

1. Mechanics

1. Chapter 1 Units and Measurement

1. [Introduction](#)
2. [The Scope and Scale of Physics](#)
3. [Unit Conversion](#)
4. [Dimensional Analysis](#)
5. [Estimates and Fermi Calculations](#)
6. [Significant Figures](#)
7. [Solving Problems in Physics](#)

2. Chapter 2 Vectors

1. [Introduction](#)
2. [Scalars and Vectors](#)
3. [Coordinate Systems and Components of a Vector](#)
4. [Algebra of Vectors](#)
5. [Products of Vectors](#)

3. Chapter 3 Motion Along a Straight Line

1. [Introduction](#)
2. [Position, Displacement, and Average Velocity](#)
3. [Instantaneous Velocity and Speed](#)
4. [Average and Instantaneous Acceleration](#)
5. [Motion with Constant Acceleration](#)
6. [Free Fall](#)
7. [Finding Velocity and Displacement from Acceleration](#)

4. Chapter 4 Motion in Two and Three Dimensions

1. [Introduction](#)
2. [Displacement and Velocity Vectors](#)
3. [Acceleration Vector](#)
4. [Projectile Motion](#)
5. [Uniform Circular Motion](#)
6. [Relative Motion in One and Two Dimensions](#)

5. Chapter 5 Newton's Laws of Motion

1. [Introduction](#)
2. [Forces](#)
3. [Newton's First Law](#)
4. [Newton's Second Law](#)
5. [Mass and Weight](#)
6. [Newton's Third Law](#)
7. [Common Forces](#)
8. [Drawing Free-Body Diagrams](#)
6. Chapter 6 Applications of Newton's Laws
 1. [Introduction](#)
 2. [Solving Problems with Newton's Laws](#)
 3. [Friction](#)
 4. [Centripetal Force](#)
 5. [Drag Force and Terminal Speed](#)
7. Chapter 7 Work and Kinetic Energy
 1. [Introduction](#)
 2. [Work](#)
 3. [Kinetic Energy](#)
 4. [Work-Energy Theorem](#)
 5. [Power](#)
8. Chapter 8 Potential Energy and Conservation of Energy
 1. [Introduction](#)
 2. [Potential Energy of a System](#)
 3. [Conservative and Non-Conservative Forces](#)
 4. [Conservation of Energy](#)
 5. [Potential Energy Diagrams and Stability](#)
 6. [Sources of Energy](#)
9. Chapter 9 Linear Momentum and Collisions
 1. [Introduction](#)
 2. [Linear Momentum](#)
 3. [Impulse and Collisions](#)
 4. [Conservation of Linear Momentum](#)

5. [Types of Collisions](#)
6. [Collisions in Multiple Dimensions](#)
7. [Center of Mass](#)
8. [Rocket Propulsion](#)
10. Chapter 10 Fixed-Axis Rotation
 1. [Introduction](#)
 2. [Rotational Variables](#)
 3. [Rotation with Constant Angular Acceleration](#)
 4. [Relating Angular and Translational Quantities](#)
 5. [Moment of Inertia and Rotational Kinetic Energy](#)
 6. [Calculating Moments of Inertia](#)
 7. [Torque](#)
 8. [Newton's Second Law for Rotation](#)
 9. [Work and Power for Rotational Motion](#)
11. Chapter 11 Angular Momentum
 1. [Introduction](#)
 2. [Rolling Motion](#)
 3. [Angular Momentum](#)
 4. [Conservation of Angular Momentum](#)
 5. [Precession of a Gyroscope](#)
12. Chapter 12 Static Equilibrium and Elasticity
 1. [Introduction](#)
 2. [Conditions for Static Equilibrium](#)
 3. [Stress, Strain, and Elastic Modulus](#)
 4. [Elasticity and Plasticity](#)
13. Chapter 13 Gravitation
 1. [Introduction](#)
 2. [Newton's Law of Universal Gravitation](#)
 3. [Gravitation Near Earth's Surface](#)
 4. [Gravitational Potential Energy and Total Energy](#)
 5. [Satellite Orbits and Energy](#)
 6. [Kepler's Laws of Planetary Motion](#)

7. [Einstein's Theory of Gravity](#)
14. Chapter 14 Fluid Mechanics
 1. [Introduction](#)
 2. [Fluids, Density, and Pressure](#)
 3. [Measuring Pressure](#)
 4. [Pascal's Principle and Hydraulics](#)
 5. [Archimedes' Principle and Buoyancy](#)
 6. [Fluid Dynamics](#)
 7. [Bernoulli's Equation](#)
 8. [Viscosity and Turbulence](#)
2. Waves and Acoustics
 1. Chapter 15 Oscillations
 1. [Introduction](#)
 2. [Simple Harmonic Motion](#)
 3. [Energy in Simple Harmonic Motion](#)
 4. [Comparing Simple Harmonic Motion and Circular Motion](#)
 5. [Pendulums](#)
 6. [Damped Oscillations](#)
 7. [Forced Oscillations](#)
 2. Chapter 16 Waves
 1. [Introduction](#)
 2. [Traveling Waves](#)
 3. [Mathematics of Waves](#)
 4. [Wave Speed on a Stretched String](#)
 5. [Energy and Power of a Wave](#)
 6. [Interference of Waves](#)
 7. [Standing Waves and Resonance](#)
 3. Chapter 17 Sound
 1. [Introduction](#)
 2. [Sound Waves](#)
 3. [Speed of Sound](#)

4. [Sound Intensity](#)
5. [Normal Modes of a Standing Sound Wave](#)
6. [Sources of Musical Sound](#)
7. [Beats](#)
8. [The Doppler Effect](#)
9. [Shock Waves](#)

3. Thermodynamics

1. Chapter 18 Temperature and Heat
 1. [Introduction](#)
 2. [Temperature and Thermal Equilibrium](#)
 3. [Thermometers and Temperature Scales](#)
 4. [Thermal Expansion](#)
 5. [Heat Transfer, Specific Heat, and Calorimetry](#)
 6. [Phase Changes](#)
 7. [Mechanisms of Heat Transfer](#)
2. Chapter 19 The Kinetic Theory of Gases
 1. [Introduction](#)
 2. [Molecular Model of an Ideal Gas](#)
 3. [Pressure, Temperature, and RMS Speed](#)
 4. [Heat Capacity and Equipartition of Energy](#)
 5. [Distribution of Molecular Speeds](#)
3. Chapter 20 The First Law of Thermodynamics
 1. [Introduction](#)
 2. [Thermodynamic Systems](#)
 3. [Work, Heat, and Internal Energy](#)
 4. [First Law of Thermodynamics](#)
 5. [Thermodynamic Processes](#)
 6. [Heat Capacities of an Ideal Gas](#)
 7. [Adiabatic Processes for an Ideal Gas](#)
4. Chapter 21 The Second Law of Thermodynamics
 1. [Introduction](#)
 2. [Reversible and Irreversible Processes](#)

3. [Heat Engines](#)
4. [Refrigerators and Heat Pumps](#)
5. [Statements of the Second Law of Thermodynamics](#)
6. [The Carnot Cycle](#)
7. [Entropy](#)
8. [Entropy on a Microscopic Scale](#)
9. [Entropy and Availability of Energy](#)
4. Electricity and Magnetism
 1. Chapter 22 Electric Charges and Fields
 1. [Introduction](#)
 2. [Electric Charge](#)
 3. [Conductors, Insulators, and Charging by Induction](#)
 4. [Coulomb's Law](#)
 5. [Electric Field](#)
 6. [Calculating Electric Fields of Charge Distributions](#)
 7. [Electric Field Lines](#)
 8. [Electric Dipoles](#)
 2. Chapter 23 Gauss's Law
 1. [Introduction](#)
 2. [Electric Flux](#)
 3. [Explaining Gauss's Law](#)
 4. [Applying Gauss's Law](#)
 5. [Conductors in Electrostatic Equilibrium](#)
 3. Chapter 24 Electric Potential
 1. [Introduction](#)
 2. [Electric Potential Energy](#)
 3. [Electric Potential and Potential Difference](#)
 4. [Calculations of Electric Potential](#)
 5. [Determining Field from Potential](#)
 6. [Equipotential Surfaces and Conductors](#)
 7. [Applications of Electrostatics](#)
 4. Chapter 25 Capacitance

1. [Introduction](#)
2. [Capacitors and Capacitance](#)
3. [Capacitors in Series and in Parallel](#)
4. [Energy Stored in a Capacitor](#)
5. [Capacitor with a Dielectric](#)
6. [Molecular Model of a Dielectric](#)
5. Chapter 26 Current and Resistance
 1. [Introduction](#)
 2. [Electrical Current](#)
 3. [Model of Conduction in Metals](#)
 4. [Resistivity and Resistance](#)
 5. [Ohm's Law](#)
 6. [Electrical Energy and Power](#)
 7. [Superconductors](#)
6. Chapter 27 Direct-Current Circuits
 1. [Introduction](#)
 2. [Electromotive Force](#)
 3. [Resistors in Series and Parallel](#)
 4. [Kirchhoff's Rules](#)
 5. [Electrical Measuring Instruments](#)
 6. [RC Circuits](#)
 7. [Household Wiring and Electrical Safety](#)
7. Chapter 28 Magnetic Forces and Fields
 1. [Introduction](#)
 2. [Magnetism and Its Historical Discoveries](#)
 3. [Magnetic Fields and Lines](#)
 4. [Motion of a Charged Particle in a Magnetic Field](#)
 5. [Magnetic Force on a Current-Carrying Conductor](#)
 6. [Force and Torque on a Current Loop](#)
 7. [The Hall Effect](#)
 8. [Applications of Magnetic Forces and Fields](#)
8. Chapter 29 Sources of Magnetic Fields

1. [Introduction](#)
2. [The Biot-Savart Law](#)
3. [Magnetic Field Due to a Thin Straight Wire](#)
4. [Magnetic Force between Two Parallel Currents](#)
5. [Magnetic Field of a Current Loop](#)
6. [Ampère's Law](#)
7. [Solenoids and Toroids](#)
8. [Magnetism in Matter](#)
9. Chapter 30 Electromagnetic Induction
 1. [Introduction](#)
 2. [Faraday's Law](#)
 3. [Lenz's Law](#)
 4. [Motional Emf](#)
 5. [Induced Electric Fields](#)
 6. [Eddy Currents](#)
 7. [Electric Generators and Back Emf](#)
 8. [Applications of Electromagnetic Induction](#)
10. Chapter 31 Inductance
 1. [Introduction](#)
 2. [Mutual Inductance](#)
 3. [Self-Inductance and Inductors](#)
 4. [Energy in a Magnetic Field](#)
 5. [RL Circuits](#)
 6. [Oscillations in an LC Circuit](#)
 7. [RLC Series Circuits](#)
11. Chapter 32 Alternating-Current Circuits
 1. [Introduction](#)
 2. [AC Sources](#)
 3. [Simple AC Circuits](#)
 4. [RLC Series Circuits with AC](#)
 5. [Power in an AC Circuit](#)
 6. [Resonance in an AC Circuit](#)

7. [Transformers](#)
12. Chapter 33 Electromagnetic Waves
 1. [Introduction](#)
 2. [Maxwell's Equations and Electromagnetic Waves](#)
 3. [Plane Electromagnetic Waves](#)
 4. [Energy Carried by Electromagnetic Waves](#)
 5. [Momentum and Radiation Pressure](#)
 6. [The Electromagnetic Spectrum](#)
5. Optics
 1. Chapter 34 The Nature of Light
 1. [Introduction](#)
 2. [The Propagation of Light](#)
 3. [The Law of Reflection](#)
 4. [Refraction](#)
 5. [Total Internal Reflection](#)
 6. [Dispersion](#)
 7. [Huygens's Principle](#)
 8. [Polarization](#)
 2. Chapter 35 Geometric Optics and Image Formation
 1. [Introduction](#)
 2. [Images Formed by Plane Mirrors](#)
 3. [Spherical Mirrors](#)
 4. [Images Formed by Refraction](#)
 5. [The Eye](#)
 6. [The Camera](#)
 7. [The Simple Magnifier](#)
 8. [Microscopes and Telescopes](#)
 3. Chapter 36 Interference
 1. [Introduction](#)
 2. [Young's Double-Slit Interference](#)
 3. [Mathematics of Interference](#)
 4. [Multiple-Slit Interference](#)

5. [Interference in Thin Films](#)
6. [The Michelson Interferometer](#)
4. Chapter 37 Diffraction
 1. [Introduction](#)
 2. [Single-Slit Diffraction](#)
 3. [Intensity in Single-Slit Diffraction](#)
 4. [Double-Slit Diffraction](#)
 5. [Circular Apertures and Resolution](#)
 6. [X-Ray Diffraction](#)
 7. [Holography](#)
6. Modern Physics
 1. Chapter 38 Relativity
 1. [Introduction](#)
 2. [Invariance of Physical Laws](#)
 3. [Relativity of Simultaneity](#)
 4. [Time Dilation](#)
 5. [Length Contraction](#)
 6. [The Lorentz Transformation](#)
 7. [Relativistic Velocity Transformation](#)
 8. [Doppler Effect for Light](#)
 9. [Relativistic Momentum](#)
 10. [Relativistic Energy](#)
 2. Chapter 39 Photons and Matter Waves
 1. [Introduction](#)
 2. [Blackbody Radiation](#)
 3. [Photoelectric Effect](#)
 4. [The Compton Effect](#)
 5. [Bohr's Model of the Hydrogen Atom](#)
 6. [De Broglie's Matter Waves](#)
 7. [Wave-Particle Duality](#)
 3. Chapter 40 Quantum Mechanics
 1. [Introduction](#)

2. [Wave Functions](#)
 3. [The Heisenberg Uncertainty Principle](#)
 4. [The Schrödinger Equation](#)
 5. [The Quantum Particle in a Box](#)
 6. [The Quantum Harmonic Oscillator](#)
 7. [The Quantum Tunneling of Particles through Potential Barriers](#)
4. Chapter 41 Atomic Structure
1. [Introduction](#)
 2. [The Hydrogen Atom](#)
 3. [Orbital Magnetic Dipole Moment of the Electron](#)
 4. [Electron Spin](#)
 5. [The Exclusion Principle and the Periodic Table](#)
 6. [Atomic Spectra and X-rays](#)
5. Chapter 42 Condensed Matter Physics
1. [Introduction](#)
 2. [Types of Molecular Bonds](#)
 3. [Molecular Spectra](#)
 4. [Bonding in Crystalline Solids](#)
 5. [Free Electron Model of Metals](#)
 6. [Band Theory of Solids](#)
 7. [Semiconductors and Doping](#)
 8. [Semiconductor Devices](#)
 9. [Superconductivity](#)
6. Chapter 43 Nuclear Physics
1. [Introduction](#)
 2. [Properties of Nuclei](#)
 3. [Nuclear Binding Energy](#)
 4. [Radioactive Decay](#)
 5. [Nuclear Reactions](#)
 6. [Fission](#)
 7. [Nuclear Fusion](#)

8. [Medical Applications and Biological Effects of Nuclear Radiation](#)
7. Chapter 44 Particle Physics and Cosmology
 1. [Introduction](#)
 2. [Introduction to Particle Physics](#)
 3. [Particle Conservation Laws](#)
 4. [Quarks](#)
 5. [Particle Accelerators and Detectors](#)
 6. [The Standard Model](#)
 7. [The Big Bang](#)
 8. [Evolution of the Early Universe](#)
7. [Units](#)
8. [Conversion Factors](#)
9. [Fundamental Constants](#)
10. [Astronomical Data](#)
11. [Chemistry](#)
12. [The Greek Alphabet](#)

Introduction

class="introduction"

This image might be showing any number of things. It might be a whirlpool in a tank of water or perhaps a collage of paint and shiny beads done for art class. Without knowing the size of the object in units we all recognize, such as meters or inches, it is difficult to know what we're looking at. In fact, this image shows the Whirlpool Galaxy (and its companion galaxy), which is about 60,000 light-years in diameter (about 6×10^{17} km across). (credit:

modification of
work by S.
Beckwith
(STScI) Hubble
Heritage Team,
(STScI/AURA)
, ESA, NASA)



As noted in the figure caption, the chapter-opening image is of the Whirlpool Galaxy, which we examine in the first section of this chapter. Galaxies are as immense as atoms are small, yet the same laws of physics describe both, along with all the rest of nature—an indication of the underlying unity in the universe. The laws of physics are surprisingly few, implying an underlying simplicity to nature’s apparent complexity. In this text, you learn about the laws of physics. Galaxies and atoms may seem far removed from your daily life, but as you begin to explore this broad-ranging subject, you may soon come to realize that physics plays a much larger role in your life than you first thought, no matter your life goals or career choice.

The Scope and Scale of Physics

By the end of this section, you will be able to:

- Describe the scope of physics.
- Calculate the order of magnitude of a quantity.
- Compare measurable length, mass, and timescales quantitatively.
- Describe the relationships among models, theories, and laws.

Physics is devoted to the understanding of all natural phenomena. In physics, we try to understand physical phenomena at all scales—from the world of subatomic particles to the entire universe. Despite the breadth of the subject, the various subfields of physics share a common core. The same basic training in physics will prepare you to work in any area of physics and the related areas of science and engineering. In this section, we investigate the scope of physics; the scales of length, mass, and time over which the laws of physics have been shown to be applicable; and the process by which science in general, and physics in particular, operates.

The Scope of Physics

Take another look at the chapter-opening image. The Whirlpool Galaxy contains billions of individual stars as well as huge clouds of gas and dust. Its companion galaxy is also visible to the right. This pair of galaxies lies a staggering billion trillion miles (1.4×10^{21} mi) from our own galaxy (which is called the *Milky Way*). The stars and planets that make up the Whirlpool Galaxy might seem to be the furthest thing from most people's everyday lives, but the Whirlpool is a great starting point to think about the forces that hold the universe together. The forces that cause the Whirlpool Galaxy to act as it does are thought to be the same forces we contend with here on Earth, whether we are planning to send a rocket into space or simply planning to raise the walls for a new home. The gravity that causes the stars of the Whirlpool Galaxy to rotate and revolve is thought to be the same as what causes water to flow over hydroelectric dams here on Earth. When you look up at the stars, realize the forces out there are the same as the ones here on Earth. Through a study of physics, you may gain a greater

understanding of the interconnectedness of everything we can see and know in this universe.

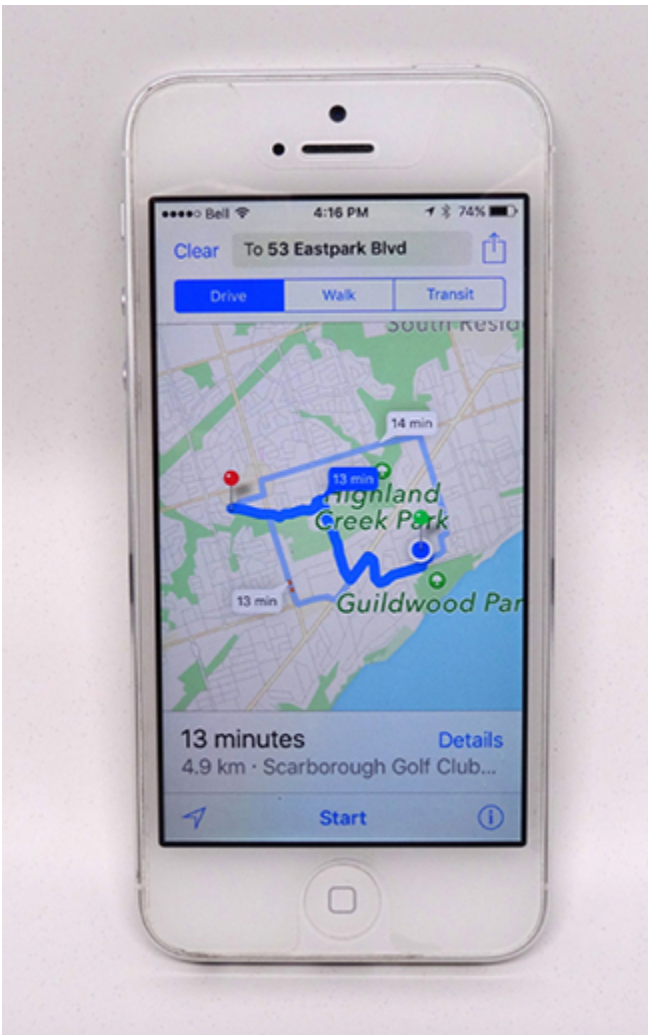
Think, now, about all the technological devices you use on a regular basis. Computers, smartphones, global positioning systems (GPSs), MP3 players, and satellite radio might come to mind. Then, think about the most exciting modern technologies you have heard about in the news, such as trains that levitate above tracks, “invisibility cloaks” that bend light around them, and microscopic robots that fight cancer cells in our bodies. All these groundbreaking advances, commonplace or unbelievable, rely on the principles of physics. Aside from playing a significant role in technology, professionals such as engineers, pilots, physicians, physical therapists, electricians, and computer programmers apply physics concepts in their daily work. For example, a pilot must understand how wind forces affect a flight path; a physical therapist must understand how the muscles in the body experience forces as they move and bend. As you will learn in this text, the principles of physics are propelling new, exciting technologies, and these principles are applied in a wide range of careers.

The underlying order of nature makes science in general, and physics in particular, interesting and enjoyable to study. For example, what do a bag of chips and a car battery have in common? Both contain energy that can be converted to other forms. The law of conservation of energy (which says that energy can change form but is never lost) ties together such topics as food calories, batteries, heat, light, and watch springs. Understanding this law makes it easier to learn about the various forms energy takes and how they relate to one another. Apparently unrelated topics are connected through broadly applicable physical laws, permitting an understanding beyond just the memorization of lists of facts.

Science consists of theories and laws that are the general truths of nature, as well as the body of knowledge they encompass. Scientists are continuously trying to expand this body of knowledge and to perfect the expression of the laws that describe it. **Physics**, which comes from the Greek *phúsis*, meaning “nature,” is concerned with describing the interactions of energy, matter, space, and time to uncover the fundamental mechanisms that underlie every

phenomenon. This concern for describing the basic phenomena in nature essentially defines the *scope of physics*.

Physics aims to understand the world around us at the most basic level. It emphasizes the use of a small number of quantitative laws to do this, which can be useful to other fields pushing the performance boundaries of existing technologies. Consider a smartphone ([\[link\]](#)). Physics describes how electricity interacts with the various circuits inside the device. This knowledge helps engineers select the appropriate materials and circuit layout when building a smartphone. Knowledge of the physics underlying these devices is required to shrink their size or increase their processing speed. Or, think about a GPS. Physics describes the relationship between the speed of an object, the distance over which it travels, and the time it takes to travel that distance. When you use a GPS in a vehicle, it relies on physics equations to determine the travel time from one location to another.



The Apple iPhone is a common smartphone with a GPS function.

Physics describes the way that electricity flows through the circuits of this device. Engineers use their knowledge of physics to construct an iPhone with features that consumers will enjoy. One specific feature of an iPhone is the GPS function. A GPS uses physics equations to determine the drive time between two locations on a map. (credit: Jane Whitney)

Knowledge of physics is useful in everyday situations as well as in nonscientific professions. It can help you understand how microwave ovens work, why metals should not be put into them, and why they might affect pacemakers. Physics allows you to understand the hazards of radiation and to evaluate these hazards rationally and more easily. Physics also explains the reason why a black car radiator helps remove heat in a car engine, and it explains why a white roof helps keep the inside of a house cool. Similarly, the operation of a car's ignition system as well as the transmission of electrical signals throughout our body's nervous system are much easier to understand when you think about them in terms of basic physics.

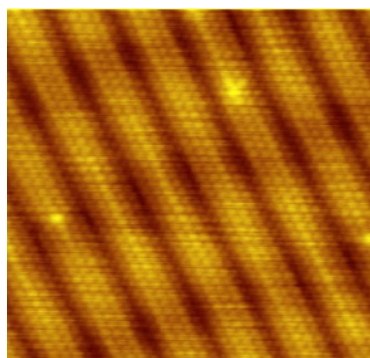
Physics is a key element of many important disciplines and contributes directly to others. Chemistry, for example—since it deals with the interactions of atoms and molecules—has close ties to atomic and molecular physics. Most branches of engineering are concerned with designing new technologies, processes, or structures within the constraints set by the laws of physics. In architecture, physics is at the heart of structural stability and is involved in the acoustics, heating, lighting, and cooling of buildings. Parts of geology rely heavily on physics, such as radioactive dating of rocks, earthquake analysis, and heat transfer within Earth. Some disciplines, such as biophysics and geophysics, are hybrids of physics and other disciplines.

Physics has many applications in the biological sciences. On the microscopic level, it helps describe the properties of cells and their environments. On the macroscopic level, it explains the heat, work, and power associated with the human body and its various organ systems. Physics is involved in medical diagnostics, such as radiographs, magnetic resonance imaging, and ultrasonic blood flow measurements. Medical therapy sometimes involves physics directly; for example, cancer radiotherapy uses ionizing radiation. Physics also explains sensory phenomena, such as how musical instruments make sound, how the eye detects color, and how lasers transmit information.

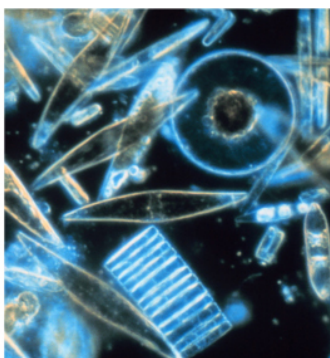
It is not necessary to study all applications of physics formally. What is most useful is knowing the basic laws of physics and developing skills in the analytical methods for applying them. The study of physics also can improve your problem-solving skills. Furthermore, physics retains the most basic aspects of science, so it is used by all the sciences, and the study of physics makes other sciences easier to understand.

The Scale of Physics

From the discussion so far, it should be clear that to accomplish your goals in any of the various fields within the natural sciences and engineering, a thorough grounding in the laws of physics is necessary. The reason for this is simply that the laws of physics govern everything in the observable universe at all measurable scales of length, mass, and time. Now, that is easy enough to say, but to come to grips with what it really means, we need to get a little bit quantitative. So, before surveying the various scales that physics allows us to explore, let's first look at the concept of "order of magnitude," which we use to come to terms with the vast ranges of length, mass, and time that we consider in this text ([\[link\]](#)).



(a)



(b)



(c)

(a) Using a scanning tunneling microscope, scientists can see the individual atoms (diameters around 10^{-10} m) that compose this sheet of gold. (b) Tiny phytoplankton swim among crystals of ice in the Antarctic Sea. They range from a few micrometers ($1\text{ }\mu\text{m}$ is 10^{-6} m) to as much as 2 mm (1 mm is 10^{-3} m) in length. (c) These two colliding

galaxies, known as NGC 4676A (right) and NGC 4676B (left), are nicknamed “The Mice” because of the tail of gas emanating from each one. They are located 300 million light-years from Earth in the constellation Coma Berenices. Eventually, these two galaxies will merge into one. (credit a: modification of work by "Erwinrossen"/Wikimedia Commons; credit b: modification of work by Prof. Gordon T. Taylor, Stony Brook University; NOAA Corps Collections; credit c: modification of work by NASA, H. Ford (JHU), G. Illingworth (UCSC/LO), M. Clampin (STScI), G. Hartig (STScI), the ACS Science Team, and ESA)

Order of magnitude

The **order of magnitude** of a number is the power of 10 that most closely approximates it. Thus, the order of magnitude refers to the scale (or size) of a value. Each power of 10 represents a different order of magnitude. For example, 10^1 , 10^2 , 10^3 , and so forth, are all different orders of magnitude, as are $10^0 = 1$, 10^{-1} , 10^{-2} , and 10^{-3} . To find the order of magnitude of a number, take the base-10 logarithm of the number and round it to the nearest integer, then the order of magnitude of the number is simply the resulting power of 10. For example, the order of magnitude of 800 is 10^3 because $\log_{10} 800 \approx 2.903$, which rounds to 3. Similarly, the order of magnitude of 450 is 10^3 because $\log_{10} 450 \approx 2.653$, which rounds to 3 as well. Thus, we say the numbers 800 and 450 are of the same order of magnitude: 10^3 . However, the order of magnitude of 250 is 10^2 because $\log_{10} 250 \approx 2.397$, which rounds to 2.

An equivalent but quicker way to find the order of magnitude of a number is first to write it in scientific notation and then check to see whether the first factor is greater than or less than $\sqrt{10} = 10^{0.5} \approx 3$. The idea is that $\sqrt{10} = 10^{0.5}$ is halfway between $1 = 10^0$ and $10 = 10^1$ on a log base-10 scale. Thus, if the first factor is less than $\sqrt{10}$, then we round it down to 1 and the order of magnitude is simply whatever power of 10 is required to

write the number in scientific notation. On the other hand, if the first factor is greater than $\sqrt{10}$, then we round it up to 10 and the order of magnitude is one power of 10 higher than the power needed to write the number in scientific notation. For example, the number 800 can be written in scientific notation as 8×10^2 . Because 8 is bigger than $\sqrt{10} \approx 3$, we say the order of magnitude of 800 is $10^{2+1} = 10^3$. The number 450 can be written as 4.5×10^2 , so its order of magnitude is also 10^3 because 4.5 is greater than 3. However, 250 written in scientific notation is 2.5×10^2 and 2.5 is less than 3, so its order of magnitude is 10^2 .

The order of magnitude of a number is designed to be a ballpark estimate for the scale (or size) of its value. It is simply a way of rounding numbers consistently to the nearest power of 10. This makes doing rough mental math with very big and very small numbers easier. For example, the diameter of a hydrogen atom is on the order of 10^{-10} m, whereas the diameter of the Sun is on the order of 10^9 m, so it would take roughly $10^9/10^{-10} = 10^{19}$ hydrogen atoms to stretch across the diameter of the Sun. This is much easier to do in your head than using the more precise values of 1.06×10^{-10} m for a hydrogen atom diameter and 1.39×10^9 m for the Sun's diameter, to find that it would take 1.31×10^{19} hydrogen atoms to stretch across the Sun's diameter. In addition to being easier, the rough estimate is also nearly as informative as the precise calculation.

Known ranges of length, mass, and time

The vastness of the universe and the breadth over which physics applies are illustrated by the wide range of examples of known lengths, masses, and times (given as orders of magnitude) in [\[link\]](#). Examining this table will give you a feeling for the range of possible topics in physics and numerical values. A good way to appreciate the vastness of the ranges of values in [\[link\]](#) is to try to answer some simple comparative questions, such as the following:

- How many hydrogen atoms does it take to stretch across the diameter of the Sun?

(Answer: $10^9 \text{ m}/10^{-10} \text{ m} = 10^{19}$ hydrogen atoms)

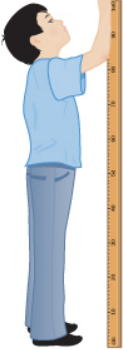
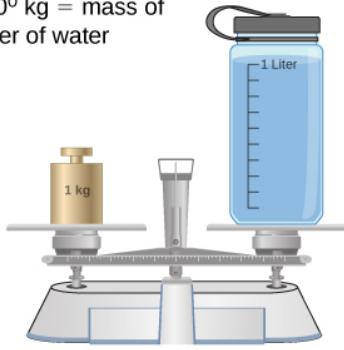
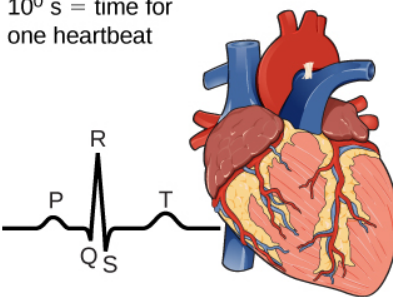
- How many protons are there in a bacterium?

(Answer: $10^{-15} \text{ kg}/10^{-27} \text{ kg} = 10^{12}$ protons)

- How many floating-point operations can a supercomputer do in 1 day?

(Answer: $10^5 \text{ s}/10^{-17} \text{ s} = 10^{22}$ floating-point operations)

In studying [\[link\]](#), take some time to come up with similar questions that interest you and then try answering them. Doing this can breathe some life into almost any table of numbers.

Length in Meters (m)	Masses in Kilograms (kg)	Time in Seconds (s)
10^{-15} m = diameter of proton	10^{-30} kg = mass of electron	10^{-22} s = mean lifetime of very unstable nucleus
10^{-14} m = diameter of large nucleus	10^{-27} kg = mass of proton	10^{-17} s = time for single floating-point operation in a supercomputer
10^{-10} m = diameter of hydrogen atom	10^{-15} kg = mass of bacterium	10^{-15} s = time for one oscillation of visible light
10^{-7} m = diameter of typical virus	10^{-5} kg = mass of mosquito	10^{-13} s = time for one vibration of an atom in a solid
10^{-2} m = pinky fingernail width	10^{-2} kg = mass of hummingbird	10^{-3} s = duration of a nerve impulse
10^0 m = height of 4 year old child 	10^0 kg = mass of liter of water 	10^0 s = time for one heartbeat 
10^2 m = length of football field	10^2 kg = mass of person	10^5 s = one day
10^7 m = diameter of Earth	10^{19} kg = mass of atmosphere	10^7 s = one year
10^{13} m = diameter of solar system	10^{22} kg = mass of Moon	10^9 s = human lifetime
10^{16} m = distance light travels in a year (one light-year)	10^{25} kg = mass of Earth	10^{11} s = recorded human history
10^{21} m = Milky Way diameter	10^{30} kg = mass of Sun	10^{17} s = age of Earth
10^{26} m = distance to edge of observable universe	10^{53} kg = upper limit on mass of known universe	10^{18} s = age of the universe

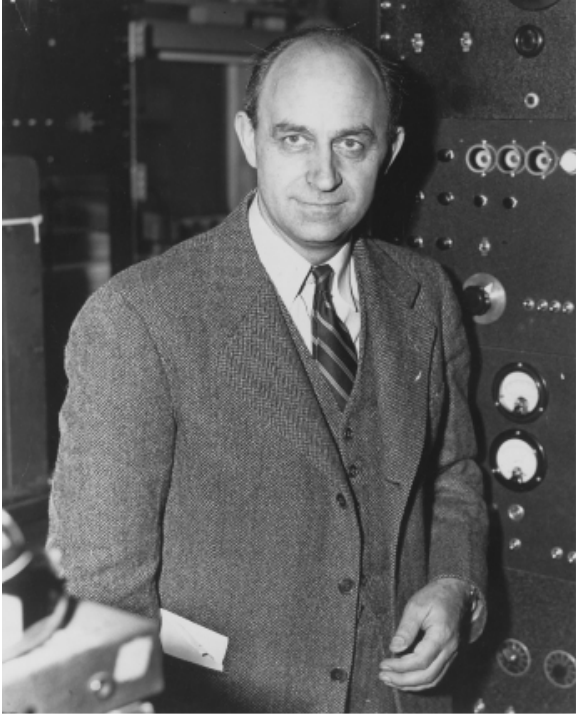
This table shows the orders of magnitude of length, mass, and time.

Note:

Visit [this site](#) to explore interactively the vast range of length scales in our universe. Scroll down and up the scale to view hundreds of organisms and objects, and click on the individual objects to learn more about each one.

Building Models

How did we come to know the laws governing natural phenomena? What we refer to as the laws of nature are concise descriptions of the universe around us. They are human statements of the underlying laws or rules that all natural processes follow. Such laws are intrinsic to the universe; humans did not create them and cannot change them. We can only discover and understand them. Their discovery is a very human endeavor, with all the elements of mystery, imagination, struggle, triumph, and disappointment inherent in any creative effort ([\[link\]](#)). The cornerstone of discovering natural laws is observation; scientists must describe the universe as it is, not as we imagine it to be.



(a) Enrico Fermi

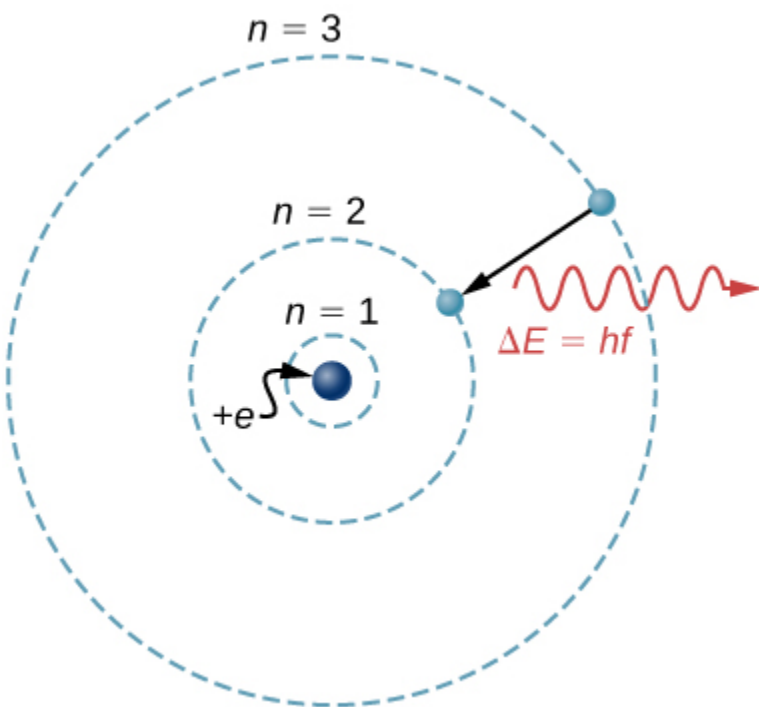


(b) Marie Curie

(a) Enrico Fermi (1901–1954) was born in Italy. On accepting the Nobel Prize in Stockholm in 1938 for his work on artificial radioactivity produced by neutrons, he took his family to America rather than return home to the government in power at the time. He became an American citizen and was a leading participant in the Manhattan Project. (b) Marie Curie (1867–1934) sacrificed monetary assets to help finance her early research and damaged her physical well-being with radiation exposure. She is the only person to win Nobel prizes in both physics and chemistry. One of her daughters also won a Nobel Prize. (credit a: modification of work by United States Department of Energy)

A **model** is a representation of something that is often too difficult (or impossible) to display directly. Although a model is justified by experimental tests, it is only accurate in describing certain aspects of a physical system. An example is the Bohr model of single-electron atoms, in which the electron is pictured as orbiting the nucleus, analogous to the way

planets orbit the Sun ([link](#)). We cannot observe electron orbits directly, but the mental image helps explain some of the observations we can make, such as the emission of light from hot gases (atomic spectra). However, other observations show that the picture in the Bohr model is not really what atoms look like. The model is “wrong,” but is still useful for some purposes. Physicists use models for a variety of purposes. For example, models can help physicists analyze a scenario and perform a calculation or models can be used to represent a situation in the form of a computer simulation. Ultimately, however, the results of these calculations and simulations need to be double-checked by other means—namely, observation and experimentation.



What is a model? The Bohr model of a single-electron atom shows the electron orbiting the nucleus in one of several possible circular orbits. Like all models, it captures some, but not all, aspects of the physical system.

The word *theory* means something different to scientists than what is often meant when the word is used in everyday conversation. In particular, to a scientist a theory is not the same as a “guess” or an “idea” or even a “hypothesis.” The phrase “it’s just a theory” seems meaningless and silly to scientists because science is founded on the notion of theories. To a scientist, a **theory** is a testable explanation for patterns in nature supported by scientific evidence and verified multiple times by various groups of researchers. Some theories include models to help visualize phenomena whereas others do not. Newton’s theory of gravity, for example, does not require a model or mental image, because we can observe the objects directly with our own senses. The kinetic theory of gases, on the other hand, is a model in which a gas is viewed as being composed of atoms and molecules. Atoms and molecules are too small to be observed directly with our senses—thus, we picture them mentally to understand what the instruments tell us about the behavior of gases. Although models are meant only to describe certain aspects of a physical system accurately, a theory should describe all aspects of any system that falls within its domain of applicability. In particular, any experimentally testable implication of a theory should be verified. If an experiment ever shows an implication of a theory to be false, then the theory is either thrown out or modified suitably (for example, by limiting its domain of applicability).

A **law** uses concise language to describe a generalized pattern in nature supported by scientific evidence and repeated experiments. Often, a law can be expressed in the form of a single mathematical equation. Laws and theories are similar in that they are both scientific statements that result from a tested hypothesis and are supported by scientific evidence. However, the designation *law* is usually reserved for a concise and very general statement that describes phenomena in nature, such as the law that energy is conserved during any process, or Newton’s second law of motion, which relates force (F), mass (m), and acceleration (a) by the simple equation $F = ma$. A theory, in contrast, is a less concise statement of observed behavior. For example, the theory of evolution and the theory of relativity cannot be expressed concisely enough to be considered laws. The biggest difference between a law and a theory is that a theory is much more complex and dynamic. A law describes a single action whereas a theory

explains an entire group of related phenomena. Less broadly applicable statements are usually called principles (such as Pascal's principle, which is applicable only in fluids), but the distinction between laws and principles often is not made carefully.

The models, theories, and laws we devise sometimes imply the existence of objects or phenomena that are as yet unobserved. These predictions are remarkable triumphs and tributes to the power of science. It is the underlying order in the universe that enables scientists to make such spectacular predictions. However, if experimentation does not verify our predictions, then the theory or law is wrong, no matter how elegant or convenient it is. Laws can never be known with absolute certainty because it is impossible to perform every imaginable experiment to confirm a law for every possible scenario. Physicists operate under the assumption that all scientific laws and theories are valid until a counterexample is observed. If a good-quality, verifiable experiment contradicts a well-established law or theory, then the law or theory must be modified or overthrown completely.

The study of science in general, and physics in particular, is an adventure much like the exploration of an uncharted ocean. Discoveries are made; models, theories, and laws are formulated; and the beauty of the physical universe is made more sublime for the insights gained.

Summary

- Physics is about trying to find the simple laws that describe all natural phenomena.
- Physics operates on a vast range of scales of length, mass, and time. Scientists use the concept of the order of magnitude of a number to track which phenomena occur on which scales. They also use orders of magnitude to compare the various scales.
- Scientists attempt to describe the world by formulating models, theories, and laws.

Conceptual Questions

Exercise:

Problem: What is physics?

Solution:

Physics is the science concerned with describing the interactions of energy, matter, space, and time to uncover the fundamental mechanisms that underlie every phenomenon.

Exercise:

Problem:

Some have described physics as a “search for simplicity.” Explain why this might be an appropriate description.

Exercise:

Problem:

If two different theories describe experimental observations equally well, can one be said to be more valid than the other (assuming both use accepted rules of logic)?

Solution:

No, neither of these two theories is more valid than the other. Experimentation is the ultimate decider. If experimental evidence does not suggest one theory over the other, then both are equally valid. A given physicist might prefer one theory over another on the grounds that one seems more simple, more natural, or more beautiful than the other, but that physicist would quickly acknowledge that he or she cannot say the other theory is invalid. Rather, he or she would be honest about the fact that more experimental evidence is needed to determine which theory is a better description of nature.

Exercise:

Problem: What determines the validity of a theory?

Exercise:**Problem:**

Certain criteria must be satisfied if a measurement or observation is to be believed. Will the criteria necessarily be as strict for an expected result as for an unexpected result?

Solution:

Probably not. As the saying goes, “Extraordinary claims require extraordinary evidence.”

Exercise:**Problem:**

Can the validity of a model be limited or must it be universally valid? How does this compare with the required validity of a theory or a law?

Problems**Exercise:****Problem:**

Find the order of magnitude of the following physical quantities. (a) The mass of Earth’s atmosphere: $5.1 \times 10^{18}\text{kg}$; (b) The mass of the Moon’s atmosphere: 25,000 kg; (c) The mass of Earth’s hydrosphere: $1.4 \times 10^{21}\text{kg}$; (d) The mass of Earth: $5.97 \times 10^{24}\text{kg}$; (e) The mass of the Moon: $7.34 \times 10^{22}\text{kg}$; (f) The Earth–Moon distance (semimajor axis): $3.84 \times 10^8\text{m}$; (g) The mean Earth–Sun distance: $1.5 \times 10^{11}\text{m}$; (h) The equatorial radius of Earth: $6.38 \times 10^6\text{m}$; (i) The mass of an electron: $9.11 \times 10^{-31}\text{kg}$; (j) The mass of a proton: $1.67 \times 10^{-27}\text{kg}$; (k) The mass of the Sun: $1.99 \times 10^{30}\text{kg}$.

Exercise:

Problem:

Use the orders of magnitude you found in the previous problem to answer the following questions to within an order of magnitude. (a) How many electrons would it take to equal the mass of a proton? (b) How many Earths would it take to equal the mass of the Sun? (c) How many Earth–Moon distances would it take to cover the distance from Earth to the Sun? (d) How many Moon atmospheres would it take to equal the mass of Earth’s atmosphere? (e) How many moons would it take to equal the mass of Earth? (f) How many protons would it take to equal the mass of the Sun?

Solution:

a. 10^3 ; b. 10^5 ; c. 10^2 ; d. 10^{15} ; e. 10^2 ; f. 10^{57}

For the remaining questions, you need to use [\[link\]](#) to obtain the necessary orders of magnitude of lengths, masses, and times.

Exercise:

Problem: Roughly how many heartbeats are there in a lifetime?

Exercise:**Problem:**

A generation is about one-third of a lifetime. Approximately how many generations have passed since the year 0?

Solution:

10^2 generations

Exercise:**Problem:**

Roughly how many times longer than the mean life of an extremely unstable atomic nucleus is the lifetime of a human?

Exercise:**Problem:**

Calculate the approximate number of atoms in a bacterium. Assume the average mass of an atom in the bacterium is 10 times the mass of a proton.

Solution:

10^{11} atoms

Exercise:**Problem:**

(a) Calculate the number of cells in a hummingbird assuming the mass of an average cell is 10 times the mass of a bacterium. (b) Making the same assumption, how many cells are there in a human?

Exercise:**Problem:**

Assuming one nerve impulse must end before another can begin, what is the maximum firing rate of a nerve in impulses per second?

Solution:

10^3 nerve impulses/s

Exercise:**Problem:**

About how many floating-point operations can a supercomputer perform each year?

Exercise:

Problem:

Roughly how many floating-point operations can a supercomputer perform in a human lifetime?

Solution:

10^{26} floating-point operations per human lifetime

Glossary

law

description, using concise language or a mathematical formula, of a generalized pattern in nature supported by scientific evidence and repeated experiments

model

representation of something often too difficult (or impossible) to display directly

order of magnitude

the size of a quantity as it relates to a power of 10

physics

science concerned with describing the interactions of energy, matter, space, and time; especially interested in what fundamental mechanisms underlie every phenomenon

theory

testable explanation for patterns in nature supported by scientific evidence and verified multiple times by various groups of researchers

Unit Conversion

By the end of this section, you will be able to:

- Use conversion factors to express the value of a given quantity in different units.

It is often necessary to convert from one unit to another. For example, if you are reading a European cookbook, some quantities may be expressed in units of liters and you need to convert them to cups. Or perhaps you are reading walking directions from one location to another and you are interested in how many miles you will be walking. In this case, you may need to convert units of feet or meters to miles.

Let's consider a simple example of how to convert units. Suppose we want to convert 80 m to kilometers. The first thing to do is to list the units you have and the units to which you want to convert. In this case, we have units in *meters* and we want to convert to *kilometers*. Next, we need to determine a conversion factor relating meters to kilometers. A **conversion factor** is a ratio that expresses how many of one unit are equal to another unit. For example, there are 12 in. in 1 ft, 1609 m in 1 mi, 100 cm in 1 m, 60 s in 1 min, and so on. Refer to [Appendix B](#) for a more complete list of conversion factors. In this case, we know that there are 1000 m in 1 km. Now we can set up our unit conversion. We write the units we have and then multiply them by the conversion factor so the units cancel out, as shown:

Equation:

$$80 \cancel{\text{m}} \times \frac{1 \text{ km}}{1000 \cancel{\text{m}}} = 0.080 \text{ km}.$$

Note that the unwanted meter unit cancels, leaving only the desired kilometer unit. You can use this method to convert between any type of unit. Now, the conversion of 80 m to kilometers is simply the use of a metric prefix, as we saw in the preceding section, so we can get the same answer just as easily by noting that

Equation:

$$80 \text{ m} = 8.0 \times 10^1 \text{ m} = 8.0 \times 10^{-2} \text{ km} = 0.080 \text{ km},$$

since “kilo-” means 10^3 (see [\[link\]](#)) and $1 = -2 + 3$. However, using conversion factors is handy when converting between units that are not metric or when converting between derived units, as the following examples illustrate.

Example:

Converting Nonmetric Units to Metric

The distance from the university to home is 10 mi and it usually takes 20 min to drive this distance. Calculate the average speed in meters per second (m/s). (*Note:* Average speed is distance traveled divided by time of travel.)

Strategy

First we calculate the average speed using the given units, then we can get the average speed into the desired units by picking the correct conversion factors and multiplying by them. The correct conversion factors are those that cancel the unwanted units and leave the desired units in their place. In this case, we want to convert miles to meters, so we need to know the fact that there are 1609 m in 1 mi. We also want to convert minutes to seconds, so we use the conversion of 60 s in 1 min.

Solution

1. Calculate average speed. Average speed is distance traveled divided by time of travel. (Take this definition as a given for now. Average speed and other motion concepts are covered in later chapters.) In equation form,

Equation:

$$\text{Average speed} = \frac{\text{Distance}}{\text{Time}}.$$

2. Substitute the given values for distance and time:

Equation:

$$\text{Average speed} = \frac{10 \text{ mi}}{20 \text{ min}} = 0.50 \frac{\text{mi}}{\text{min}}.$$

3. Convert miles per minute to meters per second by multiplying by the conversion factor that cancels miles and leave meters, and also by the conversion factor that cancels minutes and leave seconds:

Equation:

$$0.50 \frac{\cancel{\text{mile}}}{\cancel{\text{min}}} \times \frac{1609 \text{ m}}{1 \cancel{\text{mile}}} \times \frac{1 \cancel{\text{min}}}{60 \text{ s}} = \frac{(0.50)(1609)}{60} \text{ m/s} = 13 \text{ m/s}.$$

Significance

Check the answer in the following ways:

1. Be sure the units in the unit conversion cancel correctly. If the unit conversion factor was written upside down, the units do not cancel correctly in the equation. We see the “miles” in the numerator in 0.50 mi/min cancels the “mile” in the

denominator in the first conversion factor. Also, the “min” in the denominator in 0.50 mi/min cancels the “min” in the numerator in the second conversion factor.

2. Check that the units of the final answer are the desired units. The problem asked us to solve for average speed in units of meters per second and, after the cancellations, the only units left are a meter (m) in the numerator and a second (s) in the denominator, so we have indeed obtained these units.

Note:**Exercise:****Problem:**

Check Your Understanding Light travels about 9 Pm in a year. Given that a year is about 3×10^7 s, what is the speed of light in meters per second?

Solution:

$$3 \times 10^8 \text{ m/s}$$

Example:**Converting between Metric Units**

The density of iron is 7.86 g/cm^3 under standard conditions. Convert this to kg/m^3 .

Strategy

We need to convert grams to kilograms and cubic centimeters to cubic meters. The conversion factors we need are $1 \text{ kg} = 10^3 \text{ g}$ and $1 \text{ cm} = 10^{-2} \text{ m}$. However, we are dealing with cubic centimeters ($\text{cm}^3 = \text{cm} \times \text{cm} \times \text{cm}$), so we have to use the second conversion factor three times (that is, we need to cube it). The idea is still to multiply by the conversion factors in such a way that they cancel the units we want to get rid of and introduce the units we want to keep.

Solution**Equation:**

$$7.86 \frac{\cancel{\text{g}}}{\cancel{\text{cm}}^3} \times \frac{\text{kg}}{10^3 \cancel{\text{g}}} \times \left(\frac{\cancel{\text{cm}}}{10^{-2} \text{ m}} \right)^3 = \frac{7.86}{(10^3)(10^{-6})} \text{ kg/m}^3 = 7.86 \times 10^3 \text{ kg/m}^3$$

Significance

Remember, it's always important to check the answer.

1. Be sure to cancel the units in the unit conversion correctly. We see that the gram (“g”) in the numerator in 7.86 g/cm^3 cancels the “g” in the denominator in the first conversion factor. Also, the three factors of “cm” in the denominator in 7.86 g/cm^3 cancel with the three factors of “cm” in the numerator that we get by cubing the second conversion factor.
2. Check that the units of the final answer are the desired units. The problem asked for us to convert to kilograms per cubic meter. After the cancellations just described, we see the only units we have left are “kg” in the numerator and three factors of “m” in the denominator (that is, one factor of “m” cubed, or “m³”). Therefore, the units on the final answer are correct.

Note:**Exercise:****Problem:**

Check Your Understanding We know from [\[link\]](#) that the diameter of Earth is on the order of 10^7 m , so the order of magnitude of its surface area is 10^{14} m^2 . What is that in square kilometers (that is, km^2)? (Try doing this both by converting 10^7 m to km and then squaring it and then by converting 10^{14} m^2 directly to square kilometers. You should get the same answer both ways.)

Solution:

$$10^8 \text{ km}^2$$

Unit conversions may not seem very interesting, but not doing them can be costly. One famous example of this situation was seen with the *Mars Climate Orbiter*. This probe was launched by NASA on December 11, 1998. On September 23, 1999, while attempting to guide the probe into its planned orbit around Mars, NASA lost contact with it. Subsequent investigations showed a piece of software called SM_FORCES (or “small forces”) was recording thruster performance data in the English units of pound-seconds (lb-s). However, other pieces of software that used these values for course corrections expected them to be recorded in the SI units of newton-seconds (N-s), as dictated in the software interface protocols. This error caused the probe to follow a very different trajectory from what NASA thought it was following, which most likely caused the probe either to burn up in the Martian atmosphere or to shoot out into space. This failure to pay attention to unit conversions cost hundreds of millions of dollars,

not to mention all the time invested by the scientists and engineers who worked on the project.

Note:

Exercise:

Problem:

Check Your Understanding Given that 1 lb (pound) is 4.45 N, were the numbers being output by SM_FORCES too big or too small?

Solution:

The numbers were too small, by a factor of 4.45.

Summary

- To convert a quantity from one unit to another, multiply by conversions factors in such a way that you cancel the units you want to get rid of and introduce the units you want to end up with.
- Be careful with areas and volumes. Units obey the rules of algebra so, for example, if a unit is squared we need two factors to cancel it.

Problems

Exercise:

Problem:

The volume of Earth is on the order of 10^{21} m^3 . (a) What is this in cubic kilometers (km^3)? (b) What is it in cubic miles (mi^3)? (c) What is it in cubic centimeters (cm^3)?

Exercise:

Problem:

The speed limit on some interstate highways is roughly 100 km/h. (a) What is this in meters per second? (b) How many miles per hour is this?

Solution:

a. 27.8 m/s; b. 62 mi/h

Exercise:

Problem:

A car is traveling at a speed of 33 m/s. (a) What is its speed in kilometers per hour? (b) Is it exceeding the 90 km/h speed limit?

Exercise:

Problem:

In SI units, speeds are measured in meters per second (m/s). But, depending on where you live, you're probably more comfortable of thinking of speeds in terms of either kilometers per hour (km/h) or miles per hour (mi/h). In this problem, you will see that 1 m/s is roughly 4 km/h or 2 mi/h, which is handy to use when developing your physical intuition. More precisely, show that (a) $1.0 \text{ m/s} = 3.6 \text{ km/h}$ and (b) $1.0 \text{ m/s} = 2.2 \text{ mi/h}$.

Solution:

a. 3.6 km/h; b. 2.2 mi/h

Exercise:

Problem:

American football is played on a 100-yd-long field, excluding the end zones. How long is the field in meters? (Assume that $1 \text{ m} = 3.281 \text{ ft}$.)

Exercise:

Problem:

Soccer fields vary in size. A large soccer field is 115 m long and 85.0 m wide. What is its area in square feet? (Assume that $1 \text{ m} = 3.281 \text{ ft}$.)

Solution:

$$1.05 \times 10^5 \text{ ft}^2$$

Exercise:

Problem: What is the height in meters of a person who is 6 ft 1.0 in. tall?

Exercise:

Problem:

Mount Everest, at 29,028 ft, is the tallest mountain on Earth. What is its height in kilometers? (Assume that 1 m = 3.281 ft.)

Solution:

8.847 km

Exercise:**Problem:**

The speed of sound is measured to be 342 m/s on a certain day. What is this measurement in kilometers per hour?

Exercise:**Problem:**

Tectonic plates are large segments of Earth's crust that move slowly. Suppose one such plate has an average speed of 4.0 cm/yr. (a) What distance does it move in 1.0 s at this speed? (b) What is its speed in kilometers per million years?

Solution:

a. 1.3×10^{-9} m; b. 40 km/My

Exercise:**Problem:**

The average distance between Earth and the Sun is 1.5×10^{11} m. (a) Calculate the average speed of Earth in its orbit (assumed to be circular) in meters per second. (b) What is this speed in miles per hour?

Exercise:**Problem:**

The density of nuclear matter is about 10^{18} kg/m³. Given that 1 mL is equal in volume to cm³, what is the density of nuclear matter in megagrams per microliter (that is, Mg/ L)?

Solution:

10^6 Mg/ L

Exercise:

Problem:

The density of aluminum is 2.7 g/cm^3 . What is the density in kilograms per cubic meter?

Exercise:

Problem:

A commonly used unit of mass in the English system is the pound-mass, abbreviated lbm, where $1 \text{ lbm} = 0.454 \text{ kg}$. What is the density of water in pound-mass per cubic foot?

Solution:

62.4 lbm/ft^3

Exercise:

Problem:

A furlong is 220 yd. A fortnight is 2 weeks. Convert a speed of one furlong per fortnight to millimeters per second.

Exercise:

Problem:

It takes 2 radians (rad) to get around a circle, which is the same as 360° . How many radians are in 1° ?

Solution:

0.017 rad

Exercise:

Problem:

Light travels a distance of about $3 \times 10^8 \text{ m/s}$. A light-minute is the distance light travels in 1 min. If the Sun is $1.5 \times 10^{11} \text{ m}$ from Earth, how far away is it in light-minutes?

Exercise:

Problem:

A light-nanosecond is the distance light travels in 1 ns. Convert 1 ft to light-nanoseconds.

Solution:

1 light-nanosecond

Exercise:**Problem:**

An electron has a mass of 9.11×10^{-31} kg. A proton has a mass of 1.67×10^{-27} kg. What is the mass of a proton in electron-masses?

Exercise:**Problem:**

A fluid ounce is about 30 mL. What is the volume of a 12 fl-oz can of soda pop in cubic meters?

Solution:

$3.6 \times 10^{-4} \text{ m}^3$

Glossary

conversion factor

a ratio that expresses how many of one unit are equal to another unit

Dimensional Analysis

By the end of this section, you will be able to:

- Find the dimensions of a mathematical expression involving physical quantities.
- Determine whether an equation involving physical quantities is dimensionally consistent.

The **dimension** of any physical quantity expresses its dependence on the base quantities as a product of symbols (or powers of symbols) representing the base quantities. [\[link\]](#) lists the base quantities and the symbols used for their dimension. For example, a measurement of length is said to have dimension L or L^1 , a measurement of mass has dimension M or M^1 , and a measurement of time has dimension T or T^1 . Like units, dimensions obey the rules of algebra. Thus, area is the product of two lengths and so has dimension L^2 , or length squared. Similarly, volume is the product of three lengths and has dimension L^3 , or length cubed. Speed has dimension length over time, L/T or LT^{-1} . Volumetric mass density has dimension M/L^3 or ML^{-3} , or mass over length cubed. In general, the dimension of any physical quantity can be written as $L^a M^b T^c I^d \Theta^e N^f J^g$ for some powers a, b, c, d, e, f , and g . We can write the dimensions of a length in this form with $a = 1$ and the remaining six powers all set equal to zero: $L^1 = L^1 M^0 T^0 I^0 \Theta^0 N^0 J^0$. Any quantity with a dimension that can be written so that all seven powers are zero (that is, its dimension is $L^0 M^0 T^0 I^0 \Theta^0 N^0 J^0$) is called **dimensionless** (or sometimes “of dimension 1,” because anything raised to the zero power is one). Physicists often call dimensionless quantities *pure numbers*.

Base Quantity	Symbol for Dimension
Length	L

Base Quantity	Symbol for Dimension
Mass	M
Time	T
Current	I
Thermodynamic temperature	Θ
Amount of substance	N
Luminous intensity	J

Base Quantities and Their Dimensions

Physicists often use square brackets around the symbol for a physical quantity to represent the dimensions of that quantity. For example, if r is the radius of a cylinder and h is its height, then we write $[r] = L$ and $[h] = L$ to indicate the dimensions of the radius and height are both those of length, or L . Similarly, if we use the symbol A for the surface area of a cylinder and V for its volume, then $[A] = L^2$ and $[V] = L^3$. If we use the symbol m for the mass of the cylinder and ρ for the density of the material from which the cylinder is made, then $[m] = M$ and $[\rho] = ML^{-3}$.

The importance of the concept of dimension arises from the fact that any mathematical equation relating physical quantities must be **dimensionally consistent**, which means the equation must obey the following rules:

- Every term in an expression must have the same dimensions; it does not make sense to add or subtract quantities of differing dimension (think of the old saying: “You can’t add apples and oranges”). In particular, the expressions on each side of the equality in an equation must have the same dimensions.
- The arguments of any of the standard mathematical functions such as trigonometric functions (such as sine and cosine), logarithms, or exponential functions that appear in the equation must be

dimensionless. These functions require pure numbers as inputs and give pure numbers as outputs.

If either of these rules is violated, an equation is not dimensionally consistent and cannot possibly be a correct statement of physical law. This simple fact can be used to check for typos or algebra mistakes, to help remember the various laws of physics, and even to suggest the form that new laws of physics might take. This last use of dimensions is beyond the scope of this text, but is something you will undoubtedly learn later in your academic career.

Example:**Using Dimensions to Remember an Equation**

Suppose we need the formula for the area of a circle for some computation. Like many people who learned geometry too long ago to recall with any certainty, two expressions may pop into our mind when we think of circles: πr^2 and $2\pi r$. One expression is the circumference of a circle of radius r and the other is its area. But which is which?

Strategy

One natural strategy is to look it up, but this could take time to find information from a reputable source. Besides, even if we think the source is reputable, we shouldn't trust everything we read. It is nice to have a way to double-check just by thinking about it. Also, we might be in a situation in which we cannot look things up (such as during a test). Thus, the strategy is to find the dimensions of both expressions by making use of the fact that dimensions follow the rules of algebra. If either expression does not have the same dimensions as area, then it cannot possibly be the correct equation for the area of a circle.

Solution

We know the dimension of area is L^2 . Now, the dimension of the expression πr^2 is

Equation:

$$[\pi r^2] = [\pi] \cdot [r]^2 = 1 \cdot L^2 = L^2,$$

since the constant π is a pure number and the radius r is a length. Therefore, πr^2 has the dimension of area. Similarly, the dimension of the expression $2\pi r$ is

Equation:

$$[2\pi r] = [2] \cdot [\pi] \cdot [r] = 1 \cdot 1 \cdot \text{L} = \text{L},$$

since the constants 2 and π are both dimensionless and the radius r is a length. We see that $2\pi r$ has the dimension of length, which means it cannot possibly be an area.

We rule out $2\pi r$ because it is not dimensionally consistent with being an area. We see that πr^2 is dimensionally consistent with being an area, so if we have to choose between these two expressions, πr^2 is the one to choose.

Significance

This may seem like kind of a silly example, but the ideas are very general. As long as we know the dimensions of the individual physical quantities that appear in an equation, we can check to see whether the equation is dimensionally consistent. On the other hand, knowing that true equations are dimensionally consistent, we can match expressions from our imperfect memories to the quantities for which they might be expressions. Doing this will not help us remember dimensionless factors that appear in the equations (for example, if you had accidentally conflated the two expressions from the example into $2\pi r^2$, then dimensional analysis is no help), but it does help us remember the correct basic form of equations.

Note:

Exercise:

Problem:

Check Your Understanding Suppose we want the formula for the volume of a sphere. The two expressions commonly mentioned in elementary discussions of spheres are $4\pi r^2$ and $4\pi r^3/3$. One is the volume of a sphere of radius r and the other is its surface area. Which one is the volume?

Solution:

$$4\pi r^3/3$$

Example:**Checking Equations for Dimensional Consistency**

Consider the physical quantities s , v , a , and t with dimensions $[s] = \text{L}$, $[v] = \text{LT}^{-1}$, $[a] = \text{LT}^{-2}$, and $[t] = \text{T}$. Determine whether each of the following equations is dimensionally consistent: (a) $s = vt + 0.5at^2$; (b) $s = vt^2 + 0.5at$; and (c) $v = \sin(at^2/s)$.

Strategy

By the definition of dimensional consistency, we need to check that each term in a given equation has the same dimensions as the other terms in that equation and that the arguments of any standard mathematical functions are dimensionless.

Solution

- a. There are no trigonometric, logarithmic, or exponential functions to worry about in this equation, so we need only look at the dimensions of each term appearing in the equation. There are three terms, one in the left expression and two in the expression on the right, so we look at each in turn:

Equation:

$$[s] = \text{L}$$

$$[vt] = [v] \cdot [t] = \text{LT}^{-1} \cdot \text{T} = \text{LT}^0 = \text{L}$$

$$[0.5at^2] = [a] \cdot [t]^2 = \text{LT}^{-2} \cdot \text{T}^2 = \text{LT}^0 = \text{L}.$$

All three terms have the same dimension, so this equation is dimensionally consistent.

- b. Again, there are no trigonometric, exponential, or logarithmic functions, so we only need to look at the dimensions of each of the three terms appearing in the equation:

Equation:

$$[s] = \text{L}$$

$$[vt^2] = [v] \cdot [t]^2 = \text{LT}^{-1} \cdot \text{T}^2 = \text{LT}$$

$$[at] = [a] \cdot [t] = \text{LT}^{-2} \cdot \text{T} = \text{LT}^{-1}.$$

None of the three terms has the same dimension as any other, so this is about as far from being dimensionally consistent as you can get.

The technical term for an equation like this is *nonsense*.

- c. This equation has a trigonometric function in it, so first we should check that the argument of the sine function is dimensionless:

Equation:

$$\left[\frac{at^2}{s} \right] = \frac{[a] \cdot [t]^2}{[s]} = \frac{\text{LT}^{-2} \cdot \text{T}^2}{\text{L}} = \frac{\text{L}}{\text{L}} = 1.$$

The argument is dimensionless. So far, so good. Now we need to check the dimensions of each of the two terms (that is, the left expression and the right expression) in the equation:

Equation:

$$[v] = \text{LT}^{-1}$$

$$\left[\sin \left(\frac{at^2}{s} \right) \right] = 1.$$

The two terms have different dimensions—meaning, the equation is not dimensionally consistent. This equation is another example of “nonsense.”

Significance

If we are trusting people, these types of dimensional checks might seem unnecessary. But, rest assured, any textbook on a quantitative subject such as physics (including this one) almost certainly contains some equations with typos. Checking equations routinely by dimensional analysis save us the embarrassment of using an incorrect equation. Also, checking the dimensions of an equation we obtain through algebraic manipulation is a great way to make sure we did not make a mistake (or to spot a mistake, if we made one).

Note:

Exercise:

Problem:

Check Your Understanding Is the equation $v = at$ dimensionally consistent?

Solution:

yes

One further point that needs to be mentioned is the effect of the operations of calculus on dimensions. We have seen that dimensions obey the rules of algebra, just like units, but what happens when we take the derivative of one physical quantity with respect to another or integrate a physical quantity over another? The derivative of a function is just the slope of the line tangent to its graph and slopes are ratios, so for physical quantities v and t , we have that the dimension of the derivative of v with respect to t is just the ratio of the dimension of v over that of t :

Equation:

$$\left[\frac{dv}{dt} \right] = \frac{[v]}{[t]}.$$

Similarly, since integrals are just sums of products, the dimension of the integral of v with respect to t is simply the dimension of v times the dimension of t :

Equation:

$$\left[\int v dt \right] = [v] \cdot [t].$$

By the same reasoning, analogous rules hold for the units of physical quantities derived from other quantities by integration or differentiation.

Summary

- The dimension of a physical quantity is just an expression of the base quantities from which it is derived.
- All equations expressing physical laws or principles must be dimensionally consistent. This fact can be used as an aid in remembering physical laws, as a way to check whether claimed relationships between physical quantities are possible, and even to derive new physical laws.

Problems

Exercise:

Problem:

A student is trying to remember some formulas from geometry. In what follows, assume A is area, V is volume, and all other variables are lengths. Determine which formulas are dimensionally consistent. (a) $V = \pi r^2 h$; (b) $A = 2\pi r^2 + 2\pi r h$; (c) $V = 0.5bh$; (d) $V = \pi d^2$; (e) $V = \pi d^3/6$.

Exercise:

Problem:

Consider the physical quantities s , v , a , and t with dimensions $[s] = L$, $[v] = LT^{-1}$, $[a] = LT^{-2}$, and $[t] = T$. Determine whether each of the following equations is dimensionally consistent. (a) $v^2 = 2as$; (b) $s = vt^2 + 0.5at^2$; (c) $v = s/t$; (d) $a = v/t$.

Solution:

a. Yes, both terms have dimension L^2T^{-2} b. No. c. Yes, both terms have dimension LT^{-1} d. Yes, both terms have dimension LT^{-2}

Exercise:**Problem:**

Consider the physical quantities m , s , v , a , and t with dimensions $[m] = M$, $[s] = L$, $[v] = LT^{-1}$, $[a] = LT^{-2}$, and $[t] = T$. Assuming each of the following equations is dimensionally consistent, find the dimension of the quantity on the left-hand side of the equation: (a) $F = ma$; (b) $K = 0.5mv^2$; (c) $p = mv$; (d) $W = mas$; (e) $L = mvr$.

Exercise:**Problem:**

Suppose quantity s is a length and quantity t is a time. Suppose the quantities v and a are defined by $v = ds/dt$ and $a = dv/dt$. (a) What is the dimension of v ? (b) What is the dimension of the quantity a ? What are the dimensions of (c) $\int v dt$, (d) $\int a dt$, and (e) da/dt ?

Solution:

$$\text{a. } [v] = LT^{-1}; \text{ b. } [a] = LT^{-2}; \text{ c. } \left[\int v dt \right] = L; \text{ d. } \left[\int a dt \right] = LT^{-1}; \text{ e. } \left[\frac{da}{dt} \right] = LT^{-3}$$

Exercise:**Problem:**

Suppose $[V] = L^3$, $[\rho] = ML^{-3}$, and $[t] = T$. (a) What is the dimension of $\int \rho dV$? (b) What is the dimension of dV/dt ? (c) What is the dimension of $\rho(dV/dt)$?

Exercise:

Problem:

The arc length formula says the length s of arc subtended by angle Θ in a circle of radius r is given by the equation $s = r\Theta$. What are the dimensions of (a) s , (b) r , and (c) Θ ?

Solution:

a. L; b. L; c. $L^0 = 1$ (that is, it is dimensionless)

Glossary**dimension**

expression of the dependence of a physical quantity on the base quantities as a product of powers of symbols representing the base quantities; in general, the dimension of a quantity has the form $L^a M^b T^c I^d \Theta^e N^f J^g$ for some powers a, b, c, d, e, f , and g .

dimensionally consistent

equation in which every term has the same dimensions and the arguments of any mathematical functions appearing in the equation are dimensionless

dimensionless

quantity with a dimension of $L^0 M^0 T^0 I^0 \Theta^0 N^0 J^0 = 1$; also called quantity of dimension 1 or a pure number

Estimates and Fermi Calculations

By the end of this section, you will be able to:

- Estimate the values of physical quantities.

On many occasions, physicists, other scientists, and engineers need to make *estimates* for a particular quantity. Other terms sometimes used are *guesstimates*, *order-of-magnitude approximations*, *back-of-the-envelope calculations*, or *Fermi calculations*. (The physicist Enrico Fermi mentioned earlier was famous for his ability to estimate various kinds of data with surprising precision.) Will that piece of equipment fit in the back of the car or do we need to rent a truck? How long will this download take? About how large a current will there be in this circuit when it is turned on? How many houses could a proposed power plant actually power if it is built? Note that estimating does not mean guessing a number or a formula at random. Rather, **estimation** means using prior experience and sound physical reasoning to arrive at a rough idea of a quantity's value. Because the process of determining a reliable approximation usually involves the identification of correct physical principles and a good guess about the relevant variables, estimating is very useful in developing physical intuition. Estimates also allow us to perform “sanity checks” on calculations or policy proposals by helping us rule out certain scenarios or unrealistic numbers. They allow us to challenge others (as well as ourselves) in our efforts to learn truths about the world.

Many estimates are based on formulas in which the input quantities are known only to a limited precision. As you develop physics problem-solving skills (which are applicable to a wide variety of fields), you also will develop skills at estimating. You develop these skills by thinking more quantitatively and by being willing to take risks. As with any skill, experience helps. Familiarity with dimensions (see [\[link\]](#)) and units (see [\[link\]](#) and [\[link\]](#)), and the scales of base quantities (see [\[link\]](#)) also helps.

To make some progress in estimating, you need to have some definite ideas about how variables may be related. The following strategies may help you in practicing the art of estimation:

- *Get big lengths from smaller lengths.* When estimating lengths, remember that anything can be a ruler. Thus, imagine breaking a big thing into smaller things, estimate the length of one of the smaller things, and multiply to get the length of the big thing. For example, to estimate the height of a building, first count how many floors it has. Then, estimate how big a single floor is by imagining how many people would have to stand on each other's shoulders to reach the ceiling. Last, estimate the height of a person. The product of these three estimates is your estimate of the height of the building. It helps to have memorized a few length scales relevant to the sorts of problems you find yourself solving. For example, knowing some of the length scales in [\[link\]](#) might come in handy. Sometimes it also helps to do this in reverse—that is, to estimate the length of a small thing, imagine a bunch of them making up a bigger thing. For example, to estimate the thickness of a sheet of paper, estimate the thickness of a stack of paper and then divide by the number of pages in the stack. These same strategies of breaking big things into smaller things or aggregating smaller things into a bigger thing can sometimes be used to estimate other physical quantities, such as masses and times.
- *Get areas and volumes from lengths.* When dealing with an area or a volume of a complex object, introduce a simple model of the object such as a sphere or a box. Then, estimate the linear dimensions (such as the radius of the sphere or the length, width, and height of the box) first, and use your estimates to obtain the volume or area from standard geometric formulas. If you happen to have an estimate of an object's area or volume, you can also do the reverse; that is, use standard geometric formulas to get an estimate of its linear dimensions.
- *Get masses from volumes and densities.* When estimating masses of objects, it can help first to estimate its volume and then to estimate its mass from a rough estimate of its average density (recall, density has dimension mass over length cubed, so mass is density times volume). For this, it helps to remember that the density of air is around 1 kg/m^3 , the density of water is 10^3 kg/m^3 , and the densest everyday solids max out at around 10^4 kg/m^3 . Asking yourself whether an object floats or sinks in either air or water gets you a ballpark estimate of its density. You can also do this the other way around; if you have an estimate of

an object's mass and its density, you can use them to get an estimate of its volume.

- *If all else fails, bound it.* For physical quantities for which you do not have a lot of intuition, sometimes the best you can do is think something like: Well, it must be bigger than this and smaller than that. For example, suppose you need to estimate the mass of a moose. Maybe you have a lot of experience with moose and know their average mass offhand. If so, great. But for most people, the best they can do is to think something like: It must be bigger than a person (of order 10^2 kg) and less than a car (of order 10^3 kg). If you need a single number for a subsequent calculation, you can take the geometric mean of the upper and lower bound—that is, you multiply them together and then take the square root. For the moose mass example, this would be **Equation:**

$$(10^2 \times 10^3)^{0.5} = 10^{2.5} = 10^{0.5} \times 10^2 \approx 3 \times 10^2 \text{kg}.$$

The tighter the bounds, the better. Also, no rules are unbreakable when it comes to estimation. If you think the value of the quantity is likely to be closer to the upper bound than the lower bound, then you may want to bump up your estimate from the geometric mean by an order or two of magnitude.

- *One “sig. fig.” is fine.* There is no need to go beyond one significant figure when doing calculations to obtain an estimate. In most cases, the order of magnitude is good enough. The goal is just to get in the ballpark figure, so keep the arithmetic as simple as possible.
- *Ask yourself: Does this make any sense?* Last, check to see whether your answer is reasonable. How does it compare with the values of other quantities with the same dimensions that you already know or can look up easily? If you get some wacky answer (for example, if you estimate the mass of the Atlantic Ocean to be bigger than the mass of Earth, or some time span to be longer than the age of the universe), first check to see whether your units are correct. Then, check for arithmetic errors. Then, rethink the logic you used to arrive at your answer. If everything checks out, you may have just proved that some slick new idea is actually bogus.

Example:**Mass of Earth's Oceans**

Estimate the total mass of the oceans on Earth.

Strategy

We know the density of water is about 10^3 kg/m^3 , so we start with the advice to “get masses from densities and volumes.” Thus, we need to estimate the volume of the planet’s oceans. Using the advice to “get areas and volumes from lengths,” we can estimate the volume of the oceans as surface area times average depth, or $V = AD$. We know the diameter of Earth from [\[link\]](#) and we know that most of Earth’s surface is covered in water, so we can estimate the surface area of the oceans as being roughly equal to the surface area of the planet. By following the advice to “get areas and volumes from lengths” again, we can approximate Earth as a sphere and use the formula for the surface area of a sphere of diameter d —that is, $A = \pi d^2$, to estimate the surface area of the oceans. Now we just need to estimate the average depth of the oceans. For this, we use the advice: “If all else fails, bound it.” We happen to know the deepest points in the ocean are around 10 km and that it is not uncommon for the ocean to be deeper than 1 km, so we take the average depth to be around $(10^3 \times 10^4)^{0.5} \approx 3 \times 10^3 \text{ m}$. Now we just need to put it all together, heeding the advice that “one ‘sig. fig.’ is fine.”

Solution

We estimate the surface area of Earth (and hence the surface area of Earth’s oceans) to be roughly

Equation:

$$A = \pi d^2 = \pi(10^7 \text{ m})^2 \approx 3 \times 10^{14} \text{ m}^2.$$

Next, using our average depth estimate of $D = 3 \times 10^3 \text{ m}$, which was obtained by bounding, we estimate the volume of Earth’s oceans to be

Equation:

$$V = AD = (3 \times 10^{14} \text{ m}^2)(3 \times 10^3 \text{ m}) = 9 \times 10^{17} \text{ m}^3.$$

Last, we estimate the mass of the world’s oceans to be

Equation:

$$M = \rho V = (10^3 \text{ kg/m}^3)(9 \times 10^{17} \text{ m}^3) = 9 \times 10^{20} \text{ kg}.$$

Thus, we estimate that the order of magnitude of the mass of the planet's oceans is 10^{21} kg.

Significance

To verify our answer to the best of our ability, we first need to answer the question: Does this make any sense? From [\[link\]](#), we see the mass of Earth's atmosphere is on the order of 10^{19} kg and the mass of Earth is on the order of 10^{25} kg. It is reassuring that our estimate of 10^{21} kg for the mass of Earth's oceans falls somewhere between these two. So, yes, it does seem to make sense. It just so happens that we did a search on the Web for "mass of oceans" and the top search results all said 1.4×10^{21} kg, which is the same order of magnitude as our estimate. Now, rather than having to trust blindly whoever first put that number up on a website (most of the other sites probably just copied it from them, after all), we can have a little more confidence in it.

Note:

Exercise:

Problem:

Check Your Understanding [\[link\]](#) says the mass of the atmosphere is 10^{19} kg. Assuming the density of the atmosphere is 1 kg/m^3 , estimate the height of Earth's atmosphere. Do you think your answer is an underestimate or an overestimate? Explain why.

Solution:

$3 \times 10^4 \text{ m}$ or 30 km. It is probably an underestimate because the density of the atmosphere decreases with altitude. (In fact, 30 km does not even get us out of the stratosphere.)

How many piano tuners are there in New York City? How many leaves are on that tree? If you are studying photosynthesis or thinking of writing a smartphone app for piano tuners, then the answers to these questions might be of great interest to you. Otherwise, you probably couldn't care less what the answers are. However, these are exactly the sorts of estimation problems that people in various tech industries have been asking potential employees to evaluate their quantitative reasoning skills. If building physical intuition and evaluating quantitative claims do not seem like sufficient reasons for you to practice estimation problems, how about the fact that being good at them just might land you a high-paying job?

Note:

For practice estimating relative lengths, areas, and volumes, check out this [PhET](#) simulation, titled “Estimation.”

Summary

- An estimate is a rough educated guess at the value of a physical quantity based on prior experience and sound physical reasoning. Some strategies that may help when making an estimate are as follows:
 - Get big lengths from smaller lengths.
 - Get areas and volumes from lengths.
 - Get masses from volumes and densities.
 - If all else fails, bound it.
 - One “sig. fig.” is fine.
 - Ask yourself: Does this make any sense?

Problems

Exercise:

Problem:

Assuming the human body is made primarily of water, estimate the volume of a person.

Exercise:**Problem:**

Assuming the human body is primarily made of water, estimate the number of molecules in it. (Note that water has a molecular mass of 18 g/mol and there are roughly 10^{24} atoms in a mole.)

Solution:

10^{28} atoms

Exercise:

Problem: Estimate the mass of air in a classroom.

Exercise:**Problem:**

Estimate the number of molecules that make up Earth, assuming an average molecular mass of 30 g/mol. (Note there are on the order of 10^{24} objects per mole.)

Solution:

10^{51} molecules

Exercise:

Problem: Estimate the surface area of a person.

Exercise:

Problem:

Roughly how many solar systems would it take to tile the disk of the Milky Way?

Solution:

10^{16} solar systems

Exercise:**Problem:**

(a) Estimate the density of the Moon. (b) Estimate the diameter of the Moon. (c) Given that the Moon subtends at an angle of about half a degree in the sky, estimate its distance from Earth.

Exercise:**Problem:**

The average density of the Sun is on the order 10^3 kg/m^3 . (a) Estimate the diameter of the Sun. (b) Given that the Sun subtends at an angle of about half a degree in the sky, estimate its distance from Earth.

Solution:

a. Volume = 10^{27} m^3 , diameter is 10^9 m .; b. 10^{11} m

Exercise:

Problem: Estimate the mass of a virus.

Exercise:

Problem:

A floating-point operation is a single arithmetic operation such as addition, subtraction, multiplication, or division. (a) Estimate the maximum number of floating-point operations a human being could possibly perform in a lifetime. (b) How long would it take a supercomputer to perform that many floating-point operations?

Solution:

a. A reasonable estimate might be one operation per second for a total of 10^9 in a lifetime.; b. about $(10^9)(10^{-17} \text{ s}) = 10^{-8} \text{ s}$, or about 10 ns

Glossary

estimation

using prior experience and sound physical reasoning to arrive at a rough idea of a quantity's value; sometimes called an "order-of-magnitude approximation," a "guesstimate," a "back-of-the-envelope calculation", or a "Fermi calculation"

Significant Figures

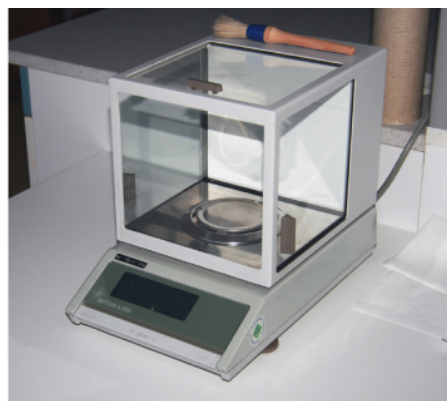
By the end of this section, you will be able to:

- Determine the correct number of significant figures for the result of a computation.
- Describe the relationship between the concepts of accuracy, precision, uncertainty, and discrepancy.
- Calculate the percent uncertainty of a measurement, given its value and its uncertainty.
- Determine the uncertainty of the result of a computation involving quantities with given uncertainties.

[\[link\]](#) shows two instruments used to measure the mass of an object. The digital scale has mostly replaced the double-pan balance in physics labs because it gives more accurate and precise measurements. But what exactly do we mean by *accurate* and *precise*? Aren't they the same thing? In this section we examine in detail the process of making and reporting a measurement.



(a)



(b)

(a) A double-pan mechanical balance is used to compare different masses. Usually an object with unknown mass is placed in one pan and objects of known mass are placed in the other pan. When the bar that connects the two pans is horizontal, then the masses in both pans are equal. The “known masses” are typically metal cylinders of standard mass such as 1 g, 10 g, and 100 g. (b) Many mechanical balances, such as double-pan balances, have been replaced by digital scales, which can typically measure the mass of an object more precisely. A mechanical

balance may read only the mass of an object to the nearest tenth of a gram, but many digital scales can measure the mass of an object up to the nearest thousandth of a gram. (credit a: modification of work by Serge Melki; credit b: modification of work by Karel Jakubec)

Accuracy and Precision of a Measurement

Science is based on observation and experiment—that is, on measurements. **Accuracy** is how close a measurement is to the accepted reference value for that measurement. For example, let's say we want to measure the length of standard printer paper. The packaging in which we purchased the paper states that it is 11.0 in. long. We then measure the length of the paper three times and obtain the following measurements: 11.1 in., 11.2 in., and 10.9 in. These measurements are quite accurate because they are very close to the reference value of 11.0 in. In contrast, if we had obtained a measurement of 12 in., our measurement would not be very accurate. Notice that the concept of accuracy requires that an accepted reference value be given.

The **precision** of measurements refers to how close the agreement is between repeated independent measurements (which are repeated under the same conditions). Consider the example of the paper measurements. The precision of the measurements refers to the spread of the measured values. One way to analyze the precision of the measurements is to determine the range, or difference, between the lowest and the highest measured values. In this case, the lowest value was 10.9 in. and the highest value was 11.2 in. Thus, the measured values deviated from each other by, at most, 0.3 in. These measurements were relatively precise because they did not vary too much in value. However, if the measured values had been 10.9 in., 11.1 in., and 11.9 in., then the measurements would not be very precise because there would be significant variation from one measurement to another. Notice that the concept of precision depends only on the actual measurements acquired and does not depend on an accepted reference value.

The measurements in the paper example are both accurate and precise, but in some cases, measurements are accurate but not precise, or they are precise

but not accurate. Let's consider an example of a GPS attempting to locate the position of a restaurant in a city. Think of the restaurant location as existing at the center of a bull's-eye target and think of each GPS attempt to locate the restaurant as a black dot. In [\[link\]](#)(a), we see the GPS measurements are spread out far apart from each other, but they are all relatively close to the actual location of the restaurant at the center of the target. This indicates a low-precision, high-accuracy measuring system. However, in [\[link\]](#)(b), the GPS measurements are concentrated quite closely to one another, but they are far away from the target location. This indicates a high-precision, low-accuracy measuring system.



(a) High accuracy, low precision



(b) Low accuracy, high precision

A GPS attempts to locate a restaurant at the center of the bull's-eye. The black dots represent each attempt to pinpoint the location of the restaurant. (a) The dots are spread out quite far apart from one another, indicating low precision, but they are each rather close to the actual location of the restaurant, indicating high accuracy. (b) The dots are concentrated rather closely to one another, indicating high precision, but they are rather far away from the actual location of the restaurant, indicating low accuracy. (credit a and credit b: modification of works by "DarkEvil"/Wikimedia Commons)

Accuracy, Precision, Uncertainty, and Discrepancy

The precision of a measuring system is related to the **uncertainty** in the measurements whereas the accuracy is related to the **discrepancy** from the accepted reference value. Uncertainty is a quantitative measure of how much your measured values deviate from one another. There are many different methods of calculating uncertainty, each of which is appropriate to different situations. Some examples include taking the range (that is, the biggest less the smallest) or finding the standard deviation of the measurements.

Discrepancy (or “measurement error”) is the difference between the measured value and a given standard or expected value. If the measurements are not very precise, then the uncertainty of the values is high. If the measurements are not very accurate, then the discrepancy of the values is high.

Recall our example of measuring paper length; we obtained measurements of 11.1 in., 11.2 in., and 10.9 in., and the accepted value was 11.0 in. We might average the three measurements to say our best guess is 11.1 in.; in this case, our discrepancy is $11.1 - 11.0 = 0.1$ in., which provides a quantitative measure of accuracy. We might calculate the uncertainty in our best guess by using half of the range of our measured values: 0.15 in. Then we would say the length of the paper is 11.1 in. plus or minus 0.15 in. The uncertainty in a measurement, A , is often denoted as δA (read “delta A ”), so the measurement result would be recorded as $A \pm \delta A$. Returning to our paper example, the measured length of the paper could be expressed as 11.1 ± 0.15 in. Since the discrepancy of 0.1 in. is less than the uncertainty of 0.15 in., we might say the measured value agrees with the accepted reference value to within experimental uncertainty.

Some factors that contribute to uncertainty in a measurement include the following:

- Limitations of the measuring device
- The skill of the person taking the measurement
- Irregularities in the object being measured
- Any other factors that affect the outcome (highly dependent on the situation)

In our example, such factors contributing to the uncertainty could be the smallest division on the ruler is 1/16 in., the person using the ruler has bad eyesight, the ruler is worn down on one end, or one side of the paper is slightly longer than the other. At any rate, the uncertainty in a measurement must be calculated to quantify its precision. If a reference value is known, it makes sense to calculate the discrepancy as well to quantify its accuracy.

Percent uncertainty

Another method of expressing uncertainty is as a percent of the measured value. If a measurement A is expressed with uncertainty δA , the **percent uncertainty** is defined as

Equation:

$$\text{Percent uncertainty} = \frac{\delta A}{A} \times 100\%.$$

Example:

Calculating Percent Uncertainty: A Bag of Apples

A grocery store sells 5-lb bags of apples. Let's say we purchase four bags during the course of a month and weigh the bags each time. We obtain the following measurements:

- Week 1 weight: 4.8 lb
- Week 2 weight: 5.3 lb
- Week 3 weight: 4.9 lb
- Week 4 weight: 5.4 lb

We then determine the average weight of the 5-lb bag of apples is 5.1 ± 0.3 lb from using half of the range. What is the percent uncertainty of the bag's weight?

Strategy

First, observe that the average value of the bag's weight, A , is 5.1 lb. The uncertainty in this value, δA , is 0.3 lb. We can use the following equation to

determine the percent uncertainty of the weight:

Note:

Equation:

$$\text{Percent uncertainty} = \frac{\delta A}{A} \times 100\%.$$

Solution

Substitute the values into the equation:

Equation:

$$\text{Percent uncertainty} = \frac{\delta A}{A} \times 100\% = \frac{0.3 \text{ lb}}{5.1 \text{ lb}} \times 100\% = 5.9\% \approx 6\%.$$

Significance

We can conclude the average weight of a bag of apples from this store is $5.1 \text{ lb} \pm 6\%$. Notice the percent uncertainty is dimensionless because the units of weight in $\delta A = 0.2 \text{ lb}$ canceled those in $A = 5.1 \text{ lb}$ when we took the ratio.

Note:

Exercise:

Problem:

Check Your Understanding A high school track coach has just purchased a new stopwatch. The stopwatch manual states the stopwatch has an uncertainty of $\pm 0.05 \text{ s}$. Runners on the track coach's team regularly clock 100-m sprints of 11.49 s to 15.01 s. At the school's last track meet, the first-place sprinter came in at 12.04 s and the second-place sprinter came in at 12.07 s. Will the coach's new stopwatch be helpful in timing the sprint team? Why or why not?

Solution:

No, the coach's new stopwatch will not be helpful. The uncertainty in the stopwatch is too great to differentiate between the sprint times effectively.

Uncertainties in calculations

Uncertainty exists in anything calculated from measured quantities. For example, the area of a floor calculated from measurements of its length and width has an uncertainty because the length and width have uncertainties. How big is the uncertainty in something you calculate by multiplication or division? If the measurements going into the calculation have small uncertainties (a few percent or less), then the **method of adding percents** can be used for multiplication or division. This method states *the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation*. For example, if a floor has a length of 4.00 m and a width of 3.00 m, with uncertainties of 2% and 1%, respectively, then the area of the floor is 12.0 m² and has an uncertainty of 3%. (Expressed as an area, this is 0.36 m² [12.0 m² × 0.03], which we round to 0.4 m² since the area of the floor is given to a tenth of a square meter.)

Precision of Measuring Tools and Significant Figures

An important factor in the precision of measurements involves the precision of the measuring tool. In general, a precise measuring tool is one that can measure values in very small increments. For example, a standard ruler can measure length to the nearest millimeter whereas a caliper can measure length to the nearest 0.01 mm. The caliper is a more precise measuring tool because it can measure extremely small differences in length. The more precise the measuring tool, the more precise the measurements.

When we express measured values, we can only list as many digits as we measured initially with our measuring tool. For example, if we use a standard ruler to measure the length of a stick, we may measure it to be 36.7 cm. We can't express this value as 36.71 cm because our measuring tool is not precise enough to measure a hundredth of a centimeter. It should be noted that the last digit in a measured value has been estimated in some way by the person performing the measurement. For example, the person measuring the length of a stick with a ruler notices the stick length seems to be somewhere in between 36.6 cm and 36.7 cm, and he or she must estimate the value of the last digit. Using the method of **significant figures**, the rule is that *the last digit written down in a measurement is the first digit with some uncertainty*. To determine the number of significant digits in a value, start with the first measured value at the left and count the number of digits through the last digit written on the right. For example, the measured value 36.7 cm has three digits, or three significant figures. Significant figures indicate the precision of the measuring tool used to measure a value.

Zeros

Special consideration is given to zeros when counting significant figures. The zeros in 0.053 are not significant because they are placeholders that locate the decimal point. There are two significant figures in 0.053. The zeros in 10.053 are not placeholders; they are significant. This number has five significant figures. The zeros in 1300 may or may not be significant, depending on the style of writing numbers. They could mean the number is known to the last digit or they could be placeholders. So 1300 could have two, three, or four significant figures. To avoid this ambiguity, we should write 1300 in scientific notation as 1.3×10^3 , 1.30×10^3 , or 1.300×10^3 , depending on whether it has two, three, or four significant figures. *Zeros are significant except when they serve only as placeholders.*

Significant figures in calculations

When combining measurements with different degrees of precision, *the number of significant digits in the final answer can be no greater than the*

number of significant digits in the least-precise measured value. There are two different rules, one for multiplication and division and the other for addition and subtraction.

1. *For multiplication and division, the result should have the same number of significant figures as the quantity with the least number of significant figures entering into the calculation.* For example, the area of a circle can be calculated from its radius using $A = \pi r^2$. Let's see how many significant figures the area has if the radius has only two—say, $r = 1.2$ m. Using a calculator with an eight-digit output, we would calculate
Equation:

$$A = \pi r^2 = (3.1415927\dots) \times (1.2 \text{ m})^2 = 4.5238934 \text{ m}^2.$$

But because the radius has only two significant figures, it limits the calculated quantity to two significant figures, or

Equation:

$$A = 4.5 \text{ m}^2,$$

although π is good to at least eight digits.

2. *For addition and subtraction, the answer can contain no more decimal places than the least-precise measurement.* Suppose we buy 7.56 kg of potatoes in a grocery store as measured with a scale with precision 0.01 kg, then we drop off 6.052 kg of potatoes at your laboratory as measured by a scale with precision 0.001 kg. Then, we go home and add 13.7 kg of potatoes as measured by a bathroom scale with precision 0.1 kg. How many kilograms of potatoes do we now have and how many significant figures are appropriate in the answer? The mass is found by simple addition and subtraction:

Equation:

$$\begin{array}{r} 7.56 \text{ kg} \\ -6.052 \text{ kg} \end{array}$$

$$\frac{+13.7 \text{ kg}}{15.208 \text{ kg}} = 15.2 \text{ kg}.$$

Next, we identify the least-precise measurement: 13.7 kg. This measurement is expressed to the 0.1 decimal place, so our final answer must also be expressed to the 0.1 decimal place. Thus, the answer is rounded to the tenths place, giving us 15.2 kg.

Significant figures in this text

In this text, most numbers are assumed to have three significant figures. Furthermore, consistent numbers of significant figures are used in all worked examples. An answer given to three digits is based on input good to at least three digits, for example. If the input has fewer significant figures, the answer will also have fewer significant figures. Care is also taken that the number of significant figures is reasonable for the situation posed. In some topics, particularly in optics, more accurate numbers are needed and we use more than three significant figures. Finally, if a number is *exact*, such as the two in the formula for the circumference of a circle, $C = 2\pi r$, it does not affect the number of significant figures in a calculation. Likewise, conversion factors such as 100 cm/1 m are considered exact and do not affect the number of significant figures in a calculation.

Summary

- Accuracy of a measured value refers to how close a measurement is to an accepted reference value. The discrepancy in a measurement is the amount by which the measurement result differs from this value.
- Precision of measured values refers to how close the agreement is between repeated measurements. The uncertainty of a measurement is a quantification of this.
- The precision of a measuring tool is related to the size of its measurement increments. The smaller the measurement increment, the more precise the tool.
- Significant figures express the precision of a measuring tool.
- When multiplying or dividing measured values, the final answer can contain only as many significant figures as the value with the least number of significant figures.

- When adding or subtracting measured values, the final answer cannot contain more decimal places than the least-precise value.

Key Equations

Percent uncertainty

$$\text{Percent uncertainty} = \frac{\delta A}{A} \times 100\%$$

Conceptual Questions

Exercise:

Problem:

(a) What is the relationship between the precision and the uncertainty of a measurement? (b) What is the relationship between the accuracy and the discrepancy of a measurement?

Solution:

a. Uncertainty is a quantitative measure of precision. b. Discrepancy is a quantitative measure of accuracy.

Problems

Exercise:

Problem:

Consider the equation $4000/400 = 10.0$. Assuming the number of significant figures in the answer is correct, what can you say about the number of significant figures in 4000 and 400?

Exercise:**Problem:**

Suppose your bathroom scale reads your mass as 65 kg with a 3% uncertainty. What is the uncertainty in your mass (in kilograms)?

Solution:

2 kg

Exercise:**Problem:**

A good-quality measuring tape can be off by 0.50 cm over a distance of 20 m. What is its percent uncertainty?

Exercise:**Problem:**

An infant's pulse rate is measured to be 130 ± 5 beats/min. What is the percent uncertainty in this measurement?

Solution:

4%

Exercise:**Problem:**

(a) Suppose that a person has an average heart rate of 72.0 beats/min. How many beats does he or she have in 2.0 years? (b) In 2.00 years? (c) In 2.000 years?

Exercise:**Problem:**

A can contains 375 mL of soda. How much is left after 308 mL is removed?

Solution:

67 mL

Exercise:**Problem:**

State how many significant figures are proper in the results of the following calculations: (a) $(106.7)(98.2)/(46.210)(1.01)$; (b) $(18.7)^2$; (c) $(1.60 \times 10^{-19})(3712)$

Exercise:**Problem:**

(a) How many significant figures are in the numbers 99 and 100.? (b) If the uncertainty in each number is 1, what is the percent uncertainty in each? (c) Which is a more meaningful way to express the accuracy of these two numbers: significant figures or percent uncertainties?

Solution:

a. The number 99 has 2 significant figures; 100. has 3 significant figures. b. 1.00%; c. percent uncertainties

Exercise:**Problem:**

(a) If your speedometer has an uncertainty of 2.0 km/h at a speed of 90 km/h, what is the percent uncertainty? (b) If it has the same percent uncertainty when it reads 60 km/h, what is the range of speeds you could be going?

Exercise:**Problem:**

(a) A person's blood pressure is measured to be 120 ± 2 mm Hg. What is its percent uncertainty? (b) Assuming the same percent uncertainty, what is the uncertainty in a blood pressure measurement of 80 mm Hg?

Solution:

a. 2%; b. 1 mm Hg

Exercise:**Problem:**

A person measures his or her heart rate by counting the number of beats in 30 s. If 40 ± 1 beats are counted in 30.0 ± 0.5 s, what is the heart rate and its uncertainty in beats per minute?

Exercise:

Problem: What is the area of a circle 3.102 cm in diameter?

Solution:

7.557 cm²

Exercise:**Problem:**

Determine the number of significant figures in the following measurements: (a) 0.0009, (b) 15,450.0, (c) 6×10^3 , (d) 87.990, and (e) 30.42.

Exercise:**Problem:**

Perform the following calculations and express your answer using the correct number of significant digits. (a) A woman has two bags weighing 13.5 lb and one bag with a weight of 10.2 lb. What is the total weight of the bags? (b) The force F on an object is equal to its mass m multiplied by its acceleration a . If a wagon with mass 55 kg accelerates at a rate of 0.0255 m/s^2 , what is the force on the wagon? (The unit of force is called the *newton* and it is expressed with the symbol N.)

Solution:

a. 37.2 lb; because the number of bags is an exact value, it is not considered in the significant figures; b. 1.4 N; because the value 55 kg has only two significant figures, the final value must also contain two significant figures

Glossary

accuracy

the degree to which a measured value agrees with an accepted reference value for that measurement

discrepancy

the difference between the measured value and a given standard or expected value

method of adding percents

the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation.

percent uncertainty

the ratio of the uncertainty of a measurement to the measured value, expressed as a percentage

precision

the degree to which repeated measurements agree with each other

significant figures

used to express the precision of a measuring tool used to measure a value

uncertainty

a quantitative measure of how much measured values deviate from one another

Solving Problems in Physics

By the end of this section, you will be able to:

- Describe the process for developing a problem-solving strategy.
- Explain how to find the numerical solution to a problem.
- Summarize the process for assessing the significance of the numerical solution to a problem.



Problem-solving skills are essential to your success in physics. (credit: “scui3asteveo”/Flickr)

Problem-solving skills are clearly essential to success in a quantitative course in physics. More important, the ability to apply broad physical principles—usually represented by equations—to specific situations is a very powerful form of knowledge. It is much more powerful than memorizing a list of facts. Analytical skills and problem-solving abilities

can be applied to new situations whereas a list of facts cannot be made long enough to contain every possible circumstance. Such analytical skills are useful both for solving problems in this text and for applying physics in everyday life.

As you are probably well aware, a certain amount of creativity and insight is required to solve problems. No rigid procedure works every time. Creativity and insight grow with experience. With practice, the basics of problem solving become almost automatic. One way to get practice is to work out the text's examples for yourself as you read. Another is to work as many end-of-section problems as possible, starting with the easiest to build confidence and then progressing to the more difficult. After you become involved in physics, you will see it all around you, and you can begin to apply it to situations you encounter outside the classroom, just as is done in many of the applications in this text.

Although there is no simple step-by-step method that works for every problem, the following three-stage process facilitates problem solving and makes it more meaningful. The three stages are strategy, solution, and significance. This process is used in examples throughout the book. Here, we look at each stage of the process in turn.

Strategy

Strategy is the beginning stage of solving a problem. The idea is to figure out exactly what the problem is and then develop a strategy for solving it. Some general advice for this stage is as follows:

- *Examine the situation to determine which physical principles are involved.* It often helps to *draw a simple sketch* at the outset. You often need to decide which direction is positive and note that on your sketch. When you have identified the physical principles, it is much easier to find and apply the equations representing those principles. Although finding the correct equation is essential, keep in mind that equations represent physical principles, laws of nature, and relationships among physical quantities. Without a conceptual understanding of a problem, a numerical solution is meaningless.

- *Make a list of what is given or can be inferred from the problem as stated (identify the “knowns”).* Many problems are stated very succinctly and require some inspection to determine what is known. Drawing a sketch can be very useful at this point as well. Formally identifying the knowns is of particular importance in applying physics to real-world situations. For example, the word *stopped* means the velocity is zero at that instant. Also, we can often take initial time and position as zero by the appropriate choice of coordinate system.
- *Identify exactly what needs to be determined in the problem (identify the unknowns).* In complex problems, especially, it is not always obvious what needs to be found or in what sequence. Making a list can help identify the unknowns.
- *Determine which physical principles can help you solve the problem.* Since physical principles tend to be expressed in the form of mathematical equations, a list of knowns and unknowns can help here. It is easiest if you can find equations that contain only one unknown—that is, all the other variables are known—so you can solve for the unknown easily. If the equation contains more than one unknown, then additional equations are needed to solve the problem. In some problems, several unknowns must be determined to get at the one needed most. In such problems it is especially important to keep physical principles in mind to avoid going astray in a sea of equations. You may have to use two (or more) different equations to get the final answer.

Solution

The solution stage is when you do the math. *Substitute the knowns (along with their units) into the appropriate equation and obtain numerical solutions complete with units.* That is, do the algebra, calculus, geometry, or arithmetic necessary to find the unknown from the knowns, being sure to carry the units through the calculations. This step is clearly important because it produces the numerical answer, along with its units. Notice, however, that this stage is only one-third of the overall problem-solving process.

Significance

After having done the math in the solution stage of problem solving, it is tempting to think you are done. But, always remember that physics is not math. Rather, in doing physics, we use mathematics as a tool to help us understand nature. So, after you obtain a numerical answer, you should always assess its significance:

- *Check your units.* If the units of the answer are incorrect, then an error has been made and you should go back over your previous steps to find it. One way to find the mistake is to check all the equations you derived for dimensional consistency. However, be warned that correct units do not guarantee the numerical part of the answer is also correct.
- *Check the answer to see whether it is reasonable. Does it make sense?* This step is extremely important: –the goal of physics is to describe nature accurately. To determine whether the answer is reasonable, check both its magnitude and its sign, in addition to its units. The magnitude should be consistent with a rough estimate of what it should be. It should also compare reasonably with magnitudes of other quantities of the same type. The sign usually tells you about direction and should be consistent with your prior expectations. Your judgment will improve as you solve more physics problems, and it will become possible for you to make finer judgments regarding whether nature is described adequately by the answer to a problem. This step brings the problem back to its conceptual meaning. If you can judge whether the answer is reasonable, you have a deeper understanding of physics than just being able to solve a problem mechanically.
- *Check to see whether the answer tells you something interesting. What does it mean?* This is the flip side of the question: Does it make sense? Ultimately, physics is about understanding nature, and we solve physics problems to learn a little something about how nature operates. Therefore, assuming the answer does make sense, you should always take a moment to see if it tells you something about the world that you find interesting. Even if the answer to this particular problem is not very interesting to you, what about the method you used to solve it? Could the method be adapted to answer a question that you do find

interesting? In many ways, it is in answering questions such as these that science progresses.

Summary

The three stages of the process for solving physics problems used in this book are as follows:

- *Strategy*: Determine which physical principles are involved and develop a strategy for using them to solve the problem.
- *Solution*: Do the math necessary to obtain a numerical solution complete with units.
- *Significance*: Check the solution to make sure it makes sense (correct units, reasonable magnitude and sign) and assess its significance.

Conceptual Questions

Exercise:

Problem:

What information do you need to choose which equation or equations to use to solve a problem?

Exercise:

Problem:

What should you do after obtaining a numerical answer when solving a problem?

Solution:

Check to make sure it makes sense and assess its significance.

Additional Problems

Exercise:

Problem:

Consider the equation $y = mt + b$, where the dimension of y is length and the dimension of t is time, and m and b are constants. What are the dimensions and SI units of (a) m and (b) b ?

Exercise:**Problem:**

Consider the equation $s = s_0 + v_0t + a_0t^2/2 + j_0t^3/6 + S_0t^4/24 + ct^5/120$, where s is a length and t is a time. What are the dimensions and SI units of (a) s_0 , (b) v_0 , (c) a_0 , (d) j_0 , (e) S_0 , and (f) c ?

Solution:

a. $[s_0] = L$ and units are meters (m); b. $[v_0] = LT^{-1}$ and units are meters per second (m/s); c. $[a_0] = LT^{-2}$ and units are meters per second squared (m/s²); d. $[j_0] = LT^{-3}$ and units are meters per second cubed (m/s³); e. $[S_0] = LT^{-4}$ and units are m/s⁴; f. $[c] = LT^{-5}$ and units are m/s⁵.

Exercise:**Problem:**

(a) A car speedometer has a 5% uncertainty. What is the range of possible speeds when it reads 90 km/h? (b) Convert this range to miles per hour. Note 1 km = 0.6214 mi.

Exercise:

Problem:

A marathon runner completes a 42.188-km course in 2 h, 30 min, and 12 s. There is an uncertainty of 25 m in the distance traveled and an uncertainty of 1 s in the elapsed time. (a) Calculate the percent uncertainty in the distance. (b) Calculate the percent uncertainty in the elapsed time. (c) What is the average speed in meters per second? (d) What is the uncertainty in the average speed?

Solution:

a. 0.059%; b. 0.01%; c. 4.681 m/s; d. 0.07%, 0.003 m/s

Exercise:**Problem:**

The sides of a small rectangular box are measured to be 1.80 ± 0.1 cm, 2.05 ± 0.02 cm, and 3.1 ± 0.1 cm long. Calculate its volume and uncertainty in cubic centimeters.

Exercise:**Problem:**

When nonmetric units were used in the United Kingdom, a unit of mass called the pound-mass (lbm) was used, where $1 \text{ lbm} = 0.4539 \text{ kg}$. (a) If there is an uncertainty of 0.0001 kg in the pound-mass unit, what is its percent uncertainty? (b) Based on that percent uncertainty, what mass in pound-mass has an uncertainty of 1 kg when converted to kilograms?

Solution:

a. 0.02%; b. 1×10^4 lbm

Exercise:

Problem:

The length and width of a rectangular room are measured to be 3.955 ± 0.005 m and 3.050 ± 0.005 m. Calculate the area of the room and its uncertainty in square meters.

Exercise:**Problem:**

A car engine moves a piston with a circular cross-section of 7.500 ± 0.002 cm in diameter a distance of 3.250 ± 0.001 cm to compress the gas in the cylinder. (a) By what amount is the gas decreased in volume in cubic centimeters? (b) Find the uncertainty in this volume.

Solution:

a. 143.6 cm^3 ; b. 0.1 cm^3 or 0.084%

Challenge Problems**Exercise:**

Problem:

The first atomic bomb was detonated on July 16, 1945, at the Trinity test site about 200 mi south of Los Alamos. In 1947, the U.S. government declassified a film reel of the explosion. From this film reel, British physicist G. I. Taylor was able to determine the rate at which the radius of the fireball from the blast grew. Using dimensional analysis, he was then able to deduce the amount of energy released in the explosion, which was a closely guarded secret at the time. Because of this, Taylor did not publish his results until 1950. This problem challenges you to recreate this famous calculation. (a) Using keen physical insight developed from years of experience, Taylor decided the radius r of the fireball should depend only on time since the explosion, t , the density of the air, ρ , and the energy of the initial explosion, E . Thus, he made the educated guess that $r = kE^a\rho^bt^c$ for some dimensionless constant k and some unknown exponents a , b , and c . Given that $[E] = \text{ML}^2\text{T}^{-2}$, determine the values of the exponents necessary to make this equation dimensionally consistent. (*Hint:* Notice the equation implies that $k = rE^{-a}\rho^{-b}t^{-c}$ and that $[k] = 1$.) (b) By analyzing data from high-energy conventional explosives, Taylor found the formula he derived seemed to be valid as long as the constant k had the value 1.03. From the film reel, he was able to determine many values of r and the corresponding values of t . For example, he found that after 25.0 ms, the fireball had a radius of 130.0 m. Use these values, along with an average air density of 1.25 kg/m^3 , to calculate the initial energy release of the Trinity detonation in joules (J). (*Hint:* To get energy in joules, you need to make sure all the numbers you substitute in are expressed in terms of SI base units.) (c) The energy released in large explosions is often cited in units of “tons of TNT” (abbreviated “t TNT”), where 1 t TNT is about 4.2 GJ. Convert your answer to (b) into kilotons of TNT (that is, kt TNT). Compare your answer with the quick-and-dirty estimate of 10 kt TNT made by physicist Enrico Fermi shortly after witnessing the explosion from what was thought to be a safe distance. (Reportedly, Fermi made his estimate by dropping some shredded bits of paper right before the remnants of the shock wave hit him and looked to see how far they were carried by it.)

Exercise:**Problem:**

The purpose of this problem is to show the entire concept of dimensional consistency can be summarized by the old saying “You can’t add apples and oranges.” If you have studied power series expansions in a calculus course, you know the standard mathematical functions such as trigonometric functions, logarithms, and exponential functions can be expressed as infinite sums of the form

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots, \text{ where the } a_n \text{ are}$$

dimensionless constants for all $n = 0, 1, 2, \cdots$ and x is the argument of the function. (If you have not studied power series in calculus yet, just trust us.) Use this fact to explain why the requirement that all terms in an equation have the same dimensions is sufficient as a definition of dimensional consistency. That is, it actually implies the arguments of standard mathematical functions must be dimensionless, so it is not really necessary to make this latter condition a separate requirement of the definition of dimensional consistency as we have done in this section.

Solution:

Since each term in the power series involves the argument raised to a different power, the only way that every term in the power series can have the same dimension is if the argument is dimensionless. To see this explicitly, suppose $[x] = L^a M^b T^c$. Then, $[x^n] = [x]^n = L^{an} M^{bn} T^{cn}$. If we want $[x] = [x^n]$, then $an = a$, $bn = b$, and $cn = c$ for all n . The only way this can happen is if $a = b = c = 0$.

Introduction

class="introduction"

A signpost gives information about distances and directions to towns or to other locations relative to the location of the signpost.

Distance is a scalar quantity. Knowing the distance alone is not enough to get to the town; we must also know the direction from the signpost to the town. The direction, together with the distance, is a vector quantity commonly called the displacement vector.

A signpost, therefore, gives information about displacement vectors from the signpost to towns. (credit: modification of work by "studio tdes"/Flickr, thedailyenglishshow.com)



Vectors are essential to physics and engineering. Many fundamental physical quantities are vectors, including displacement, velocity, force, and electric and magnetic vector fields. Scalar products of vectors define other fundamental scalar physical quantities, such as energy. Vector products of vectors define still other fundamental vector physical quantities, such as torque and angular momentum. In other words, vectors are a component part of physics in much the same way as sentences are a component part of literature.

In introductory physics, vectors are Euclidean quantities that have geometric representations as arrows in one dimension (in a line), in two dimensions (in a plane), or in three dimensions (in space). They can be added, subtracted, or multiplied. In this chapter, we explore elements of vector algebra for applications in mechanics and in electricity and magnetism. Vector operations also have numerous generalizations in other branches of physics.

Scalars and Vectors

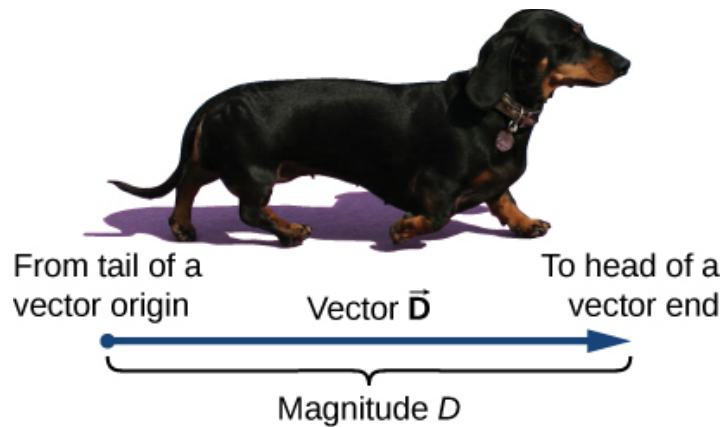
By the end of this section, you will be able to:

- Describe the difference between vector and scalar quantities.
- Identify the magnitude and direction of a vector.
- Explain the effect of multiplying a vector quantity by a scalar.
- Describe how one-dimensional vector quantities are added or subtracted.
- Explain the geometric construction for the addition or subtraction of vectors in a plane.
- Distinguish between a vector equation and a scalar equation.

Many familiar physical quantities can be specified completely by giving a single number and the appropriate unit. For example, “a class period lasts 50 min” or “the gas tank in my car holds 65 L” or “the distance between two posts is 100 m.” A physical quantity that can be specified completely in this manner is called a **scalar quantity**. Scalar is a synonym of “number.” Time, mass, distance, length, volume, temperature, and energy are examples of **scalar** quantities.

Scalar quantities that have the same physical units can be added or subtracted according to the usual rules of algebra for numbers. For example, a class ending 10 min earlier than 50 min lasts $50 \text{ min} - 10 \text{ min} = 40 \text{ min}$. Similarly, a 60-cal serving of corn followed by a 200-cal serving of donuts gives $60 \text{ cal} + 200 \text{ cal} = 260 \text{ cal}$ of energy. When we multiply a scalar quantity by a number, we obtain the same scalar quantity but with a larger (or smaller) value. For example, if yesterday’s breakfast had 200 cal of energy and today’s breakfast has four times as much energy as it had yesterday, then today’s breakfast has $4(200 \text{ cal}) = 800 \text{ cal}$ of energy. Two scalar quantities can also be multiplied or divided by each other to form a derived scalar quantity. For example, if a train covers a distance of 100 km in 1.0 h, its speed is $100.0 \text{ km}/1.0 \text{ h} = 27.8 \text{ m/s}$, where the speed is a derived scalar quantity obtained by dividing distance by time.

Many physical quantities, however, cannot be described completely by just a single number of physical units. For example, when the U.S. Coast Guard dispatches a ship or a helicopter for a rescue mission, the rescue team must know not only the distance to the distress signal, but also the direction from which the signal is coming so they can get to its origin as quickly as possible. Physical quantities specified completely by giving a number of units (magnitude) and a direction are called **vector quantities**. Examples of vector quantities include displacement, velocity, position, force, and torque. In the language of mathematics, physical vector quantities are represented by mathematical objects called **vectors** ([link](#)). We can add or subtract two vectors, and we can multiply a vector by a scalar or by another vector, but we cannot divide by a vector. The operation of division by a vector is not defined.



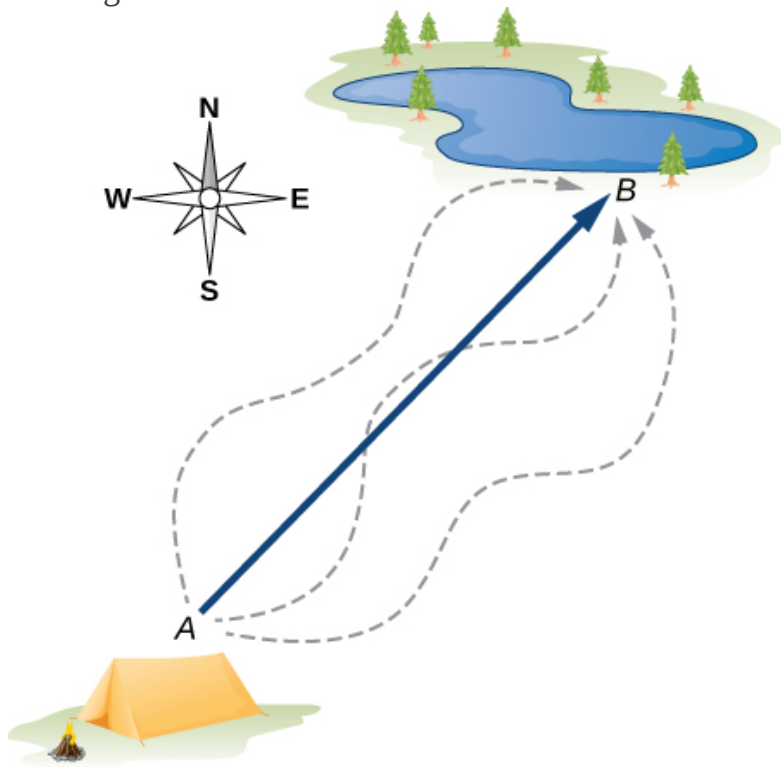
We draw a vector from the initial point or origin (called the “tail” of a vector) to the end or terminal point (called the “head” of a vector), marked by an arrowhead. Magnitude is the length of a vector and is always a positive scalar quantity. (credit "photo": modification of work by Cate Sevilla)

Let's examine vector algebra using a graphical method to be aware of basic terms and to develop a qualitative understanding. In practice, however, when it comes to solving physics problems, we use analytical methods, which we'll see in the next section. Analytical methods are more simple computationally and more accurate than graphical methods. From now on, to distinguish between a vector and a scalar quantity, we adopt the common convention that a letter in bold type with an arrow above it denotes a vector, and a letter without an arrow denotes a scalar. For example, a distance of 2.0 km, which is a scalar quantity, is denoted by $d = 2.0$ km, whereas a displacement of 2.0 km in some direction, which is a vector quantity, is denoted by \vec{d} .

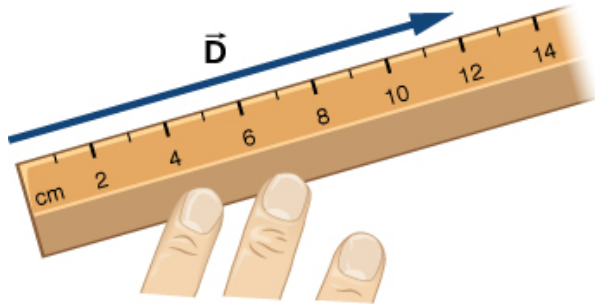
Suppose you tell a friend on a camping trip that you have discovered a terrific fishing hole 6 km from your tent. It is unlikely your friend would be able to find the hole easily unless you also communicate the direction in which it can be found with respect to your campsite. You may say, for example, “Walk about 6 km northeast from my tent.” The key concept here is that you have to give not one but *two* pieces of information—namely, the distance or magnitude (6 km) *and* the direction (northeast).

Displacement is a general term used to describe a *change in position*, such as during a trip from the tent to the fishing hole. Displacement is an example of a vector quantity. If you walk from the tent (location A) to the hole (location B), as shown in [\[link\]](#), the vector \vec{D} , representing your **displacement**, is drawn as the arrow that originates at point

A and ends at point B . The arrowhead marks the end of the vector. The direction of the displacement vector \vec{D} is the direction of the arrow. The length of the arrow represents the **magnitude** D of vector \vec{D} . Here, $D = 6$ km. Since the magnitude of a vector is its length, which is a positive number, the magnitude is also indicated by placing the absolute value notation around the symbol that denotes the vector; so, we can write equivalently that $D \equiv |\vec{D}|$. To solve a vector problem graphically, we need to draw the vector \vec{D} to scale. For example, if we assume 1 unit of distance (1 km) is represented in the drawing by a line segment of length $u = 2$ cm, then the total displacement in this example is represented by a vector of length $d = 6u = 6(2 \text{ cm}) = 12$ cm, as shown in [\[link\]](#). Notice that here, to avoid confusion, we used $D = 6$ km to denote the magnitude of the actual displacement and $d = 12$ cm to denote the length of its representation in the drawing.



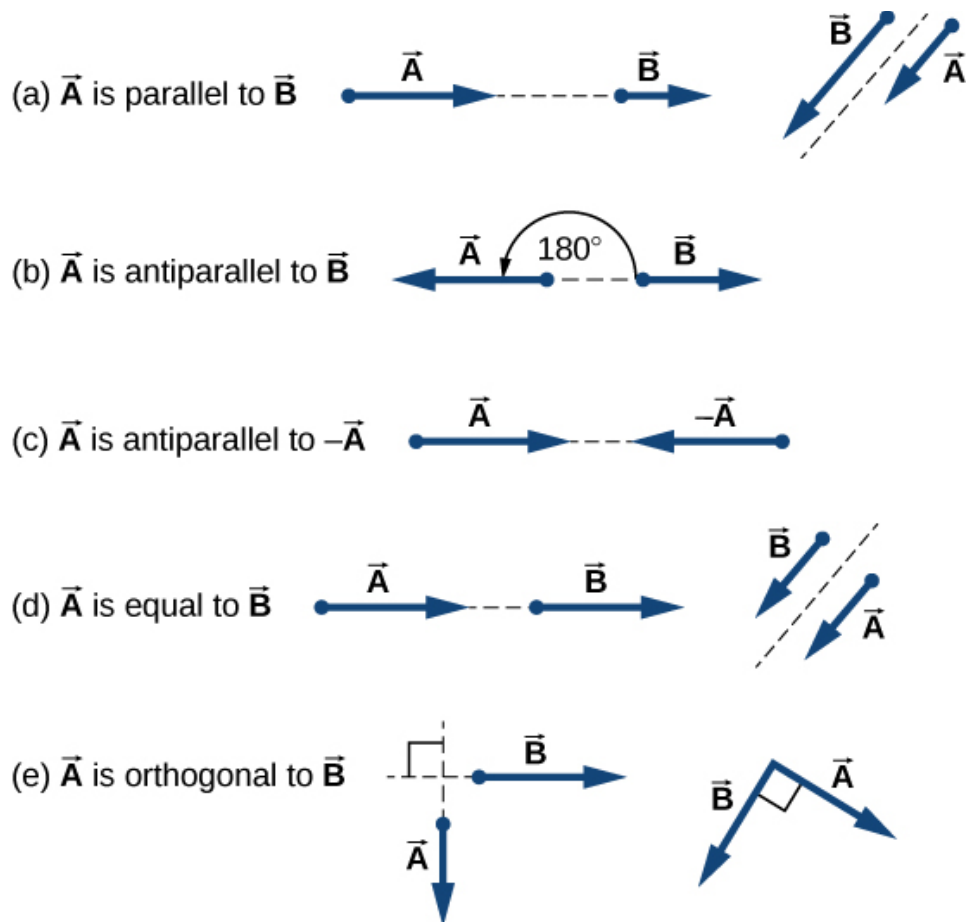
The displacement vector from point A (the initial position at the campsite) to point B (the final position at the fishing hole) is indicated by an arrow with origin at point A and end at point B . The displacement is the same for any of the actual paths (dashed curves) that may be taken between points A and B .



A displacement \vec{D} of magnitude 6 km is drawn to scale as a vector of length 12 cm when the length of 2 cm represents 1 unit of displacement (which in this case is 1 km).

Suppose your friend walks from the campsite at A to the fishing pond at B and then walks back: from the fishing pond at B to the campsite at A . The magnitude of the displacement vector \vec{D}_{AB} from A to B is the same as the magnitude of the displacement vector \vec{D}_{BA} from B to A (it equals 6 km in both cases), so we can write $D_{AB} = D_{BA}$. However, vector \vec{D}_{AB} is *not* equal to vector \vec{D}_{BA} because these two vectors have different directions: $\vec{D}_{AB} \neq \vec{D}_{BA}$. In [\[link\]](#), vector \vec{D}_{BA} would be represented by a vector with an origin at point B and an end at point A , indicating vector \vec{D}_{BA} points to the southwest, which is exactly 180° opposite to the direction of vector \vec{D}_{AB} . We say that vector \vec{D}_{BA} is **antiparallel** to vector \vec{D}_{AB} and write $\vec{D}_{AB} = -\vec{D}_{BA}$, where the minus sign indicates the antiparallel direction.

Two vectors that have identical directions are said to be **parallel vectors**—meaning, they are *parallel* to each other. Two parallel vectors \vec{A} and \vec{B} are equal, denoted by $\vec{A} = \vec{B}$, if and only if they have equal magnitudes $|\vec{A}| = |\vec{B}|$. Two vectors with directions perpendicular to each other are said to be **orthogonal vectors**. These relations between vectors are illustrated in [\[link\]](#).



Various relations between two vectors \vec{A} and \vec{B} . (a) $\vec{A} \neq \vec{B}$ because $A \neq B$. (b) $\vec{A} \neq \vec{B}$ because they are not parallel and $A \neq B$. (c) $\vec{A} \neq -\vec{A}$ because they have different directions (even though $|\vec{A}| = |-\vec{A}| = A$). (d) $\vec{A} = \vec{B}$ because they are parallel *and* have identical magnitudes $A = B$. (e) $\vec{A} \neq \vec{B}$ because they have different directions (are not parallel); here, their directions differ by 90° —meaning, they are orthogonal.

Note:

Exercise:

Problem:

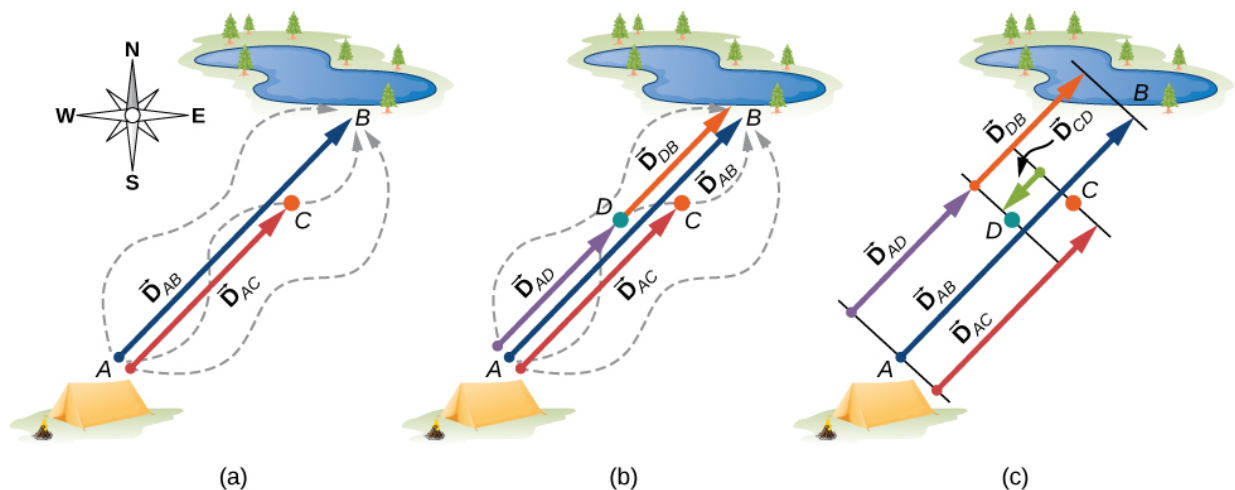
Check Your Understanding Two motorboats named *Alice* and *Bob* are moving on a lake. Given the information about their velocity vectors in each of the following situations, indicate whether their velocity vectors are equal or otherwise. (a) *Alice* moves north at 6 knots and *Bob* moves west at 6 knots. (b) *Alice* moves west at 6 knots and *Bob* moves west at 3 knots. (c) *Alice* moves northeast at 6 knots and *Bob* moves south at 3 knots. (d) *Alice* moves northeast at 6 knots and *Bob* moves southwest at 6 knots. (e) *Alice* moves northeast at 2 knots and *Bob* moves closer to the shore northeast at 2 knots.

Solution:

a. not equal because they are orthogonal; b. not equal because they have different magnitudes; c. not equal because they have different magnitudes and directions; d. not equal because they are antiparallel; e. equal.

Algebra of Vectors in One Dimension

Vectors can be multiplied by scalars, added to other vectors, or subtracted from other vectors. We can illustrate these vector concepts using an example of the fishing trip seen in [\[link\]](#).



Displacement vectors for a fishing trip. (a) Stopping to rest at point C while walking from camp (point A) to the pond (point B). (b) Going back for the dropped tackle box (point D). (c) Finishing up at the fishing pond.

Suppose your friend departs from point A (the campsite) and walks in the direction to point B (the fishing pond), but, along the way, stops to rest at some point C located three-quarters of the distance between A and B , beginning from point A ([link](#)(a)). What is his displacement vector \vec{D}_{AC} when he reaches point C ? We know that if he walks all the way to B , his displacement vector relative to A is \vec{D}_{AB} , which has magnitude $D_{AB} = 6$ km and a direction of northeast. If he walks only a 0.75 fraction of the total distance, maintaining the northeasterly direction, at point C he must be $0.75D_{AB} = 4.5$ km away from the campsite at A . So, his displacement vector at the rest point C has magnitude $D_{AC} = 4.5$ km $= 0.75D_{AB}$ and is parallel to the displacement vector \vec{D}_{AB} . All of this can be stated succinctly in the form of the following **vector equation**:

Equation:

$$\vec{D}_{AC} = 0.75\vec{D}_{AB}.$$

In a vector equation, both sides of the equation are vectors. The previous equation is an example of a vector multiplied by a positive scalar (number) $\alpha = 0.75$. The result, \vec{D}_{AC} , of such a multiplication is a new vector with a direction parallel to the direction of the original vector \vec{D}_{AB} .

In general, when a vector \vec{A} is multiplied by a *positive* scalar α , the result is a new vector \vec{B} that is *parallel* to \vec{A} :

Note:

Equation:

$$\vec{B} = \alpha\vec{A}.$$

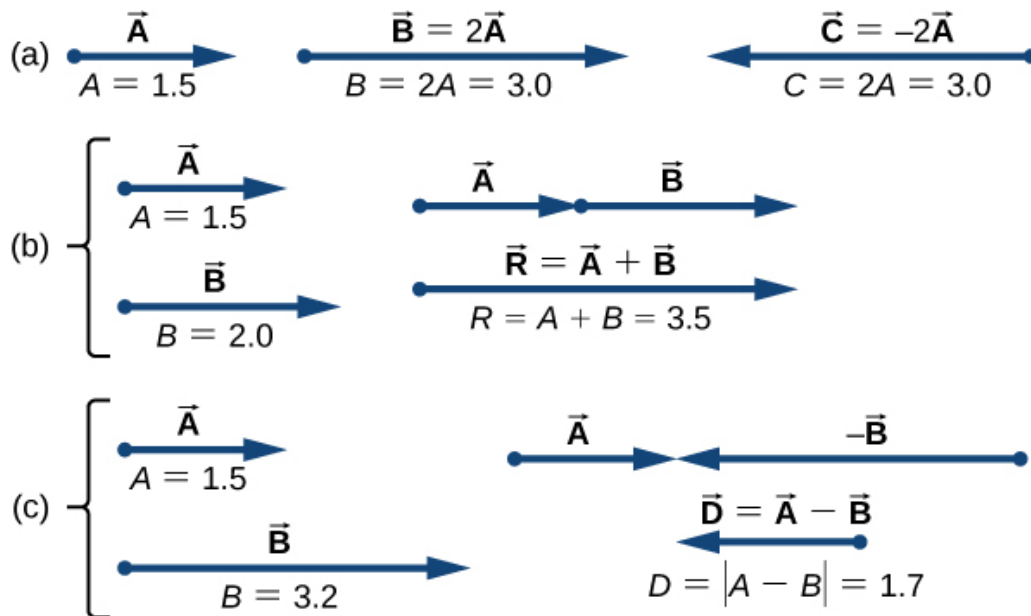
The magnitude $|\vec{B}|$ of this new vector is obtained by multiplying the magnitude $|\vec{A}|$ of the original vector, as expressed by the **scalar equation**:

Note:

Equation:

$$B = |\alpha|A.$$

In a scalar equation, both sides of the equation are numbers. [\[link\]](#) is a scalar equation because the magnitudes of vectors are scalar quantities (and positive numbers). If the scalar α is *negative* in the vector equation [\[link\]](#), then the magnitude $|\vec{B}|$ of the new vector is still given by [\[link\]](#), but the direction of the new vector \vec{B} is *antiparallel* to the direction of \vec{A} . These principles are illustrated in [\[link\]](#)(a) by two examples where the length of vector \vec{A} is 1.5 units. When $\alpha = 2$, the new vector $\vec{B} = 2\vec{A}$ has length $B = 2A = 3.0$ units (twice as long as the original vector) and is parallel to the original vector. When $\alpha = -2$, the new vector $\vec{C} = -2\vec{A}$ has length $C = |-2|A = 3.0$ units (twice as long as the original vector) and is antiparallel to the original vector.



Algebra of vectors in one dimension. (a) Multiplication by a scalar. (b) Addition of two vectors (\vec{R} is called the *resultant* of vectors \vec{A} and \vec{B}). (c) Subtraction of two vectors (\vec{D} is the difference of vectors \vec{A} and \vec{B}).

Now suppose your fishing buddy departs from point A (the campsite), walking in the direction to point B (the fishing hole), but he realizes he lost his tackle box when he stopped to rest at point C (located three-quarters of the distance between A and B , beginning from point A). So, he turns back and retraces his steps in the direction toward the campsite and finds the box lying on the path at some point D only 1.2 km away from point C (see [\[link\]](#)(b)). What is his displacement vector \vec{D}_{AD} when he finds the box at point D ? What is his displacement vector \vec{D}_{DB} from point D to the hole? We have already established that at rest point C his displacement vector is $\vec{D}_{AC} = 0.75\vec{D}_{AB}$. Starting at point C , he walks southwest (toward the campsite), which means his new displacement vector \vec{D}_{CD} from point C to point D is antiparallel to \vec{D}_{AB} . Its magnitude $|\vec{D}_{CD}|$ is $D_{CD} = 1.2 \text{ km} = 0.2D_{AB}$, so his second displacement vector is $\vec{D}_{CD} = -0.2\vec{D}_{AB}$. His total displacement \vec{D}_{AD} relative to the campsite is the **vector sum** of the two displacement vectors: vector \vec{D}_{AC} (from the campsite to the rest point) and vector \vec{D}_{CD} (from the rest point to the point where he finds his box):

Note:
Equation:

$$\vec{D}_{AD} = \vec{D}_{AC} + \vec{D}_{CD}.$$

The vector sum of two (or more) vectors is called the **resultant vector** or, for short, the *resultant*. When the vectors on the right-hand-side of [\[link\]](#) are known, we can find the resultant \vec{D}_{AD} as follows:

Equation:

$$\vec{D}_{AD} = \vec{D}_{AC} + \vec{D}_{CD} = 0.75\vec{D}_{AB} - 0.2\vec{D}_{AB} = (0.75 - 0.2)\vec{D}_{AB} = 0.55\vec{D}_{AB}.$$

When your friend finally reaches the pond at B , his displacement vector \vec{D}_{AB} from point A is the vector sum of his displacement vector \vec{D}_{AD} from point A to point D and his displacement vector \vec{D}_{DB} from point D to the fishing hole: $\vec{D}_{AB} = \vec{D}_{AD} + \vec{D}_{DB}$ (see [\[link\]](#)(c)). This means his displacement vector \vec{D}_{DB} is the **difference of two vectors**:

Equation:

$$\vec{D}_{DB} = \vec{D}_{AB} - \vec{D}_{AD} = \vec{D}_{AB} + (-\vec{D}_{AD}).$$

Notice that a difference of two vectors is nothing more than a vector sum of two vectors because the second term in [\[link\]](#) is vector $-\vec{D}_{AD}$ (which is antiparallel to \vec{D}_{AD}). When we substitute [\[link\]](#) into [\[link\]](#), we obtain the second displacement vector:

Equation:

$$\vec{D}_{DB} = \vec{D}_{AB} - \vec{D}_{AD} = \vec{D}_{AB} - 0.55\vec{D}_{AB} = (1.0 - 0.55)\vec{D}_{AB} = 0.45\vec{D}_{AB}.$$

This result means your friend walked $D_{DB} = 0.45D_{AB} = 0.45(6.0 \text{ km}) = 2.7 \text{ km}$ from the point where he finds his tackle box to the fishing hole.

When vectors \vec{A} and \vec{B} lie along a line (that is, in one dimension), such as in the camping example, their resultant $\vec{R} = \vec{A} + \vec{B}$ and their difference $\vec{D} = \vec{A} - \vec{B}$ both lie along the same direction. We can illustrate the addition or subtraction of vectors by drawing the corresponding vectors to scale in one dimension, as shown in [\[link\]](#).

To illustrate the resultant when \vec{A} and \vec{B} are two parallel vectors, we draw them along one line by placing the origin of one vector at the end of the other vector in head-to-tail fashion (see [\[link\]](#)(b)). The magnitude of this resultant is the sum of their magnitudes: $R = A + B$. The direction of the resultant is parallel to both vectors. When vector \vec{A} is antiparallel to vector \vec{B} , we draw them along one line in either head-to-head fashion ([\[link\]](#)(c)) or tail-to-tail fashion. The magnitude of the vector difference, then, is the *absolute value* $D = |A - B|$ of the difference of their magnitudes. The direction of the difference vector \vec{D} is parallel to the direction of the longer vector.

In general, in one dimension—as well as in higher dimensions, such as in a plane or in space—we can add any number of vectors and we can do so in any order because the addition of vectors is **commutative**,

Note:

Equation:

$$\vec{A} + \vec{B} = \vec{B} + \vec{A},$$

and **associative**,

Note:

Equation:

$$(\vec{\mathbf{A}} + \vec{\mathbf{B}}) + \vec{\mathbf{C}} = \vec{\mathbf{A}} + (\vec{\mathbf{B}} + \vec{\mathbf{C}}).$$

Moreover, multiplication by a scalar is **distributive**:

Note:

Equation:

$$\alpha_1 \vec{\mathbf{A}} + \alpha_2 \vec{\mathbf{A}} = (\alpha_1 + \alpha_2) \vec{\mathbf{A}}.$$

We used the distributive property in [\[link\]](#) and [\[link\]](#).

When adding many vectors in one dimension, it is convenient to use the concept of a **unit vector**. A unit vector, which is denoted by a letter symbol with a hat, such as $\hat{\mathbf{u}}$, has a magnitude of one and does not have any physical unit so that $|\hat{\mathbf{u}}| \equiv u = 1$. The only role of a unit vector is to specify direction. For example, instead of saying vector $\vec{\mathbf{D}}_{AB}$ has a magnitude of 6.0 km and a direction of northeast, we can introduce a unit vector $\hat{\mathbf{u}}$ that points to the northeast and say succinctly that $\vec{\mathbf{D}}_{AB} = (6.0 \text{ km})\hat{\mathbf{u}}$. Then the southwesterly direction is simply given by the unit vector $-\hat{\mathbf{u}}$. In this way, the displacement of 6.0 km in the southwesterly direction is expressed by the vector

Equation:

$$\vec{\mathbf{D}}_{BA} = (-6.0 \text{ km})\hat{\mathbf{u}}.$$

Example:

A Ladybug Walker

A long measuring stick rests against a wall in a physics laboratory with its 200-cm end at the floor. A ladybug lands on the 100-cm mark and crawls randomly along the stick. It first walks 15 cm toward the floor, then it walks 56 cm toward the wall, then it walks 3 cm toward the floor again. Then, after a brief stop, it continues for 25 cm toward the floor and then, again, it crawls up 19 cm toward the wall before coming to a complete rest ([link](#)). Find the vector of its total displacement and its final resting position on the stick.

Strategy

If we choose the direction along the stick toward the floor as the direction of unit vector $\hat{\mathbf{u}}$, then the direction toward the floor is $+\hat{\mathbf{u}}$ and the direction toward the wall is $-\hat{\mathbf{u}}$.

The ladybug makes a total of five displacements:

Equation:

$$\vec{\mathbf{D}}_1 = (15 \text{ cm})(+\hat{\mathbf{u}}),$$

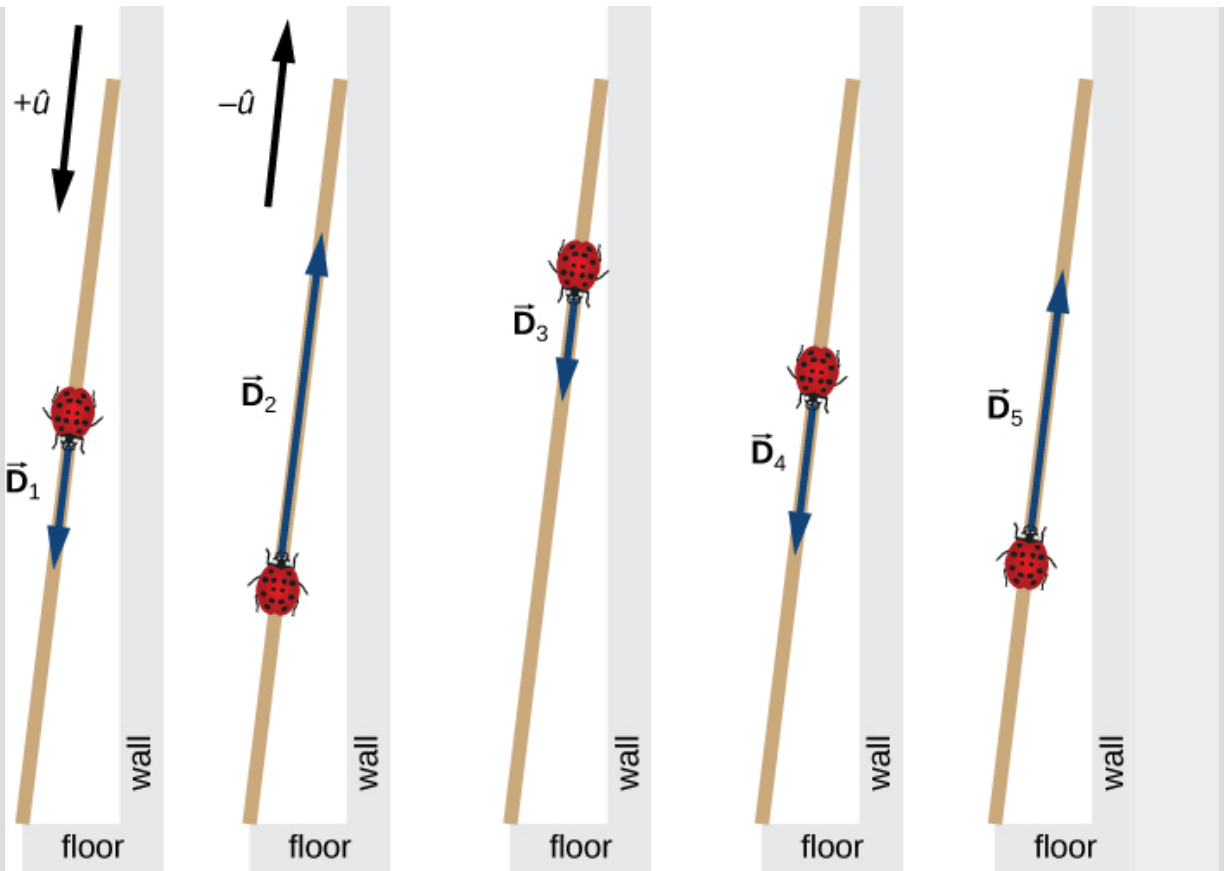
$$\vec{\mathbf{D}}_2 = (56 \text{ cm})(-\hat{\mathbf{u}}),$$

$$\vec{\mathbf{D}}_3 = (3 \text{ cm})(+\hat{\mathbf{u}}),$$

$$\vec{\mathbf{D}}_4 = (25 \text{ cm})(+\hat{\mathbf{u}}), \text{ and}$$

$$\vec{\mathbf{D}}_5 = (19 \text{ cm})(-\hat{\mathbf{u}}).$$

The total displacement $\vec{\mathbf{D}}$ is the resultant of all its displacement vectors.



Five displacements of the ladybug. Note that in this schematic drawing, magnitudes of displacements are not drawn to scale. (credit "ladybug": modification of work by “Persian Poet Gal”/Wikimedia Commons)

Solution

The resultant of all the displacement vectors is

Equation:

$$\begin{aligned}
 \vec{D} &= \vec{D}_1 + \vec{D}_2 + \vec{D}_3 + \vec{D}_4 + \vec{D}_5 \\
 &= (15 \text{ cm})(+\hat{u}) + (56 \text{ cm})(-\hat{u}) + (3 \text{ cm})(+\hat{u}) + (25 \text{ cm})(+\hat{u}) + (19 \text{ cm})(-\hat{u}) \\
 &= (15 - 56 + 3 + 25 - 19)\text{cm}\hat{u} \\
 &= -32 \text{ cm}\hat{u}.
 \end{aligned}$$

In this calculation, we use the distributive law given by [\[link\]](#). The result reads that the total displacement vector points away from the 100-cm mark (initial landing site) toward the end of the meter stick that touches the wall. The end that touches the wall is marked 0 cm, so the final position of the ladybug is at the $(100 - 32)\text{cm} = 68\text{-cm}$ mark.

Note:

Exercise:

Problem:

Check Your Understanding A cave diver enters a long underwater tunnel. When her displacement with respect to the entry point is 20 m, she accidentally drops her camera, but she doesn't notice it missing until she is some 6 m farther into the tunnel. She swims back 10 m but cannot find the camera, so she decides to end the dive. How far from the entry point is she? Taking the positive direction out of the tunnel, what is her displacement vector relative to the entry point?

Solution:

$$16 \text{ m}; \vec{D} = -16 \text{ m}\hat{u}$$

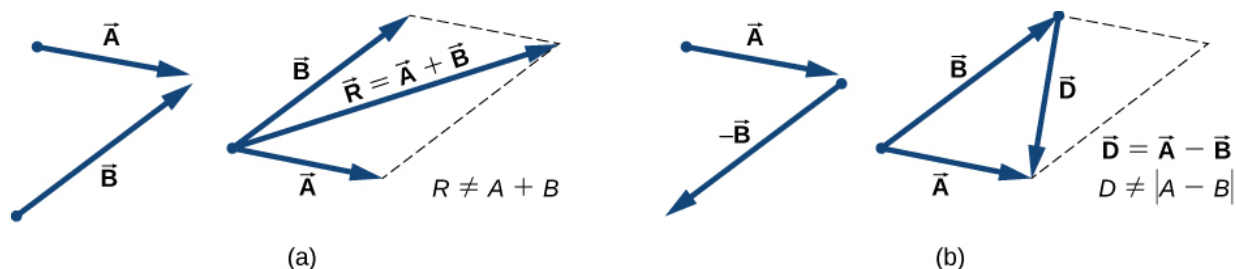
Algebra of Vectors in Two Dimensions

When vectors lie in a plane—that is, when they are in two dimensions—they can be multiplied by scalars, added to other vectors, or subtracted from other vectors in accordance with the general laws expressed by [\[link\]](#), [\[link\]](#), [\[link\]](#), and [\[link\]](#). However, the addition rule for two vectors in a plane becomes more complicated than the rule for vector addition in one dimension. We have to use the laws of geometry to construct resultant vectors, followed by trigonometry to find vector magnitudes and directions. This geometric approach is commonly used in navigation ([\[link\]](#)). In this section, we need to have at hand two rulers, a triangle, a protractor, a pencil, and an eraser for drawing vectors to scale by geometric constructions.



In navigation, the laws of geometry are used to draw resultant displacements on nautical maps.

For a geometric construction of the sum of two vectors in a plane, we follow **the parallelogram rule**. Suppose two vectors \vec{A} and \vec{B} are at the arbitrary positions shown in [\[link\]](#). Translate either one of them in parallel to the beginning of the other vector, so that after the translation, both vectors have their origins at the same point. Now, at the end of vector \vec{A} we draw a line parallel to vector \vec{B} and at the end of vector \vec{B} we draw a line parallel to vector \vec{A} (the dashed lines in [\[link\]](#)). In this way, we obtain a parallelogram. From the origin of the two vectors we draw a diagonal that is the resultant \vec{R} of the two vectors: $\vec{R} = \vec{A} + \vec{B}$ ([\[link\]](#)(a)). The other diagonal of this parallelogram is the vector difference of the two vectors $\vec{D} = \vec{A} - \vec{B}$, as shown in [\[link\]](#)(b). Notice that the end of the difference vector is placed at the end of vector \vec{A} .

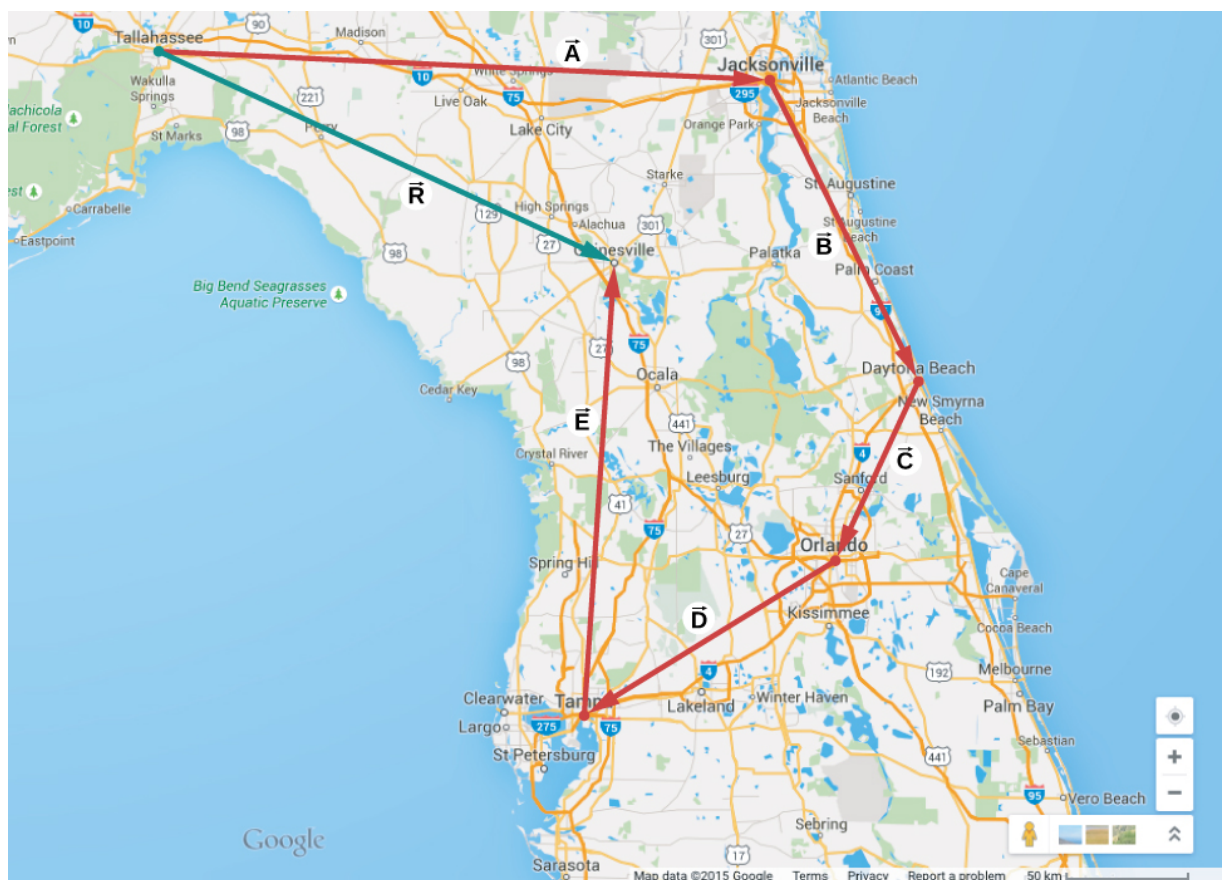


The parallelogram rule for the addition of two vectors. Make the parallel translation of each vector to a point where their origins (marked by the dot) coincide and construct a parallelogram with two sides on the vectors and the other two sides (indicated by dashed lines) parallel to the vectors. (a) Draw the resultant vector \vec{R} along the diagonal of the parallelogram from the common point to the opposite corner. Length R of the resultant vector is *not* equal to the sum of the magnitudes of the two vectors. (b) Draw the difference vector $\vec{D} = \vec{A} - \vec{B}$ along the diagonal connecting the ends of the vectors. Place the origin of vector \vec{D} at the end of vector \vec{B} and the end (arrowhead) of vector \vec{D} at the end of vector \vec{A} . Length D of the difference vector is *not* equal to the difference of magnitudes of the two vectors.

It follows from the parallelogram rule that neither the magnitude of the resultant vector nor the magnitude of the difference vector can be expressed as a simple sum or difference of magnitudes A and B , because the length of a diagonal cannot be expressed as a simple sum of side lengths. When using a geometric construction to find magnitudes $|\vec{R}|$ and $|\vec{D}|$, we have to use trigonometry laws for triangles, which may lead to complicated algebra. There are two ways to circumvent this algebraic complexity. One way is to use the method of components, which we examine in the next section. The other way is to draw the vectors to scale, as is done in navigation, and read approximate vector lengths and angles (directions) from the graphs. In this section we examine the second approach.

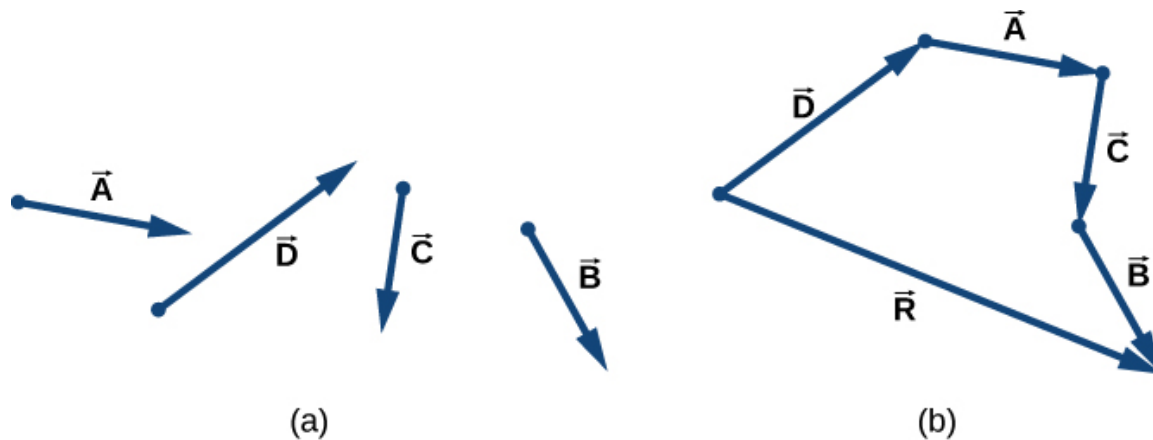
If we need to add three or more vectors, we repeat the parallelogram rule for the pairs of vectors until we find the resultant of all of the resultants. For three vectors, for example, we first find the resultant of vector 1 and vector 2, and then we find the resultant of this resultant and vector 3. The order in which we select the pairs of vectors does not matter because the operation of vector addition is commutative and associative (see [\[link\]](#) and [\[link\]](#)). Before we state a general rule that follows from repetitive applications of the parallelogram rule, let's look at the following example.

Suppose you plan a vacation trip in Florida. Departing from Tallahassee, the state capital, you plan to visit your uncle Joe in Jacksonville, see your cousin Vinny in Daytona Beach, stop for a little fun in Orlando, see a circus performance in Tampa, and visit the University of Florida in Gainesville. Your route may be represented by five displacement vectors \vec{A} , \vec{B} , \vec{C} , \vec{D} , and \vec{E} , which are indicated by the red vectors in [\[link\]](#). What is your total displacement when you reach Gainesville? The total displacement is the vector sum of all five displacement vectors, which may be found by using the parallelogram rule four times. Alternatively, recall that the displacement vector has its beginning at the initial position (Tallahassee) and its end at the final position (Gainesville), so the total displacement vector can be drawn directly as an arrow connecting Tallahassee with Gainesville (see the green vector in [\[link\]](#)). When we use the parallelogram rule four times, the resultant \vec{R} we obtain is exactly this green vector connecting Tallahassee with Gainesville: $\vec{R} = \vec{A} + \vec{B} + \vec{C} + \vec{D} + \vec{E}$.



When we use the parallelogram rule four times, we obtain the resultant vector $\vec{R} = \vec{A} + \vec{B} + \vec{C} + \vec{D} + \vec{E}$, which is the green vector connecting Tallahassee with Gainesville.

Drawing the resultant vector of many vectors can be generalized by using the following **tail-to-head geometric construction**. Suppose we want to draw the resultant vector \vec{R} of four vectors \vec{A} , \vec{B} , \vec{C} , and \vec{D} ([link](a)). We select any one of the vectors as the first vector and make a parallel translation of a second vector to a position where the origin (“tail”) of the second vector coincides with the end (“head”) of the first vector. Then, we select a third vector and make a parallel translation of the third vector to a position where the origin of the third vector coincides with the end of the second vector. We repeat this procedure until all the vectors are in a head-to-tail arrangement like the one shown in [link]. We draw the resultant vector \vec{R} by connecting the origin (“tail”) of the first vector with the end (“head”) of the last vector. The end of the resultant vector is at the end of the last vector. Because the addition of vectors is associative and commutative, we obtain the same resultant vector regardless of which vector we choose to be first, second, third, or fourth in this construction.

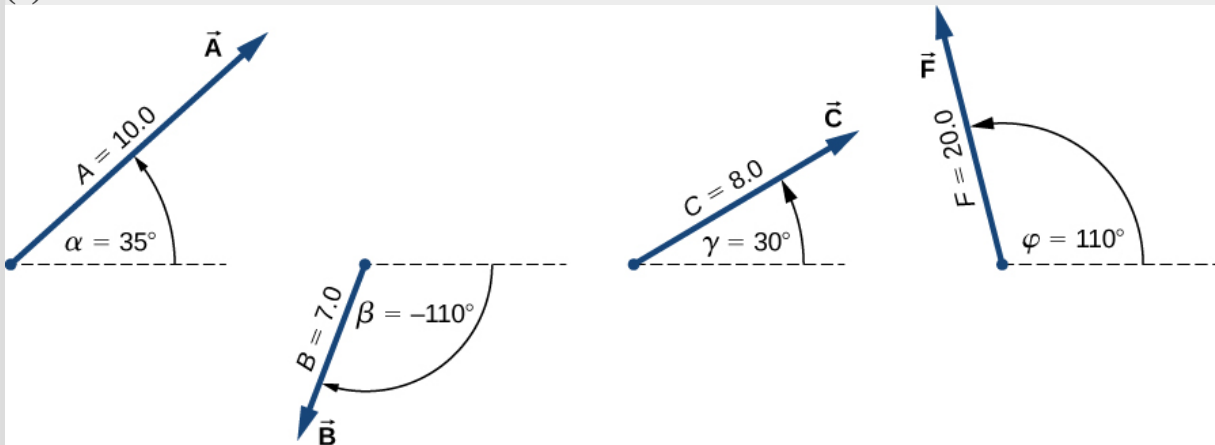


Tail-to-head method for drawing the resultant vector $\vec{R} = \vec{A} + \vec{B} + \vec{C} + \vec{D}$.
 (a) Four vectors of different magnitudes and directions. (b) Vectors in (a) are translated to new positions where the origin (“tail”) of one vector is at the end (“head”) of another vector. The resultant vector is drawn from the origin (“tail”) of the first vector to the end (“head”) of the last vector in this arrangement.

Example:

Geometric Construction of the Resultant

The three displacement vectors \vec{A} , \vec{B} , and \vec{C} in [\[link\]](#) are specified by their magnitudes $A = 10.0$, $B = 7.0$, and $C = 8.0$, respectively, and by their respective direction angles with the horizontal direction $\alpha = 35^\circ$, $\beta = -110^\circ$, and $\gamma = 30^\circ$. The physical units of the magnitudes are centimeters. Choose a convenient scale and use a ruler and a protractor to find the following vector sums: (a) $\vec{R} = \vec{A} + \vec{B}$, (b) $\vec{D} = \vec{A} - \vec{B}$, and (c) $\vec{S} = \vec{A} - 3\vec{B} + \vec{C}$.



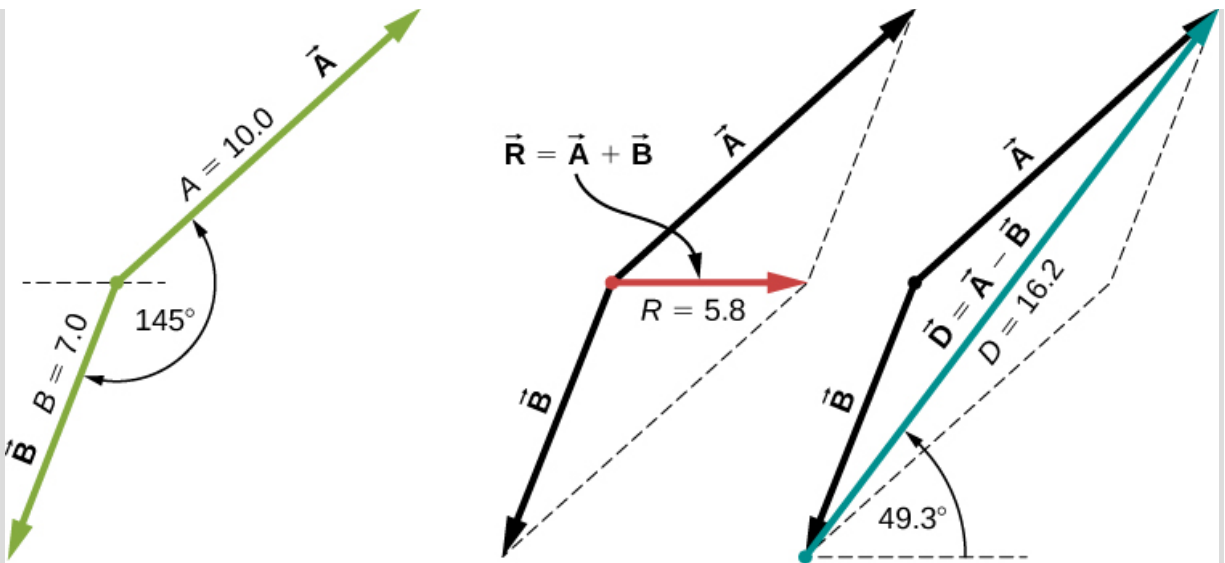
Vectors used in [\[link\]](#) and in the Check Your Understanding feature that follows.

Strategy

In geometric construction, to find a vector means to find its magnitude and its direction angle with the horizontal direction. The strategy is to draw to scale the vectors that appear on the right-hand side of the equation and construct the resultant vector. Then, use a ruler and a protractor to read the magnitude of the resultant and the direction angle. For parts (a) and (b) we use the parallelogram rule. For (c) we use the tail-to-head method.

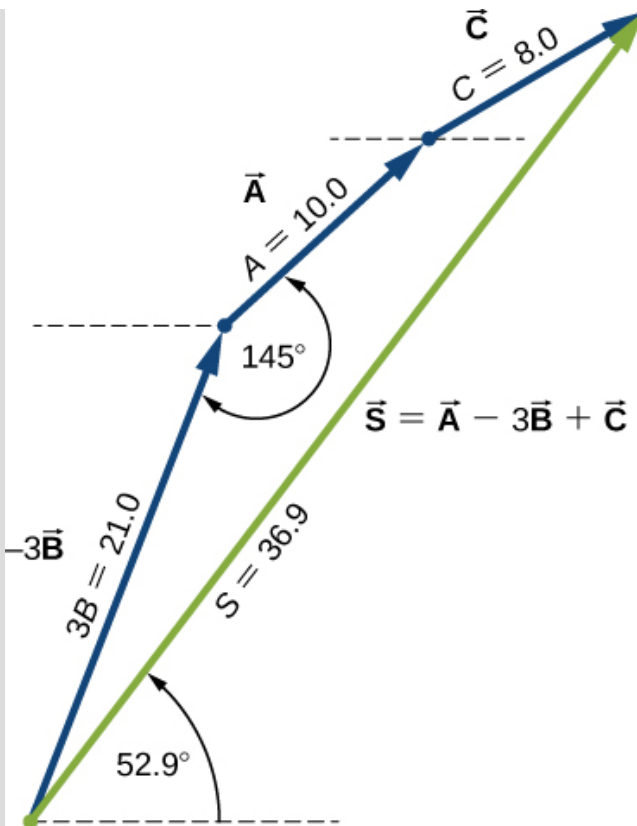
Solution

For parts (a) and (b), we attach the origin of vector \vec{B} to the origin of vector \vec{A} , as shown in [\[link\]](#), and construct a parallelogram. The shorter diagonal of this parallelogram is the sum $\vec{A} + \vec{B}$. The longer of the diagonals is the difference $\vec{A} - \vec{B}$. We use a ruler to measure the lengths of the diagonals, and a protractor to measure the angles with the horizontal. For the resultant \vec{R} , we obtain $R = 5.8$ cm and $\theta_R \approx 0^\circ$. For the difference \vec{D} , we obtain $D = 16.2$ cm and $\theta_D = 49.3^\circ$, which are shown in [\[link\]](#).



Using the parallelogram rule to solve (a) (finding the resultant, red) and (b) (finding the difference, blue).

For (c), we can start with vector $-3\vec{B}$ and draw the remaining vectors tail-to-head as shown in [\[link\]](#). In vector addition, the order in which we draw the vectors is unimportant, but drawing the vectors to scale is very important. Next, we draw vector \vec{S} from the origin of the first vector to the end of the last vector and place the arrowhead at the end of \vec{S} . We use a ruler to measure the length of \vec{S} , and find that its magnitude is $S = 36.9$ cm. We use a protractor and find that its direction angle is $\theta_S = 52.9^\circ$. This solution is shown in [\[link\]](#).



Using the tail-to-head method to solve
(c) (finding vector \vec{S} , green).

Note:

Exercise:

Problem:

Check Your Understanding Using the three displacement vectors \vec{A} , \vec{B} , and \vec{F} in [\[link\]](#), choose a convenient scale, and use a ruler and a protractor to find vector \vec{G} given by the vector equation $\vec{G} = \vec{A} + 2\vec{B} - \vec{F}$.

Solution:

$$G = 28.2 \text{ cm}, \theta_G = 291^\circ$$

Note:

Observe the addition of vectors in a plane by visiting this [vector calculator](#) and this [Phet simulation](#).

Summary

- A vector quantity is any quantity that has magnitude and direction, such as displacement or velocity. Vector quantities are represented by mathematical objects called vectors.
- Geometrically, vectors are represented by arrows, with the end marked by an arrowhead. The length of the vector is its magnitude, which is a positive scalar. On a plane, the direction of a vector is given by the angle the vector makes with a reference direction, often an angle with the horizontal. The direction angle of a vector is a scalar.
- Two vectors are equal if and only if they have the same magnitudes and directions. Parallel vectors have the same direction angles but may have different magnitudes. Antiparallel vectors have direction angles that differ by 180° . Orthogonal vectors have direction angles that differ by 90° .
- When a vector is multiplied by a scalar, the result is another vector of a different length than the length of the original vector. Multiplication by a positive scalar does not change the original direction; only the magnitude is affected. Multiplication by a negative scalar reverses the original direction. The resulting vector is antiparallel to the original vector. Multiplication by a scalar is distributive. Vectors can be divided by nonzero scalars but cannot be divided by vectors.
- Two or more vectors can be added to form another vector. The vector sum is called the resultant vector. We can add vectors to vectors or scalars to scalars, but we cannot add scalars to vectors. Vector addition is commutative and associative.
- To construct a resultant vector of two vectors in a plane geometrically, we use the parallelogram rule. To construct a resultant vector of many vectors in a plane geometrically, we use the tail-to-head method.

Conceptual Questions

Exercise:**Problem:**

A weather forecast states the temperature is predicted to be -5°C the following day. Is this temperature a vector or a scalar quantity? Explain.

Solution:

scalar

Exercise:

Problem:

Which of the following is a vector: a person's height, the altitude on Mt. Everest, the velocity of a fly, the age of Earth, the boiling point of water, the cost of a book, Earth's population, or the acceleration of gravity?

Exercise:

Problem:

Give a specific example of a vector, stating its magnitude, units, and direction.

Solution:

answers may vary

Exercise:

Problem: What do vectors and scalars have in common? How do they differ?

Exercise:

Problem:

Suppose you add two vectors \vec{A} and \vec{B} . What relative direction between them produces the resultant with the greatest magnitude? What is the maximum magnitude? What relative direction between them produces the resultant with the smallest magnitude? What is the minimum magnitude?

Solution:

parallel, sum of magnitudes, antiparallel, zero

Exercise:

Problem: Is it possible to add a scalar quantity to a vector quantity?

Exercise:

Problem:

Is it possible for two vectors of different magnitudes to add to zero? Is it possible for three vectors of different magnitudes to add to zero? Explain.

Solution:

no, yes

Exercise:

Problem:

Does the odometer in an automobile indicate a scalar or a vector quantity?

Exercise:

Problem:

When a 10,000-m runner competing on a 400-m track crosses the finish line, what is the runner's net displacement? Can this displacement be zero? Explain.

Solution:

zero, yes

Exercise:

Problem:

A vector has zero magnitude. Is it necessary to specify its direction? Explain.

Exercise:

Problem: Can a magnitude of a vector be negative?

Solution:

no

Exercise:

Problem:

Can the magnitude of a particle's displacement be greater than the distance traveled?

Exercise:

Problem:

If two vectors are equal, what can you say about their components? What can you say about their magnitudes? What can you say about their directions?

Solution:

equal, equal, the same

Exercise:**Problem:**

If three vectors sum up to zero, what geometric condition do they satisfy?

Problems**Exercise:****Problem:**

A scuba diver makes a slow descent into the depths of the ocean. His vertical position with respect to a boat on the surface changes several times. He makes the first stop 9.0 m from the boat but has a problem with equalizing the pressure, so he ascends 3.0 m and then continues descending for another 12.0 m to the second stop. From there, he ascends 4 m and then descends for 18.0 m, ascends again for 7 m and descends again for 24.0 m, where he makes a stop, waiting for his buddy. Assuming the positive direction up to the surface, express his net vertical displacement vector in terms of the unit vector. What is his distance to the boat?

Solution:

$$\vec{h} = -49 \text{ m}\hat{u}, 49 \text{ m}$$

Exercise:**Problem:**

In a tug-of-war game on one campus, 15 students pull on a rope at both ends in an effort to displace the central knot to one side or the other. Two students pull with force 196 N each to the right, four students pull with force 98 N each to the left, five students pull with force 62 N each to the left, three students pull with force 150 N each to the right, and one student pulls with force 250 N to the left. Assuming the positive direction to the right, express the net pull on the knot in terms of the unit vector. How big is the net pull on the knot? In what direction?

Exercise:**Problem:**

Suppose you walk 18.0 m straight west and then 25.0 m straight north. How far are you from your starting point and what is the compass direction of a line connecting your starting point to your final position? Use a graphical method.

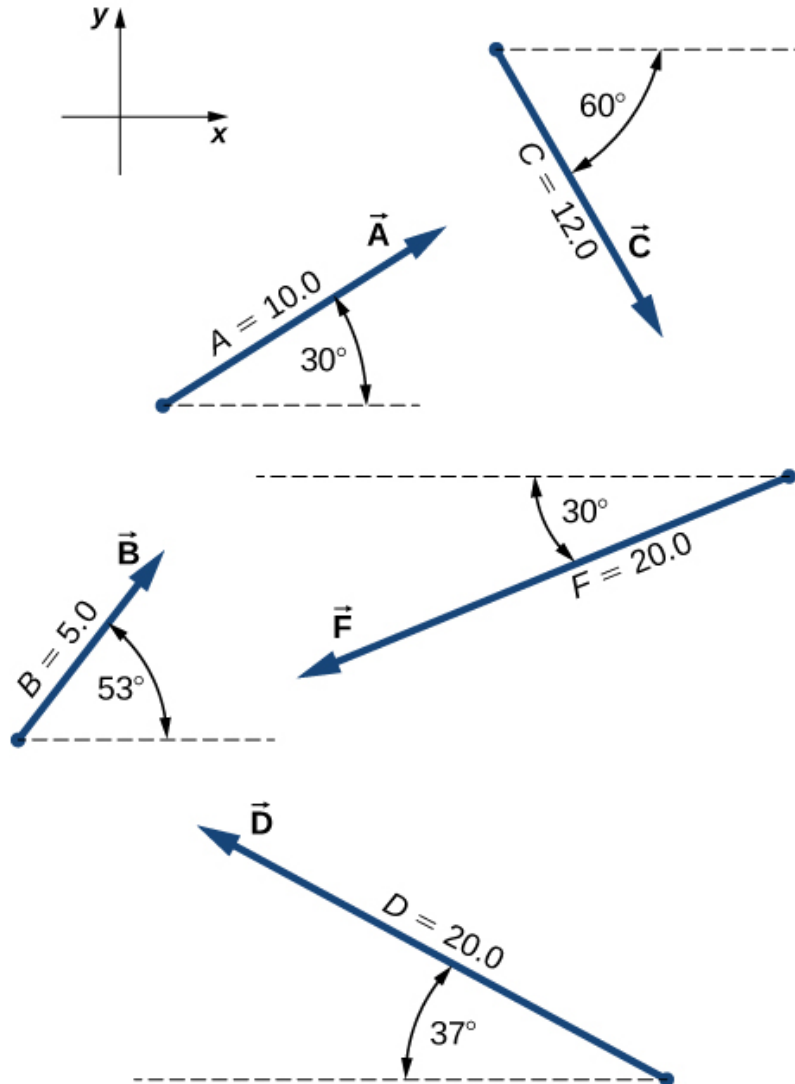
Solution:

30.8 m, 35.7° west of north

Exercise:

Problem:

For the vectors given in the following figure, use a graphical method to find the following resultants: (a) $\vec{A} + \vec{B}$, (b) $\vec{C} + \vec{B}$, (c) $\vec{D} + \vec{F}$, (d) $\vec{A} - \vec{B}$, (e) $\vec{D} - \vec{F}$, (f) $\vec{A} + 2\vec{F}$, (g) $\vec{C} - 2\vec{D} + 3\vec{F}$; and (h) $\vec{A} - 4\vec{D} + 2\vec{F}$.



Exercise:

Problem:

A delivery man starts at the post office, drives 40 km north, then 20 km west, then 60 km northeast, and finally 50 km north to stop for lunch. Use a graphical method to find his net displacement vector.

Solution:

134 km, 80°

Exercise:**Problem:**

An adventurous dog strays from home, runs three blocks east, two blocks north, one block east, one block north, and two blocks west. Assuming that each block is about 100 m, how far from home and in what direction is the dog? Use a graphical method.

Exercise:**Problem:**

In an attempt to escape a desert island, a castaway builds a raft and sets out to sea. The wind shifts a great deal during the day and he is blown along the following directions: 2.50 km and 45.0° north of west, then 4.70 km and 60.0° south of east, then 1.30 km and 25.0° south of west, then 5.10 km straight east, then 1.70 km and 5.00° east of north, then 7.20 km and 55.0° south of west, and finally 2.80 km and 10.0° north of east. Use a graphical method to find the castaway's final position relative to the island.

Solution:

7.34 km, 63.5° south of east

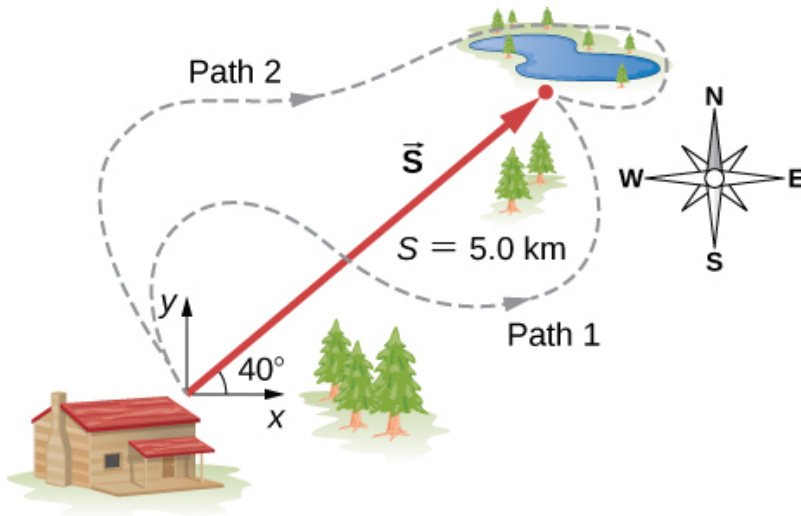
Exercise:**Problem:**

A small plane flies 40.0 km in a direction 60° north of east and then flies 30.0 km in a direction 15° north of east. Use a graphical method to find the total distance the plane covers from the starting point and the direction of the path to the final position.

Exercise:

Problem:

A trapper walks a 5.0-km straight-line distance from his cabin to the lake, as shown in the following figure. Use a graphical method (the parallelogram rule) to determine the trapper's displacement directly to the east and displacement directly to the north that sum up to his resultant displacement vector. If the trapper walked only in directions east and north, zigzagging his way to the lake, how many kilometers would he have to walk to get to the lake?



Solution:

3.8 km east, 3.2 km north, 7.0 km

Exercise:**Problem:**

A surveyor measures the distance across a river that flows straight north by the following method. Starting directly across from a tree on the opposite bank, the surveyor walks 100 m along the river to establish a baseline. She then sights across to the tree and reads that the angle from the baseline to the tree is 35° . How wide is the river?

Exercise:**Problem:**

A pedestrian walks 6.0 km east and then 13.0 km north. Use a graphical method to find the pedestrian's resultant displacement and geographic direction.

Solution:

14.3 km, 65°

Exercise:

Problem:

The magnitudes of two displacement vectors are $A = 20$ m and $B = 6$ m. What are the largest and the smallest values of the magnitude of the resultant $\vec{R} = \vec{A} + \vec{B}$?

Glossary

antiparallel vectors

two vectors with directions that differ by 180°

associative

terms can be grouped in any fashion

commutative

operations can be performed in any order

difference of two vectors

vector sum of the first vector with the vector antiparallel to the second

displacement

change in position

distributive

multiplication can be distributed over terms in summation

magnitude

length of a vector

orthogonal vectors

two vectors with directions that differ by exactly 90° , synonymous with perpendicular vectors

parallelogram rule

geometric construction of the vector sum in a plane

parallel vectors

two vectors with exactly the same direction angles

resultant vector

vector sum of two (or more) vectors

scalar

a number, synonymous with a scalar quantity in physics

scalar equation

equation in which the left-hand and right-hand sides are numbers

scalar quantity

quantity that can be specified completely by a single number with an appropriate physical unit

tail-to-head geometric construction

geometric construction for drawing the resultant vector of many vectors

unit vector

vector of a unit magnitude that specifies direction; has no physical unit

vector

mathematical object with magnitude and direction

vector equation

equation in which the left-hand and right-hand sides are vectors

vector quantity

physical quantity described by a mathematical vector—that is, by specifying both its magnitude and its direction; synonymous with a vector in physics

vector sum

resultant of the combination of two (or more) vectors

Coordinate Systems and Components of a Vector

By the end of this section, you will be able to:

- Describe vectors in two and three dimensions in terms of their components, using unit vectors along the axes.
- Distinguish between the vector components of a vector and the scalar components of a vector.
- Explain how the magnitude of a vector is defined in terms of the components of a vector.
- Identify the direction angle of a vector in a plane.
- Explain the connection between polar coordinates and Cartesian coordinates in a plane.

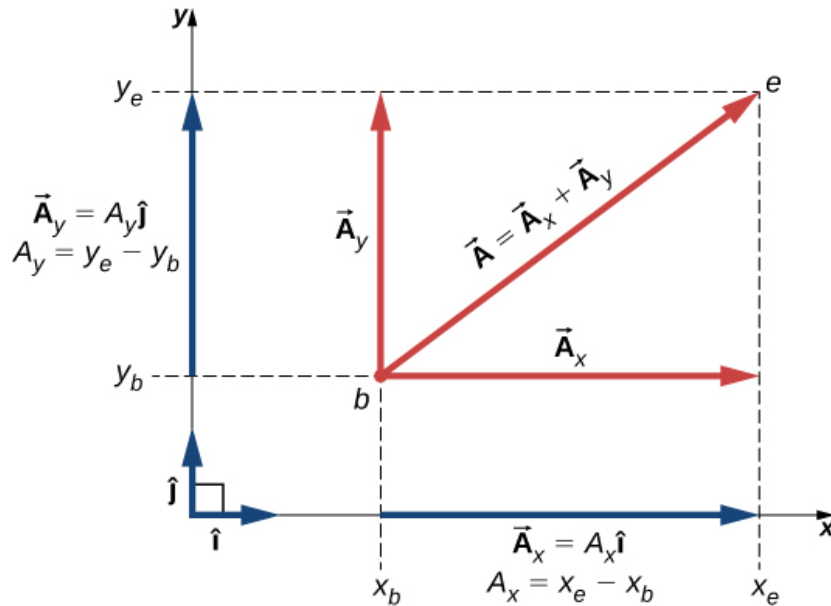
Vectors are usually described in terms of their components in a coordinate system. Even in everyday life we naturally invoke the concept of orthogonal projections in a rectangular coordinate system. For example, if you ask someone for directions to a particular location, you will more likely be told to go 40 km east and 30 km north than 50 km in the direction 37° north of east.

In a rectangular (Cartesian) xy -coordinate system in a plane, a point in a plane is described by a pair of coordinates (x, y) . In a similar fashion, a vector $\vec{\mathbf{A}}$ in a plane is described by a pair of its *vector* coordinates. The x -coordinate of vector $\vec{\mathbf{A}}$ is called its x -component and the y -coordinate of vector $\vec{\mathbf{A}}$ is called its y -component. The vector x -component is a vector denoted by $\vec{\mathbf{A}}_x$. The vector y -component is a vector denoted by $\vec{\mathbf{A}}_y$. In the Cartesian system, the x and y **vector components** of a vector are the orthogonal projections of this vector onto the x - and y -axes, respectively. In this way, following the parallelogram rule for vector addition, each vector on a Cartesian plane can be expressed as the vector sum of its vector components:

Equation:

$$\vec{\mathbf{A}} = \vec{\mathbf{A}}_x + \vec{\mathbf{A}}_y.$$

As illustrated in [\[link\]](#), vector $\vec{\mathbf{A}}$ is the diagonal of the rectangle where the x -component $\vec{\mathbf{A}}_x$ is the side parallel to the x -axis and the y -component $\vec{\mathbf{A}}_y$ is the side parallel to the y -axis. Vector component $\vec{\mathbf{A}}_x$ is orthogonal to vector component $\vec{\mathbf{A}}_y$.



Vector $\vec{\mathbf{A}}$ in a plane in the Cartesian coordinate system is the vector sum of its vector x - and y -components. The x -vector component $\vec{\mathbf{A}}_x$ is the orthogonal projection of vector $\vec{\mathbf{A}}$ onto the x -axis. The y -vector component $\vec{\mathbf{A}}_y$ is the orthogonal projection of vector $\vec{\mathbf{A}}$ onto the y -axis. The numbers A_x and A_y that multiply the unit vectors are the scalar components of the vector.

It is customary to denote the positive direction on the x -axis by the unit vector $\hat{\mathbf{i}}$ and the positive direction on the y -axis by the unit vector $\hat{\mathbf{j}}$. **Unit vectors of the axes, $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$** , define two orthogonal directions in the plane. As shown in [\[link\]](#), the x - and y - components of a vector can now be written in terms of the unit vectors of the axes:

Equation:

$$\begin{cases} \vec{\mathbf{A}}_x = A_x \hat{\mathbf{i}} \\ \vec{\mathbf{A}}_y = A_y \hat{\mathbf{j}}. \end{cases}$$

The vectors $\vec{\mathbf{A}}_x$ and $\vec{\mathbf{A}}_y$ defined by [\[link\]](#) are the *vector components* of vector $\vec{\mathbf{A}}$. The numbers A_x and A_y that define the vector components in [\[link\]](#) are the **scalar components** of vector $\vec{\mathbf{A}}$. Combining [\[link\]](#) with [\[link\]](#), we obtain **the component form of a vector**:

Note:

Equation:

$$\vec{\mathbf{A}} = A_x \hat{\mathbf{i}} + A_y \hat{\mathbf{j}}.$$

If we know the coordinates $b(x_b, y_b)$ of the origin point of a vector (where b stands for “beginning”) and the coordinates $e(x_e, y_e)$ of the end point of a vector (where e stands for “end”), we can obtain the scalar components of a vector simply by subtracting the origin point coordinates from the end point coordinates:

Note:**Equation:**

$$\begin{cases} A_x = x_e - x_b \\ A_y = y_e - y_b. \end{cases}$$

Example:**Displacement of a Mouse Pointer**

A mouse pointer on the display monitor of a computer at its initial position is at point $b(6.0 \text{ cm}, 1.6 \text{ cm})$ with respect to the lower left-side corner. If you move the pointer to an icon located at point $e(2.0 \text{ cm}, 4.5 \text{ cm})$, what is the displacement vector of the pointer?

Strategy

The origin of the xy -coordinate system is the lower left-side corner of the computer monitor. Therefore, the unit vector $\hat{\mathbf{i}}$ on the x -axis points horizontally to the right and the unit vector $\hat{\mathbf{j}}$ on the y -axis points vertically upward. The origin of the displacement vector is located at point $b(6.0, 1.6)$ and the end of the displacement vector is located at point $e(2.0, 4.5)$. Substitute the coordinates of these points into [\[link\]](#) to find the scalar components D_x and D_y of the displacement vector $\vec{\mathbf{D}}$. Finally, substitute the coordinates into [\[link\]](#) to write the displacement vector in the vector component form.

Solution

We identify $x_b = 6.0$, $x_e = 2.0$, $y_b = 1.6$, and $y_e = 4.5$, where the physical unit is 1 cm. The scalar x - and y -components of the displacement vector are

Equation:

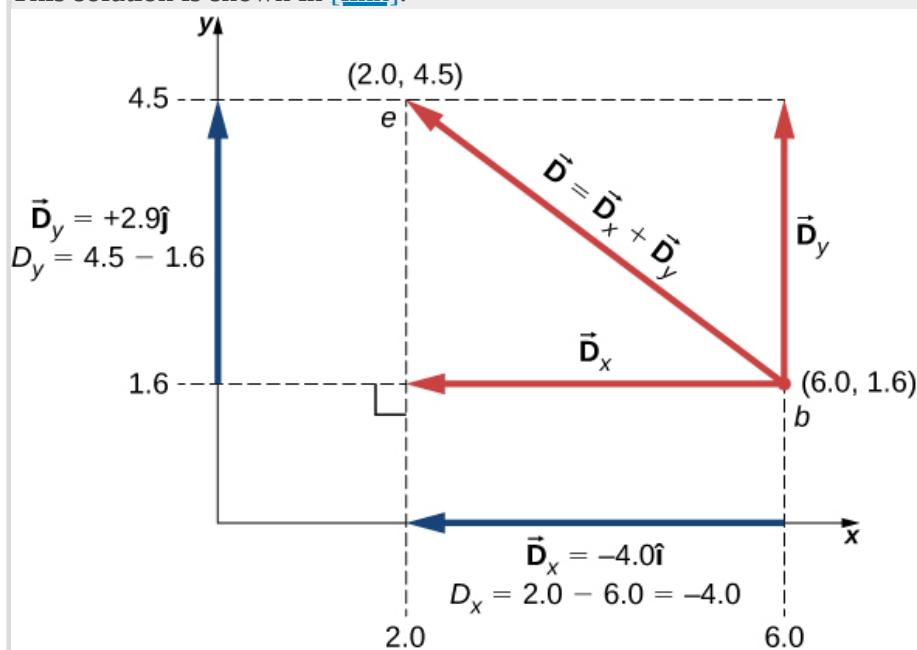
$$\begin{aligned} D_x &= x_e - x_b = (2.0 - 6.0)\text{cm} = -4.0 \text{ cm}, \\ D_y &= y_e - y_b = (4.5 - 1.6)\text{cm} = +2.9 \text{ cm}. \end{aligned}$$

The vector component form of the displacement vector is

Equation:

$$\vec{\mathbf{D}} = D_x \hat{\mathbf{i}} + D_y \hat{\mathbf{j}} = (-4.0 \text{ cm})\hat{\mathbf{i}} + (2.9 \text{ cm})\hat{\mathbf{j}} = (-4.0\hat{\mathbf{i}} + 2.9\hat{\mathbf{j}})\text{cm}.$$

This solution is shown in [\[link\]](#).



The graph of the displacement vector. The vector points from the origin point at b to the end point at e .

Significance

Notice that the physical unit—here, 1 cm—can be placed either with each component immediately before the unit vector or globally for both components, as in [\[link\]](#). Often, the latter way is more convenient because it is simpler.

The vector x -component $\vec{D}_x = -4.0\hat{i} = 4.0(-\hat{i})$ of the displacement vector has the magnitude $|\vec{D}_x| = |-4.0||\hat{i}| = 4.0$ because the magnitude of the unit vector is $|\hat{i}| = 1$. Notice, too, that the direction of the x -component is $-\hat{i}$, which is antiparallel to the direction of the $+x$ -axis; hence, the x -component vector \vec{D}_x points to the left, as shown in [\[link\]](#). The scalar x -component of vector \vec{D} is $D_x = -4.0$.

Similarly, the vector y -component $\vec{D}_y = +2.9\hat{j}$ of the displacement vector has magnitude $|\vec{D}_y| = |2.9||\hat{j}| = 2.9$ because the magnitude of the unit vector is $|\hat{j}| = 1$. The direction of the y -component is $+\hat{j}$, which is parallel to the direction of the $+y$ -axis. Therefore, the y -component vector \vec{D}_y points up, as seen in [\[link\]](#). The scalar y -component of vector \vec{D} is $D_y = +2.9$. The displacement vector \vec{D} is the resultant of its two *vector* components.

The vector component form of the displacement vector [\[link\]](#) tells us that the mouse pointer has been moved on the monitor 4.0 cm to the left and 2.9 cm upward from its initial position.

Note:

Exercise:

Problem:

Check Your Understanding A blue fly lands on a sheet of graph paper at a point located 10.0 cm to the right of its left edge and 8.0 cm above its bottom edge and walks slowly to a point located 5.0 cm from the left edge and 5.0 cm from the bottom edge. Choose the rectangular coordinate system with the origin at the lower left-side corner of the paper and find the displacement vector of the fly. Illustrate your solution by graphing.

Solution:

$\vec{D} = (-5.0\hat{i} - 3.0\hat{j})\text{cm}$; the fly moved 5.0 cm to the left and 3.0 cm down from its landing site.

When we know the scalar components A_x and A_y of a vector \vec{A} , we can find its magnitude A and its direction angle θ_A . The **direction angle**—or direction, for short—is the angle the vector forms with the positive direction on the x-axis. The angle θ_A is measured in the *counterclockwise direction* from the +x-axis to the vector ([link](#)). Because the lengths A , A_x , and A_y form a right triangle, they are related by the Pythagorean theorem:

Note:

Equation:

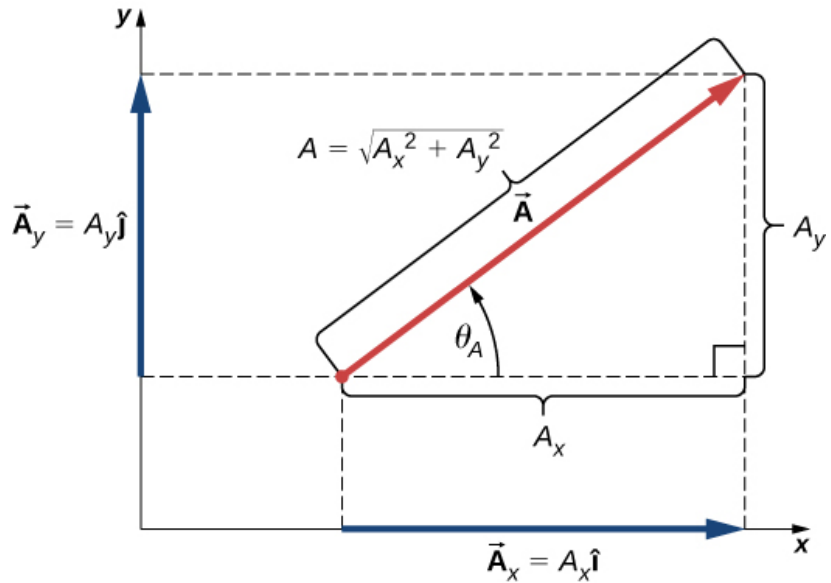
$$A^2 = A_x^2 + A_y^2 \Leftrightarrow A = \sqrt{A_x^2 + A_y^2}.$$

This equation works even if the scalar components of a vector are negative. The direction angle θ_A of a vector is defined via the tangent function of angle θ_A in the triangle shown in [link](#):

Note:

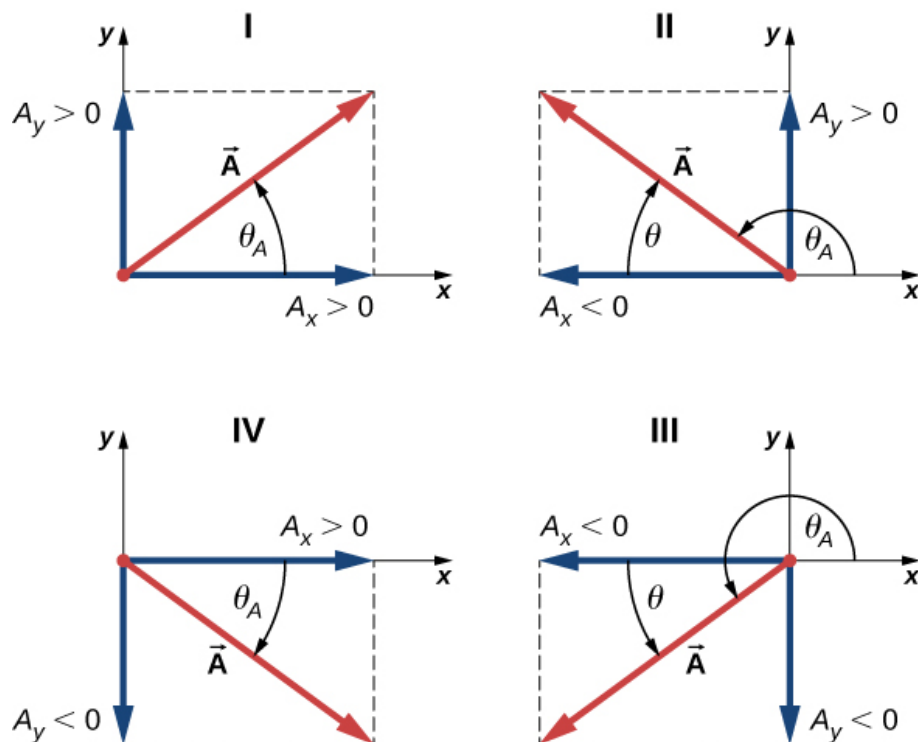
Equation:

$$\tan \theta = \frac{A_y}{A_x}$$



When the vector lies either in the first quadrant or in the fourth quadrant, where component A_x is positive (Figure 2.19), the direction angle θ_A in Equation 2.16) is identical to the angle θ .

When the vector lies either in the first quadrant or in the fourth quadrant, where component A_x is positive ([\[link\]](#)), the angle θ in [\[link\]](#) is identical to the direction angle θ_A . For vectors in the fourth quadrant, angle θ is negative, which means that for these vectors, direction angle θ_A is measured *clockwise* from the positive x -axis. Similarly, for vectors in the second quadrant, angle θ is negative. When the vector lies in either the second or third quadrant, where component A_x is negative, the direction angle is $\theta_A = \theta + 180^\circ$ ([\[link\]](#)).



Scalar components of a vector may be positive or negative. Vectors in the first quadrant (I) have both scalar components positive and vectors in the third quadrant have both scalar components negative. For vectors in quadrants II and III, the direction angle of a vector is $\theta_A = \theta + 180^\circ$.

Example:

Magnitude and Direction of the Displacement Vector

You move a mouse pointer on the display monitor from its initial position at point (6.0 cm, 1.6 cm) to an icon located at point (2.0 cm, 4.5 cm). What are the magnitude and direction of the displacement vector of the pointer?

Strategy

In [\[link\]](#), we found the displacement vector \vec{D} of the mouse pointer (see [\[link\]](#)). We identify its scalar components $D_x = -4.0$ cm and $D_y = +2.9$ cm and substitute into [\[link\]](#) and [\[link\]](#) to find the magnitude D and direction θ_D , respectively.

Solution

The magnitude of vector \vec{D} is

Equation:

$$D = \sqrt{D_x^2 + D_y^2} = \sqrt{(-4.0 \text{ cm})^2 + (2.9 \text{ cm})^2} = \sqrt{(4.0)^2 + (2.9)^2} \text{ cm} = 4.9 \text{ cm}.$$

The direction angle is

Equation:

$$\tan \theta = \frac{D_y}{D_x} = \frac{+2.9 \text{ cm}}{-4.0 \text{ cm}} = -0.725 \Rightarrow \theta = \tan^{-1}(-0.725) = -35.9^\circ.$$

Vector \vec{D} lies in the second quadrant, so its direction angle is

Equation:

$$\theta_D = \theta + 180^\circ = -35.9^\circ + 180^\circ = 144.1^\circ.$$

Note:

Exercise:

Problem:

Check Your Understanding If the displacement vector of a blue fly walking on a sheet of graph paper is $\vec{D} = (-5.00\hat{i} - 3.00\hat{j})\text{cm}$, find its magnitude and direction.

Solution:

5.83 cm, 211°

In many applications, the magnitudes and directions of vector quantities are known and we need to find the resultant of many vectors. For example, imagine 400 cars moving on the Golden Gate Bridge in San Francisco in a strong wind. Each car gives the bridge a different push in various directions and we would like to know how big the resultant push can possibly be. We have already gained some experience with the geometric construction of vector sums, so we know the task of finding the resultant by drawing the vectors and measuring their lengths and angles may become intractable pretty quickly, leading to huge errors. Worries like this do not appear when we use analytical methods. The very first step in an analytical approach is to find vector components when the direction and magnitude of a vector are known.

Let us return to the right triangle in [\[link\]](#). The quotient of the adjacent side A_x to the hypotenuse A is the cosine function of direction angle θ_A , $A_x/A = \cos \theta_A$, and the quotient of the opposite side A_y to the hypotenuse A is the sine function of θ_A , $A_y/A = \sin \theta_A$. When magnitude A and direction θ_A are known, we can solve these relations for the scalar components:

Note:

Equation:

$$\begin{cases} A_x = A \cos \theta_A \\ A_y = A \sin \theta_A \end{cases}$$

When calculating vector components with [\[link\]](#), care must be taken with the angle. The direction angle θ_A of a vector is the angle measured *counterclockwise* from the positive direction on the x -axis to the vector. The clockwise measurement gives a negative angle.

Example:

Components of Displacement Vectors

A rescue party for a missing child follows a search dog named Trooper. Trooper wanders a lot and makes many trial sniffs along many different paths. Trooper eventually finds the child and the story has a happy ending, but his displacements on various legs seem to be truly convoluted. On one of the legs he walks 200.0 m southeast, then he runs north some 300.0 m. On the third leg, he examines the scents carefully for 50.0 m in the direction 30° west of north. On the fourth leg, Trooper goes directly south for 80.0 m, picks up a fresh scent and turns 23° west of south for 150.0 m. Find the scalar components of Trooper's displacement vectors and his displacement vectors in vector component form for each leg.

Strategy

Let's adopt a rectangular coordinate system with the positive x -axis in the direction of geographic east, with the positive y -direction pointed to geographic north. Explicitly, the unit vector $\hat{\mathbf{i}}$ of the x -axis points east and the unit vector $\hat{\mathbf{j}}$ of the y -axis points north. Trooper makes five legs, so there are five displacement vectors. We start by identifying their magnitudes and direction angles, then we use [\[link\]](#) to find the scalar components of the displacements and [\[link\]](#) for the displacement vectors.

Solution

On the first leg, the displacement magnitude is $L_1 = 200.0$ m and the direction is southeast. For direction angle θ_1 we can take either 45° measured clockwise from the east direction or $45^\circ + 270^\circ$ measured counterclockwise from the east direction. With the first choice, $\theta_1 = -45^\circ$. With the second choice, $\theta_1 = +315^\circ$. We can use either one of these two angles. The components are

Equation:

$$\begin{aligned} L_{1x} &= L_1 \cos \theta_1 = (200.0 \text{ m}) \cos 315^\circ = 141.4 \text{ m}, \\ L_{1y} &= L_1 \sin \theta_1 = (200.0 \text{ m}) \sin 315^\circ = -141.4 \text{ m}. \end{aligned}$$

The displacement vector of the first leg is

Equation:

$$\vec{\mathbf{L}}_1 = L_{1x}\hat{\mathbf{i}} + L_{1y}\hat{\mathbf{j}} = (141.4\hat{\mathbf{i}} - 141.4\hat{\mathbf{j}}) \text{ m}.$$

On the second leg of Trooper's wanderings, the magnitude of the displacement is $L_2 = 300.0$ m and the direction is north. The direction angle is $\theta_2 = +90^\circ$. We obtain the following results:

Equation:

$$\begin{aligned}
 L_{2x} &= L_2 \cos \theta_2 = (300.0 \text{ m}) \cos 90^\circ = 0.0, \\
 L_{2y} &= L_2 \sin \theta_2 = (300.0 \text{ m}) \sin 90^\circ = 300.0 \text{ m}, \\
 \vec{L}_2 &= L_{2x}\hat{\mathbf{i}} + L_{2y}\hat{\mathbf{j}} = (300.0 \text{ m})\hat{\mathbf{j}}.
 \end{aligned}$$

On the third leg, the displacement magnitude is $L_3 = 50.0 \text{ m}$ and the direction is 30° west of north. The direction angle measured counterclockwise from the eastern direction is $\theta_3 = 30^\circ + 90^\circ = +120^\circ$. This gives the following answers:

Equation:

$$\begin{aligned}
 L_{3x} &= L_3 \cos \theta_3 = (50.0 \text{ m}) \cos 120^\circ = -25.0 \text{ m}, \\
 L_{3y} &= L_3 \sin \theta_3 = (50.0 \text{ m}) \sin 120^\circ = +43.3 \text{ m}, \\
 \vec{L}_3 &= L_{3x}\hat{\mathbf{i}} + L_{3y}\hat{\mathbf{j}} = (-25.0\hat{\mathbf{i}} + 43.3\hat{\mathbf{j}})\text{m}.
 \end{aligned}$$

On the fourth leg of the excursion, the displacement magnitude is $L_4 = 80.0 \text{ m}$ and the direction is south. The direction angle can be taken as either $\theta_4 = -90^\circ$ or $\theta_4 = +270^\circ$. We obtain

Equation:

$$\begin{aligned}
 L_{4x} &= L_4 \cos \theta_4 = (80.0 \text{ m}) \cos (-90^\circ) = 0, \\
 L_{4y} &= L_4 \sin \theta_4 = (80.0 \text{ m}) \sin (-90^\circ) = -80.0 \text{ m}, \\
 \vec{L}_4 &= L_{4x}\hat{\mathbf{i}} + L_{4y}\hat{\mathbf{j}} = (-80.0 \text{ m})\hat{\mathbf{j}}.
 \end{aligned}$$

On the last leg, the magnitude is $L_5 = 150.0 \text{ m}$ and the angle is $\theta_5 = -23^\circ + 270^\circ = +247^\circ$ (23° west of south), which gives

Equation:

$$\begin{aligned}
 L_{5x} &= L_5 \cos \theta_5 = (150.0 \text{ m}) \cos 247^\circ = -58.6 \text{ m}, \\
 L_{5y} &= L_5 \sin \theta_5 = (150.0 \text{ m}) \sin 247^\circ = -138.1 \text{ m}, \\
 \vec{L}_5 &= L_{5x}\hat{\mathbf{i}} + L_{5y}\hat{\mathbf{j}} = (-58.6\hat{\mathbf{i}} - 138.1\hat{\mathbf{j}})\text{m}.
 \end{aligned}$$

Note:

Exercise:

Problem:

Check Your Understanding If Trooper runs 20 m west before taking a rest, what is his displacement vector?

Solution:

$$\vec{D} = (-20 \text{ m})\hat{\mathbf{i}}$$

Polar Coordinates

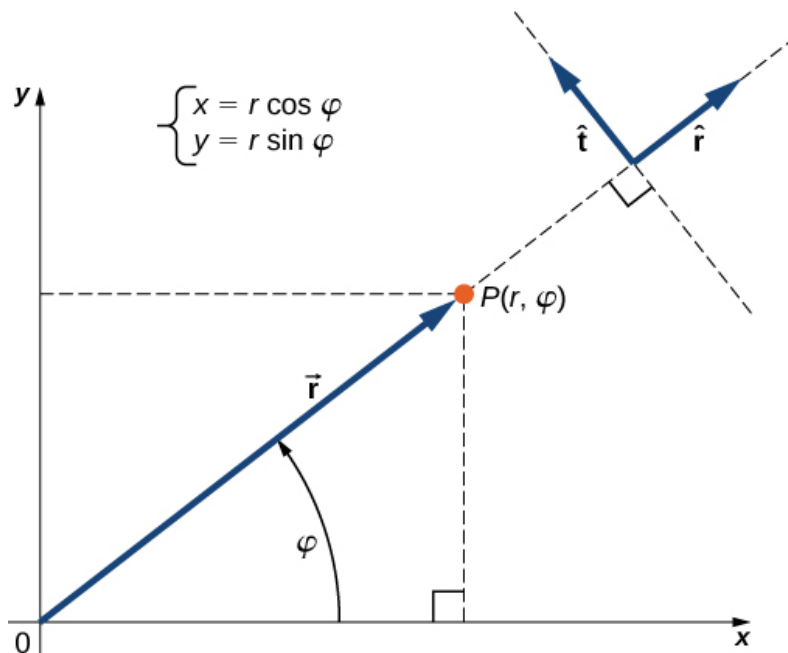
To describe locations of points or vectors in a plane, we need two orthogonal directions. In the Cartesian coordinate system these directions are given by unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ along the x -axis and the y -axis, respectively. The Cartesian coordinate system is very convenient to use in describing displacements and velocities of objects and the forces acting on them. However, it becomes cumbersome when we need to describe the rotation of objects. When describing rotation, we usually work in the **polar coordinate system**.

In the polar coordinate system, the location of point P in a plane is given by two **polar coordinates** ([\[link\]](#)). The first polar coordinate is the **radial coordinate** r , which is the distance of point P from the origin. The second polar coordinate is an angle φ that the radial vector makes with some chosen direction, usually the positive x -direction. In polar coordinates, angles are measured in radians, or rads. The radial vector is attached at the origin and points away from the origin to point P . This radial direction is described by a unit radial vector $\hat{\mathbf{r}}$. The second unit vector $\hat{\mathbf{t}}$ is a vector orthogonal to the radial direction $\hat{\mathbf{r}}$. The positive $+\hat{\mathbf{t}}$ direction indicates how the angle φ changes in the counterclockwise direction. In this way, a point P that has coordinates (x, y) in the rectangular system can be described equivalently in the polar coordinate system by the two polar coordinates (r, φ) . [\[link\]](#) is valid for any vector, so we can use it to express the x - and y -coordinates of vector $\vec{\mathbf{r}}$. In this way, we obtain the connection between the polar coordinates and rectangular coordinates of point P :

Note:

Equation:

$$\begin{cases} x = r \cos \varphi \\ y = r \sin \varphi \end{cases}.$$



Using polar coordinates, the unit vector \hat{r} defines the positive direction along the radius r (radial direction) and, orthogonal to it, the unit vector \hat{t} defines the positive direction of rotation by the angle φ .

Example:

Polar Coordinates

A treasure hunter finds one silver coin at a location 20.0 m away from a dry well in the direction 20° north of east and finds one gold coin at a location 10.0 m away from the well in the direction 20° north of west. What are the polar and rectangular coordinates of these findings with respect to the well?

Strategy

The well marks the origin of the coordinate system and east is the $+x$ -direction. We identify radial distances from the locations to the origin, which are $r_S = 20.0$ m (for the silver coin) and $r_G = 10.0$ m (for the gold coin). To find the angular coordinates, we convert 20° to radians: $20^\circ = \pi 20/180 = \pi/9$. We use [\[link\]](#) to find the x - and y -coordinates of the coins.

Solution

The angular coordinate of the silver coin is $\varphi_S = \pi/9$, whereas the angular coordinate of the gold coin is $\varphi_G = \pi - \pi/9 = 8\pi/9$. Hence, the polar coordinates of the silver coin are $(r_S, \varphi_S) = (20.0 \text{ m}, \pi/9)$ and those of the gold coin are $(r_G, \varphi_G) = (10.0 \text{ m}, 8\pi/9)$. We substitute these coordinates into [\[link\]](#) to obtain rectangular coordinates. For the gold coin, the coordinates are

Equation:

$$\begin{cases} x_G = r_G \cos \varphi_G = (10.0 \text{ m}) \cos 8\pi/9 = -9.4 \text{ m} \\ y_G = r_G \sin \varphi_G = (10.0 \text{ m}) \sin 8\pi/9 = 3.4 \text{ m} \end{cases} \Rightarrow (x_G, y_G) = (-9.4 \text{ m}, 3.4 \text{ m}).$$

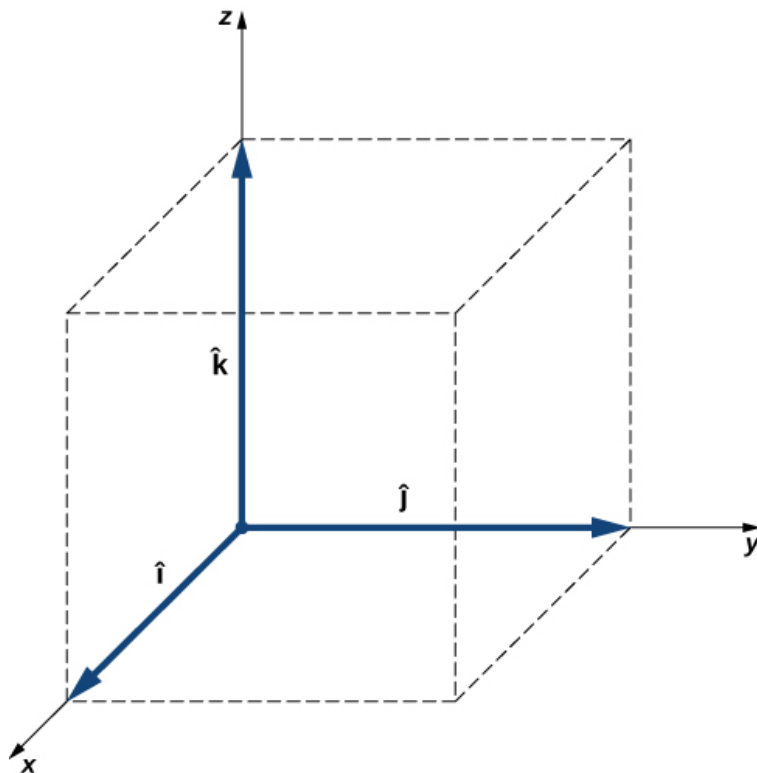
For the silver coin, the coordinates are

Equation:

$$\begin{cases} x_S = r_S \cos \varphi_S = (20.0 \text{ m}) \cos \pi/9 = 18.9 \text{ m} \\ y_S = r_S \sin \varphi_S = (20.0 \text{ m}) \sin \pi/9 = 6.8 \text{ m} \end{cases} \Rightarrow (x_S, y_S) = (18.9 \text{ m}, 6.8 \text{ m}).$$

Vectors in Three Dimensions

To specify the location of a point in space, we need three coordinates (x, y, z) , where coordinates x and y specify locations in a plane, and coordinate z gives a vertical position above or below the plane. Three-dimensional space has three orthogonal directions, so we need not two but *three* unit vectors to define a three-dimensional coordinate system. In the Cartesian coordinate system, the first two unit vectors are the unit vector of the x -axis $\hat{\mathbf{i}}$ and the unit vector of the y -axis $\hat{\mathbf{j}}$. The third unit vector $\hat{\mathbf{k}}$ is the direction of the z -axis ([\[link\]](#)). The order in which the axes are labeled, which is the order in which the three unit vectors appear, is important because it defines the orientation of the coordinate system. The order x - y - z , which is equivalent to the order $\hat{\mathbf{i}}$ - $\hat{\mathbf{j}}$ - $\hat{\mathbf{k}}$, defines the standard right-handed coordinate system (positive orientation).



Three unit vectors define a Cartesian system in three-dimensional space. The order in which these unit vectors appear defines the orientation of the coordinate system. The order shown here defines the right-handed orientation.

In three-dimensional space, vector $\vec{\mathbf{A}}$ has three vector components: the x-component $\vec{\mathbf{A}}_x = A_x \hat{\mathbf{i}}$, which is the part of vector $\vec{\mathbf{A}}$ along the x-axis; the y-component $\vec{\mathbf{A}}_y = A_y \hat{\mathbf{j}}$, which is the part of $\vec{\mathbf{A}}$ along the y-axis; and the z-component $\vec{\mathbf{A}}_z = A_z \hat{\mathbf{k}}$, which is the part of the vector along the z-axis. A vector in three-dimensional space is the vector sum of its three vector components ([\[link\]](#)):

Note:
Equation:

$$\vec{\mathbf{A}} = A_x \hat{\mathbf{i}} + A_y \hat{\mathbf{j}} + A_z \hat{\mathbf{k}}.$$

If we know the coordinates of its origin $b(x_b, y_b, z_b)$ and of its end $e(x_e, y_e, z_e)$, its scalar components are obtained by taking their differences: A_x and A_y are given by [\[link\]](#) and the z-component is given by

Note:
Equation:

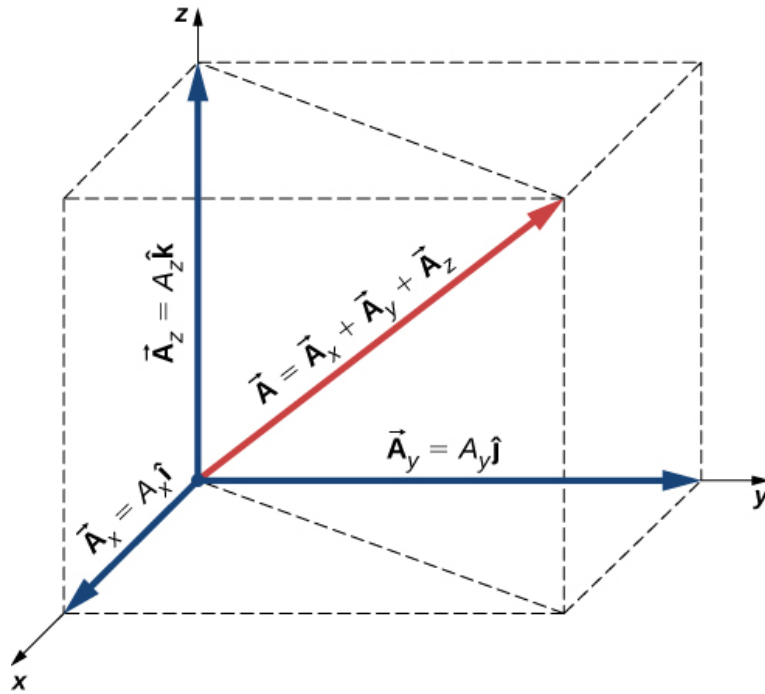
$$A_z = z_e - z_b.$$

Magnitude A is obtained by generalizing [\[link\]](#) to three dimensions:

Note:
Equation:

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2}.$$

This expression for the vector magnitude comes from applying the Pythagorean theorem twice. As seen in [\[link\]](#), the diagonal in the xy -plane has length $\sqrt{A_x^2 + A_y^2}$ and its square adds to the square A_z^2 to give A^2 . Note that when the z -component is zero, the vector lies entirely in the xy -plane and its description is reduced to two dimensions.



A vector in three-dimensional space is the vector sum of its three vector components.

Example:

Takeoff of a Drone

During a takeoff of IAI Heron ([\[link\]](#)), its position with respect to a control tower is 100 m above the ground, 300 m to the east, and 200 m to the north. One minute later, its position is 250 m above the ground, 1200 m to the east, and 2100 m to the north. What is the drone's displacement vector with respect to the control tower? What is the magnitude of its displacement vector?



The drone IAI Heron in flight. (credit: SSgt Reynaldo Ramon, USAF)

Strategy

We take the origin of the Cartesian coordinate system as the control tower. The direction of the +x-axis is given by unit vector $\hat{\mathbf{i}}$ to the east, the direction of the +y-axis is given by unit vector $\hat{\mathbf{j}}$ to the north, and the direction of the +z-axis is given by unit vector $\hat{\mathbf{k}}$, which points up from the ground. The drone's first position is the origin (or, equivalently, the beginning) of the displacement vector and its second position is the end of the displacement vector.

Solution

We identify $b(300.0 \text{ m}, 200.0 \text{ m}, 100.0 \text{ m})$ and $e(1200 \text{ m}, 2100 \text{ m}, 250 \text{ m})$, and use [\[link\]](#) and [\[link\]](#) to find the scalar components of the drone's displacement vector:

Equation:

$$\begin{cases} D_x = x_e - x_b = 1200.0 \text{ m} - 300.0 \text{ m} = 900.0 \text{ m}, \\ D_y = y_e - y_b = 2100.0 \text{ m} - 200.0 \text{ m} = 1900.0 \text{ m}, \\ D_z = z_e - z_b = 250.0 \text{ m} - 100.0 \text{ m} = 150.0 \text{ m}. \end{cases}$$

We substitute these components into [\[link\]](#) to find the displacement vector:

Equation:

$$\vec{D} = D_x \hat{\mathbf{i}} + D_y \hat{\mathbf{j}} + D_z \hat{\mathbf{k}} = 900.0 \text{ m} \hat{\mathbf{i}} + 1900.0 \text{ m} \hat{\mathbf{j}} + 150.0 \text{ m} \hat{\mathbf{k}} = (0.90 \hat{\mathbf{i}} + 1.90 \hat{\mathbf{j}} + 0.15 \hat{\mathbf{k}}) \text{ km}.$$

We substitute into [\[link\]](#) to find the magnitude of the displacement:

Equation:

$$D = \sqrt{D_x^2 + D_y^2 + D_z^2} = \sqrt{(0.90 \text{ km})^2 + (1.90 \text{ km})^2 + (0.15 \text{ km})^2} = 2.11 \text{ km}.$$

Note:

Exercise:

Problem:

Check Your Understanding If the average velocity vector of the drone in the displacement in [\[link\]](#) is $\vec{u} = (15.0\hat{i} + 31.7\hat{j} + 2.5\hat{k})\text{m/s}$, what is the magnitude of the drone's velocity vector?

Solution:

$$35.2 \text{ m/s} = 126.4 \text{ km/h}$$

Summary

- Vectors are described in terms of their components in a coordinate system. In two dimensions (in a plane), vectors have two components. In three dimensions (in space), vectors have three components.
- A vector component of a vector is its part in an axis direction. The vector component is the product of the unit vector of an axis with its scalar component along this axis. A vector is the resultant of its vector components.
- Scalar components of a vector are differences of coordinates, where coordinates of the origin are subtracted from end point coordinates of a vector. In a rectangular system, the magnitude of a vector is the square root of the sum of the squares of its components.
- In a plane, the direction of a vector is given by an angle the vector has with the positive x -axis. This direction angle is measured counterclockwise. The scalar x -component of a vector can be expressed as the product of its magnitude with the cosine of its direction angle, and the scalar y -component can be expressed as the product of its magnitude with the sine of its direction angle.
- In a plane, there are two equivalent coordinate systems. The Cartesian coordinate system is defined by unit vectors \hat{i} and \hat{j} along the x -axis and the y -axis, respectively. The polar coordinate system is defined by the radial unit vector \hat{r} , which gives the direction from the origin, and a unit vector \hat{t} , which is perpendicular (orthogonal) to the radial direction.

Conceptual Questions

Exercise:

Problem: Give an example of a nonzero vector that has a component of zero.

Solution:

a unit vector of the x -axis

Exercise:

Problem: Explain why a vector cannot have a component greater than its own magnitude.

Exercise:

Problem: If two vectors are equal, what can you say about their components?

Solution:

They are equal.

Exercise:

Problem:

If vectors \vec{A} and \vec{B} are orthogonal, what is the component of \vec{B} along the direction of \vec{A} ?
What is the component of \vec{A} along the direction of \vec{B} ?

Exercise:

Problem:

If one of the two components of a vector is not zero, can the magnitude of the other vector component of this vector be zero?

Solution:

yes

Exercise:

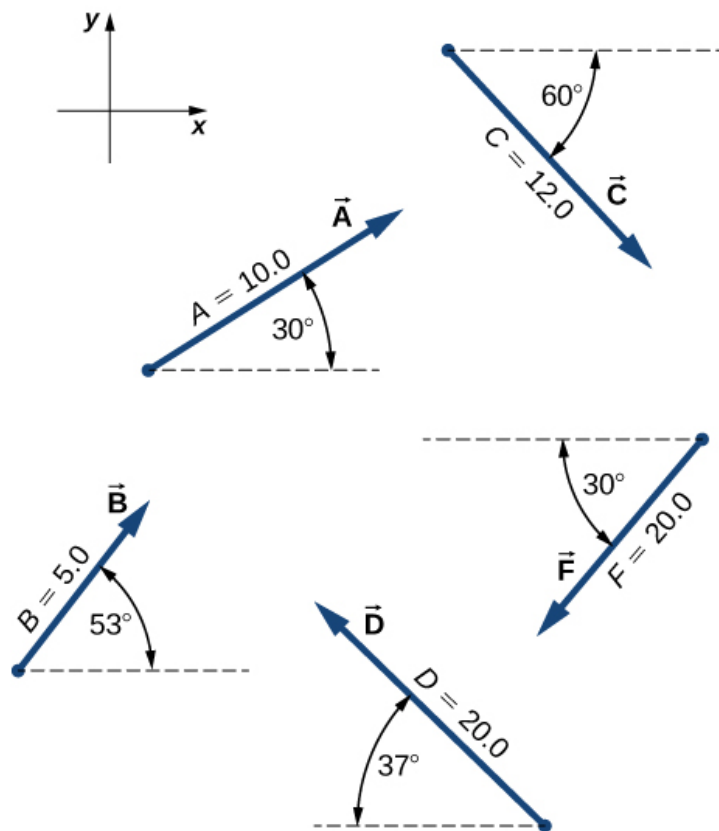
Problem: If two vectors have the same magnitude, do their components have to be the same?

Problems

Exercise:

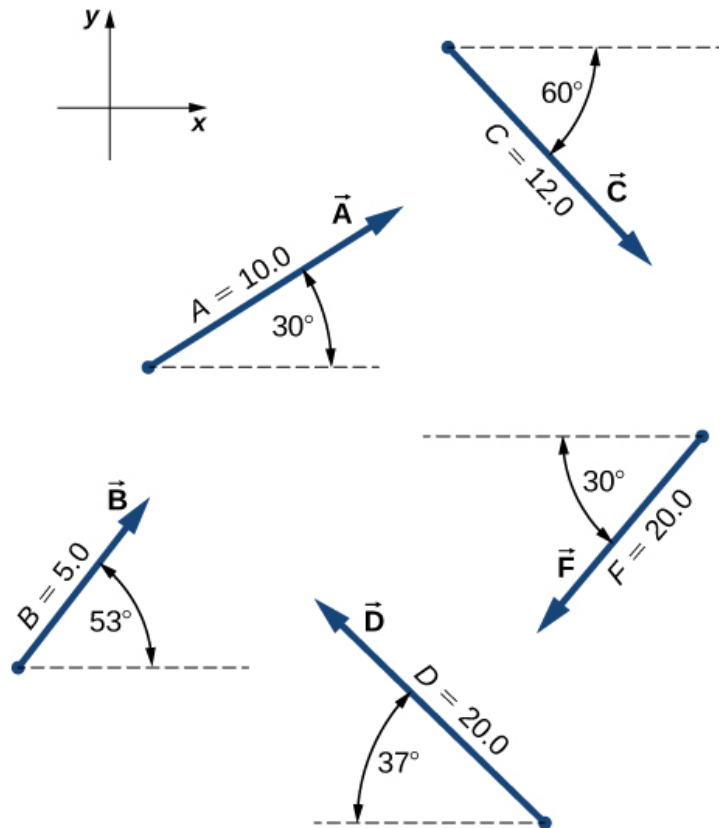
Problem:

Assuming the $+x$ -axis is horizontal and points to the right, resolve the vectors given in the following figure to their scalar components and express them in vector component form.



Solution:

a. $\vec{A} = +8.66\hat{i} + 5.00\hat{j}$, b. $\vec{B} = +3.01\hat{i} + 3.99\hat{j}$, c. $\vec{C} = +6.00\hat{i} - 10.39\hat{j}$, d.
 $\vec{D} = -15.97\hat{i} + 12.04\hat{j}$, f. $\vec{F} = -17.32\hat{i} - 10.00\hat{j}$



Exercise:

Problem:

Suppose you walk 18.0 m straight west and then 25.0 m straight north. How far are you from your starting point? What is your displacement vector? What is the direction of your displacement? Assume the +x-axis is to the east.

Exercise:

Problem:

You drive 7.50 km in a straight line in a direction 15° east of north. (a) Find the distances you would have to drive straight east and then straight north to arrive at the same point. (b) Show that you still arrive at the same point if the east and north legs are reversed in order. Assume the +x-axis is to the east.

Solution:

a. 1.94 km, 7.24 km; b. proof

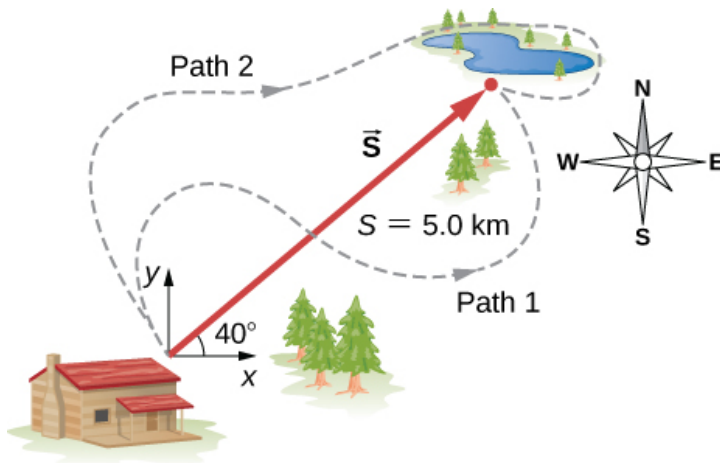
Exercise:

Problem:

A sledge is being pulled by two horses on a flat terrain. The net force on the sledge can be expressed in the Cartesian coordinate system as vector $\vec{F} = (-2980.0\hat{i} + 8200.0\hat{j})\text{N}$, where \hat{i} and \hat{j} denote directions to the east and north, respectively. Find the magnitude and direction of the pull.

Exercise:**Problem:**

A trapper walks a 5.0-km straight-line distance from her cabin to the lake, as shown in the following figure. Determine the east and north components of her displacement vector. How many more kilometers would she have to walk if she walked along the component displacements? What is her displacement vector?

**Solution:**

3.8 km east, 3.2 km north, 2.0 km, $\vec{D} = (3.8\hat{i} + 3.2\hat{j})\text{km}$

Exercise:**Problem:**

The polar coordinates of a point are $4\pi/3$ and 5.50 m. What are its Cartesian coordinates?

Exercise:**Problem:**

Two points in a plane have polar coordinates $P_1(2.500\text{ m}, \pi/6)$ and $P_2(3.800\text{ m}, 2\pi/3)$. Determine their Cartesian coordinates and the distance between them in the Cartesian coordinate system. Round the distance to a nearest centimeter.

Solution:

$P_1(2.165 \text{ m}, 1.250 \text{ m})$, $P_2(-1.900 \text{ m}, 3.290 \text{ m})$, 5.27 m

Exercise:

Problem:

A chameleon is resting quietly on a lanai screen, waiting for an insect to come by. Assume the origin of a Cartesian coordinate system at the lower left-hand corner of the screen and the horizontal direction to the right as the $+x$ -direction. If its coordinates are $(2.000 \text{ m}, 1.000 \text{ m})$, (a) how far is it from the corner of the screen? (b) What is its location in polar coordinates?

Exercise:

Problem:

Two points in the Cartesian plane are $A(2.00 \text{ m}, -4.00 \text{ m})$ and $B(-3.00 \text{ m}, 3.00 \text{ m})$. Find the distance between them and their polar coordinates.

Solution:

8.60 m , $A(2\sqrt{5} \text{ m}, 0.647\pi)$, $B(3\sqrt{2} \text{ m}, 0.75\pi)$

Exercise:

Problem:

A fly enters through an open window and zooms around the room. In a Cartesian coordinate system with three axes along three edges of the room, the fly changes its position from point $b(4.0 \text{ m}, 1.5 \text{ m}, 2.5 \text{ m})$ to point $e(1.0 \text{ m}, 4.5 \text{ m}, 0.5 \text{ m})$. Find the scalar components of the fly's displacement vector and express its displacement vector in vector component form. What is its magnitude?

Glossary

component form of a vector

a vector written as the vector sum of its components in terms of unit vectors

direction angle

in a plane, an angle between the positive direction of the x -axis and the vector, measured counterclockwise from the axis to the vector

polar coordinate system

an orthogonal coordinate system where location in a plane is given by polar coordinates

polar coordinates

a radial coordinate and an angle

radial coordinate

distance to the origin in a polar coordinate system

scalar component

a number that multiplies a unit vector in a vector component of a vector

unit vectors of the axes

unit vectors that define orthogonal directions in a plane or in space

vector components

orthogonal components of a vector; a vector is the vector sum of its vector components.

Algebra of Vectors

By the end of this section, you will be able to:

- Apply analytical methods of vector algebra to find resultant vectors and to solve vector equations for unknown vectors.
- Interpret physical situations in terms of vector expressions.

Vectors can be added together and multiplied by scalars. Vector addition is associative ([\[link\]](#)) and commutative ([\[link\]](#)), and vector multiplication by a sum of scalars is distributive ([\[link\]](#)). Also, scalar multiplication by a sum of vectors is distributive:

Note:

Equation:

$$\alpha(\vec{A} + \vec{B}) = \alpha\vec{A} + \alpha\vec{B}.$$

In this equation, α is any number (a scalar). For example, a vector antiparallel to vector $\vec{A} = A_x\hat{i} + A_y\hat{j} + A_z\hat{k}$ can be expressed simply by multiplying \vec{A} by the scalar $\alpha = -1$:

Note:

Equation:

$$-\vec{A} = -A_x\hat{i} - A_y\hat{j} - A_z\hat{k}.$$

Example:

Direction of Motion

In a Cartesian coordinate system where \hat{i} denotes geographic east, \hat{j} denotes geographic north, and \hat{k} denotes altitude above sea level, a military convoy advances its position through unknown territory with velocity $\vec{v} = (4.0\hat{i} + 3.0\hat{j} + 0.1\hat{k})\text{km/h}$. If the convoy had to retreat, in what geographic direction would it be moving?

Solution

The velocity vector has the third component $\vec{v}_z = (+0.1\text{km/h})\hat{k}$, which says the convoy is climbing at a rate of 100 m/h through mountainous terrain. At the same time, its velocity is 4.0 km/h to the east and 3.0 km/h to the north, so it moves on the ground in direction $\tan^{-1}(3/4) \approx 37^\circ$ north of east. If the convoy had to retreat, its new velocity vector \vec{u} would have to be antiparallel to \vec{v} and be in the form $\vec{u} = -\alpha\vec{v}$, where α is a positive number. Thus, the velocity of the retreat would be $\vec{u} = \alpha(-4.0\hat{i} - 3.0\hat{j} - 0.1\hat{k})\text{km/h}$. The negative sign of the third component indicates the convoy would be descending. The direction angle of the retreat velocity is $\tan^{-1}(-3\alpha/-4\alpha) \approx 37^\circ$ south of west. Therefore, the convoy would be moving on the ground in direction 37° south of west while descending on its way back.

The generalization of the number zero to vector algebra is called the **null vector**, denoted by $\vec{0}$. All components of the null vector are zero, $\vec{0} = 0\hat{i} + 0\hat{j} + 0\hat{k}$, so the null vector has no length and no direction.

Two vectors \vec{A} and \vec{B} are **equal vectors** if and only if their difference is the null vector:

Equation:

$$\vec{0} = \vec{A} - \vec{B} = (A_x\hat{i} + A_y\hat{j} + A_z\hat{k}) - (B_x\hat{i} + B_y\hat{j} + B_z\hat{k}) = (A_x - B_x)\hat{i} + (A_y - B_y)\hat{j} + (A_z - B_z)\hat{k}.$$

This vector equation means we must have simultaneously $A_x - B_x = 0$, $A_y - B_y = 0$, and $A_z - B_z = 0$. Hence, we can write $\vec{A} = \vec{B}$ if and only if the corresponding components of vectors \vec{A} and \vec{B} are equal:

Note:

Equation:

$$\vec{A} = \vec{B} \Leftrightarrow \begin{cases} A_x = B_x \\ A_y = B_y \\ A_z = B_z \end{cases}.$$

Two vectors are equal when their corresponding scalar components are equal.

Resolving vectors into their scalar components (i.e., finding their scalar components) and expressing them analytically in vector component form (given by [\[link\]](#)) allows us to use vector algebra to find sums or differences of many vectors *analytically* (i.e., without using graphical methods). For example, to find the resultant of two vectors \vec{A} and \vec{B} , we simply add them component by component, as follows:

Equation:

$$\vec{R} = \vec{A} + \vec{B} = (A_x\hat{i} + A_y\hat{j} + A_z\hat{k}) + (B_x\hat{i} + B_y\hat{j} + B_z\hat{k}) = (A_x + B_x)\hat{i} + (A_y + B_y)\hat{j} + (A_z + B_z)\hat{k}.$$

In this way, using [\[link\]](#), scalar components of the resultant vector $\vec{R} = R_x\hat{i} + R_y\hat{j} + R_z\hat{k}$ are the sums of corresponding scalar components of vectors \vec{A} and \vec{B} :

Equation:

$$\begin{cases} R_x = A_x + B_x, \\ R_y = A_y + B_y, \\ R_z = A_z + B_z. \end{cases}$$

Analytical methods can be used to find components of a resultant of many vectors. For example, if we are to sum up N vectors $\vec{F}_1, \vec{F}_2, \vec{F}_3, \dots, \vec{F}_N$, where each vector is $\vec{F}_k = F_{kx}\hat{i} + F_{ky}\hat{j} + F_{kz}\hat{k}$, the resultant vector \vec{F}_R is

Equation:

$$\begin{aligned}\vec{\mathbf{F}}_R &= \vec{\mathbf{F}}_1 + \vec{\mathbf{F}}_2 + \vec{\mathbf{F}}_3 + \dots + \vec{\mathbf{F}}_N = \sum_{k=1}^N \vec{\mathbf{F}}_k = \sum_{k=1}^N (F_{kx}\hat{\mathbf{i}} + F_{ky}\hat{\mathbf{j}} + F_{kz}\hat{\mathbf{k}}) \\ &= \left(\sum_{k=1}^N F_{kx}\right)\hat{\mathbf{i}} + \left(\sum_{k=1}^N F_{ky}\right)\hat{\mathbf{j}} + \left(\sum_{k=1}^N F_{kz}\right)\hat{\mathbf{k}}.\end{aligned}$$

Therefore, scalar components of the resultant vector are

Note:

Equation:

$$\begin{cases} F_{Rx} = \sum_{k=1}^N F_{kx} = F_{1x} + F_{2x} + \dots + F_{Nx} \\ F_{Ry} = \sum_{k=1}^N F_{ky} = F_{1y} + F_{2y} + \dots + F_{Ny} \\ F_{Rz} = \sum_{k=1}^N F_{kz} = F_{1z} + F_{2z} + \dots + F_{Nz}. \end{cases}$$

Having found the scalar components, we can write the resultant in vector component form:

Equation:

$$\vec{\mathbf{F}}_R = F_{Rx}\hat{\mathbf{i}} + F_{Ry}\hat{\mathbf{j}} + F_{Rz}\hat{\mathbf{k}}.$$

Analytical methods for finding the resultant and, in general, for solving vector equations are very important in physics because many physical quantities are vectors. For example, we use this method in kinematics to find resultant displacement vectors and resultant velocity vectors, in mechanics to find resultant force vectors and the resultants of many derived vector quantities, and in electricity and magnetism to find resultant electric or magnetic vector fields.

Example:

Analytical Computation of a Resultant

Three displacement vectors $\vec{\mathbf{A}}$, $\vec{\mathbf{B}}$, and $\vec{\mathbf{C}}$ in a plane ([link](#)) are specified by their magnitudes $A = 10.0$, $B = 7.0$, and $C = 8.0$, respectively, and by their respective direction angles with the horizontal direction $\alpha = 35^\circ$, $\beta = -110^\circ$, and $\gamma = 30^\circ$. The physical units of the magnitudes are centimeters. Resolve the vectors to their scalar components and find the following vector sums: (a) $\vec{\mathbf{R}} = \vec{\mathbf{A}} + \vec{\mathbf{B}} + \vec{\mathbf{C}}$, (b) $\vec{\mathbf{D}} = \vec{\mathbf{A}} - \vec{\mathbf{B}}$, and (c) $\vec{\mathbf{S}} = \vec{\mathbf{A}} - 3\vec{\mathbf{B}} + \vec{\mathbf{C}}$.

Strategy

First, we use ([link](#)) to find the scalar components of each vector and then we express each vector in its vector component form given by ([link](#)). Then, we use analytical methods of vector algebra to find the resultants.

Solution

We resolve the given vectors to their scalar components:

Equation:

$$\begin{cases} A_x = A \cos \alpha = (10.0 \text{ cm}) \cos 35^\circ = 8.19 \text{ cm} \\ A_y = A \sin \alpha = (10.0 \text{ cm}) \sin 35^\circ = 5.73 \text{ cm} \\ B_x = B \cos \beta = (7.0 \text{ cm}) \cos (-110^\circ) = -2.39 \text{ cm} \\ B_y = B \sin \beta = (7.0 \text{ cm}) \sin (-110^\circ) = -6.58 \text{ cm} \\ C_x = C \cos \gamma = (8.0 \text{ cm}) \cos 30^\circ = 6.93 \text{ cm} \\ C_y = C \sin \gamma = (8.0 \text{ cm}) \sin 30^\circ = 4.00 \text{ cm} \end{cases}$$

For (a) we may substitute directly into [\[link\]](#) to find the scalar components of the resultant:

Equation:

$$\begin{cases} R_x = A_x + B_x + C_x = 8.19 \text{ cm} - 2.39 \text{ cm} + 6.93 \text{ cm} = 12.73 \text{ cm} \\ R_y = A_y + B_y + C_y = 5.73 \text{ cm} - 6.58 \text{ cm} + 4.00 \text{ cm} = 3.15 \text{ cm} \end{cases}$$

Therefore, the resultant vector is $\vec{R} = R_x \hat{i} + R_y \hat{j} = (12.7\hat{i} + 3.1\hat{j})\text{cm}$.

For (b), we may want to write the vector difference as

Equation:

$$\vec{D} = \vec{A} - \vec{B} = (A_x \hat{i} + A_y \hat{j}) - (B_x \hat{i} + B_y \hat{j}) = (A_x - B_x) \hat{i} + (A_y - B_y) \hat{j}.$$

Then, the scalar components of the vector difference are

Equation:

$$\begin{cases} D_x = A_x - B_x = 8.19 \text{ cm} - (-2.39 \text{ cm}) = 10.58 \text{ cm} \\ D_y = A_y - B_y = 5.73 \text{ cm} - (-6.58 \text{ cm}) = 12.31 \text{ cm} \end{cases}$$

Hence, the difference vector is $\vec{D} = D_x \hat{i} + D_y \hat{j} = (10.6\hat{i} + 12.3\hat{j})\text{cm}$.

For (c), we can write vector \vec{S} in the following explicit form:

Equation:

$$\begin{aligned} \vec{S} &= \vec{A} - 3\vec{B} + \vec{C} = (A_x \hat{i} + A_y \hat{j}) - 3(B_x \hat{i} + B_y \hat{j}) + (C_x \hat{i} + C_y \hat{j}) \\ &= (A_x - 3B_x + C_x) \hat{i} + (A_y - 3B_y + C_y) \hat{j}. \end{aligned}$$

Then, the scalar components of \vec{S} are

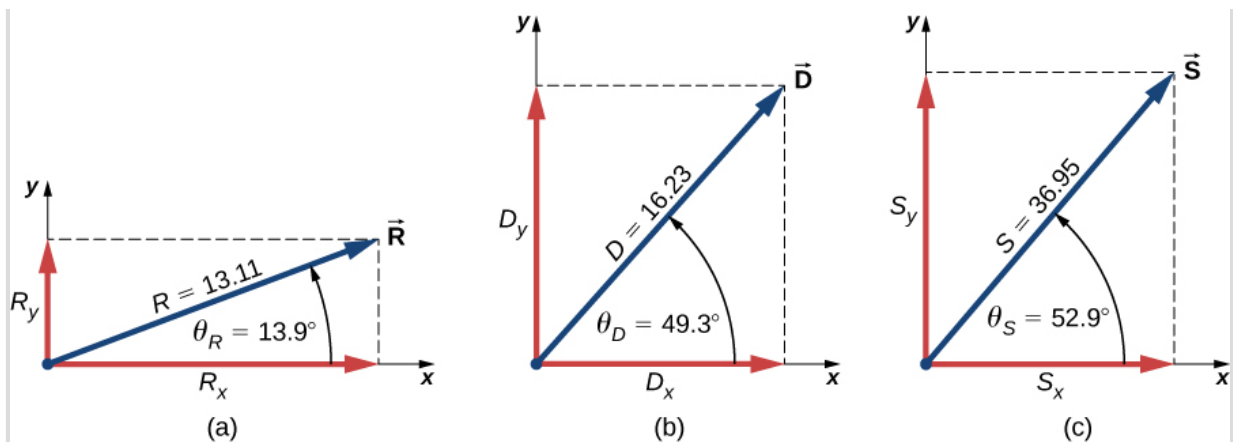
Equation:

$$\begin{cases} S_x = A_x - 3B_x + C_x = 8.19 \text{ cm} - 3(-2.39 \text{ cm}) + 6.93 \text{ cm} = 22.29 \text{ cm} \\ S_y = A_y - 3B_y + C_y = 5.73 \text{ cm} - 3(-6.58 \text{ cm}) + 4.00 \text{ cm} = 29.47 \text{ cm} \end{cases}$$

The vector is $\vec{S} = S_x \hat{i} + S_y \hat{j} = (22.3\hat{i} + 29.5\hat{j})\text{cm}$.

Significance

Having found the vector components, we can illustrate the vectors by graphing or we can compute magnitudes and direction angles, as shown in [\[link\]](#). Results for the magnitudes in (b) and (c) can be compared with results for the same problems obtained with the graphical method, shown in [\[link\]](#) and [\[link\]](#). Notice that the analytical method produces exact results and its accuracy is not limited by the resolution of a ruler or a protractor, as it was with the graphical method used in [\[link\]](#) for finding this same resultant.



Graphical illustration of the solutions obtained analytically in [\[link\]](#).

Note:

Exercise:

Problem:

Check Your Understanding Three displacement vectors \vec{A} , \vec{B} , and \vec{F} ([\[link\]](#)) are specified by their magnitudes $A = 10.00$, $B = 7.00$, and $F = 20.00$, respectively, and by their respective direction angles with the horizontal direction $\alpha = 35^\circ$, $\beta = -110^\circ$, and $\varphi = 110^\circ$. The physical units of the magnitudes are centimeters. Use the analytical method to find vector $\vec{G} = \vec{A} + 2\vec{B} - \vec{F}$. Verify that $G = 28.15$ cm and that $\theta_G = -68.65^\circ$.

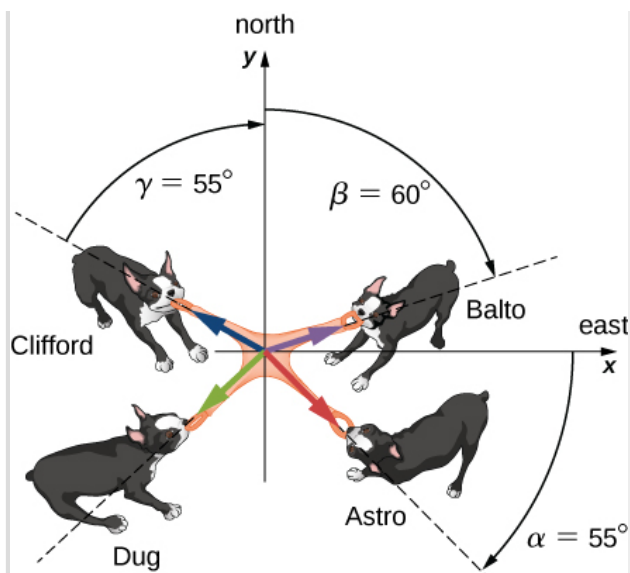
Solution:

$$\vec{G} = (10.25\hat{i} - 26.22\hat{j})\text{cm}$$

Example:

The Tug-of-War Game

Four dogs named Astro, Balto, Clifford, and Dug play a tug-of-war game with a toy ([\[link\]](#)). Astro pulls on the toy in direction $\alpha = 55^\circ$ south of east, Balto pulls in direction $\beta = 60^\circ$ east of north, and Clifford pulls in direction $\gamma = 55^\circ$ west of north. Astro pulls strongly with 160.0 units of force (N), which we abbreviate as $A = 160.0$ N. Balto pulls even stronger than Astro with a force of magnitude $B = 200.0$ N, and Clifford pulls with a force of magnitude $C = 140.0$ N. When Dug pulls on the toy in such a way that his force balances out the resultant of the other three forces, the toy does not move in any direction. With how big a force and in what direction must Dug pull on the toy for this to happen?



Four dogs play a tug-of-war game with a toy.

Strategy

We assume that east is the direction of the positive x -axis and north is the direction of the positive y -axis. As in [\[link\]](#), we have to resolve the three given forces— \vec{A} (the pull from Astro), \vec{B} (the pull from Balto), and \vec{C} (the pull from Clifford)—into their scalar components and then find the scalar components of the resultant vector $\vec{R} = \vec{A} + \vec{B} + \vec{C}$. When the pulling force \vec{D} from Dug balances out this resultant, the sum of \vec{D} and \vec{R} must give the null vector $\vec{D} + \vec{R} = \vec{0}$. This means that $\vec{D} = -\vec{R}$, so the pull from Dug must be antiparallel to \vec{R} .

Solution

The direction angles are $\theta_A = -\alpha = -55^\circ$, $\theta_B = 90^\circ - \beta = 30^\circ$, and $\theta_C = 90^\circ + \gamma = 145^\circ$, and substituting them into [\[link\]](#) gives the scalar components of the three given forces:

Equation:

$$\begin{cases} A_x = A \cos \theta_A = (160.0 \text{ N}) \cos (-55^\circ) = +91.8 \text{ N} \\ A_y = A \sin \theta_A = (160.0 \text{ N}) \sin (-55^\circ) = -131.1 \text{ N} \\ B_x = B \cos \theta_B = (200.0 \text{ N}) \cos 30^\circ = +173.2 \text{ N} \\ B_y = B \sin \theta_B = (200.0 \text{ N}) \sin 30^\circ = +100.0 \text{ N} \\ C_x = C \cos \theta_C = (140.0 \text{ N}) \cos 145^\circ = -114.7 \text{ N} \\ C_y = C \sin \theta_C = (140.0 \text{ N}) \sin 145^\circ = +80.3 \text{ N} \end{cases}.$$

Now we compute scalar components of the resultant vector $\vec{R} = \vec{A} + \vec{B} + \vec{C}$:

Equation:

$$\begin{cases} R_x = A_x + B_x + C_x = +91.8 \text{ N} + 173.2 \text{ N} - 114.7 \text{ N} = +150.3 \text{ N} \\ R_y = A_y + B_y + C_y = -131.1 \text{ N} + 100.0 \text{ N} + 80.3 \text{ N} = +49.2 \text{ N} \end{cases}.$$

The antiparallel vector to the resultant \vec{R} is

Equation:

$$\vec{D} = -\vec{R} = -R_x\hat{i} - R_y\hat{j} = (-150.3\hat{i} - 49.2\hat{j}) \text{ N}.$$

The magnitude of Dug's pulling force is

Equation:

$$D = \sqrt{D_x^2 + D_y^2} = \sqrt{(-150.3)^2 + (-49.2)^2} \text{ N} = 158.1 \text{ N}.$$

The direction of Dug's pulling force is

Equation:

$$\theta = \tan^{-1} \left(\frac{D_y}{D_x} \right) = \tan^{-1} \left(\frac{-49.2 \text{ N}}{-150.3 \text{ N}} \right) = \tan^{-1} \left(\frac{49.2}{150.3} \right) = 18.1^\circ.$$

Dug pulls in the direction 18.1° south of west because both components are negative, which means the pull vector lies in the third quadrant ([link](#)).

Note:

Exercise:

Problem:

Check Your Understanding Suppose that Balto in [link](#) leaves the game to attend to more important matters, but Astro, Clifford, and Dug continue playing. Astro and Clifford's pull on the toy does not change, but Dug runs around and bites on the toy in a different place. With how big a force and in what direction must Dug pull on the toy now to balance out the combined pulls from Clifford and Astro? Illustrate this situation by drawing a vector diagram indicating all forces involved.

Solution:

$D = 55.7 \text{ N}$; direction 65.7° north of east

Example:

Vector Algebra

Find the magnitude of the vector \vec{C} that satisfies the equation $2\vec{A} - 6\vec{B} + 3\vec{C} = 2\hat{j}$, where $\vec{A} = \hat{i} - 2\hat{k}$ and $\vec{B} = -\hat{j} + \hat{k}/2$.

Strategy

We first solve the given equation for the unknown vector \vec{C} . Then we substitute \vec{A} and \vec{B} ; group the terms along each of the three directions \hat{i} , \hat{j} , and \hat{k} ; and identify the scalar components C_x , C_y , and C_z . Finally, we substitute into [link](#) to find magnitude C .

Solution

Equation:

$$\begin{aligned}
2\vec{A} - 6\vec{B} + 3\vec{C} &= 2\hat{j} \\
3\vec{C} &= 2\hat{j} - 2\vec{A} + 6\vec{B} \\
\vec{C} &= \frac{2}{3}\hat{j} - \frac{2}{3}\vec{A} + 2\vec{B} \\
&= \frac{2}{3}\hat{j} - \frac{2}{3}(\hat{i} - 2\hat{k}) + 2\left(-\hat{j} + \frac{\hat{k}}{2}\right) = \frac{2}{3}\hat{j} - \frac{2}{3}\hat{i} + \frac{4}{3}\hat{k} - 2\hat{j} + \hat{k} \\
&= -\frac{2}{3}\hat{i} + \left(\frac{2}{3} - 2\right)\hat{j} + \left(\frac{4}{3} + 1\right)\hat{k} \\
&= -\frac{2}{3}\hat{i} - \frac{4}{3}\hat{j} + \frac{7}{3}\hat{k}.
\end{aligned}$$

The components are $C_x = -2/3$, $C_y = -4/3$, and $C_z = 7/3$, and substituting into [link](#) gives
Equation:

$$C = \sqrt{C_x^2 + C_y^2 + C_z^2} = \sqrt{(-2/3)^2 + (-4/3)^2 + (7/3)^2} = \sqrt{23/3}.$$

Example:

Displacement of a Skier

Starting at a ski lodge, a cross-country skier goes 5.0 km north, then 3.0 km west, and finally 4.0 km southwest before taking a rest. Find his total displacement vector relative to the lodge when he is at the rest point. How far and in what direction must he ski from the rest point to return directly to the lodge?

Strategy

We assume a rectangular coordinate system with the origin at the ski lodge and with the unit vector \hat{i} pointing east and the unit vector \hat{j} pointing north. There are three displacements: \vec{D}_1 , \vec{D}_2 , and \vec{D}_3 . We identify their magnitudes as $D_1 = 5.0$ km, $D_2 = 3.0$ km, and $D_3 = 4.0$ km. We identify their directions as the angles $\theta_1 = 90^\circ$, $\theta_2 = 180^\circ$, and $\theta_3 = 180^\circ + 45^\circ = 225^\circ$. We resolve each displacement vector to its scalar components and substitute the components into [link](#) to obtain the scalar components of the resultant displacement \vec{D} from the lodge to the rest point. On the way back from the rest point to the lodge, the displacement is $\vec{B} = -\vec{D}$. Finally, we find the magnitude and direction of \vec{B} .

Solution

Scalar components of the displacement vectors are

Equation:

$$\begin{cases} D_{1x} = D_1 \cos \theta_1 = (5.0 \text{ km}) \cos 90^\circ = 0 \\ D_{1y} = D_1 \sin \theta_1 = (5.0 \text{ km}) \sin 90^\circ = 5.0 \text{ km} \\ D_{2x} = D_2 \cos \theta_2 = (3.0 \text{ km}) \cos 180^\circ = -3.0 \text{ km} \\ D_{2y} = D_2 \sin \theta_2 = (3.0 \text{ km}) \sin 180^\circ = 0 \\ D_{3x} = D_3 \cos \theta_3 = (4.0 \text{ km}) \cos 225^\circ = -2.8 \text{ km} \\ D_{3y} = D_3 \sin \theta_3 = (4.0 \text{ km}) \sin 225^\circ = -2.8 \text{ km} \end{cases}.$$

Scalar components of the net displacement vector are

Equation:

$$\begin{cases} D_x = D_{1x} + D_{2x} + D_{3x} = (0 - 3.0 - 2.8)\text{km} = -5.8 \text{ km} \\ D_y = D_{1y} + D_{2y} + D_{3y} = (5.0 + 0 - 2.8)\text{km} = +2.2 \text{ km} \end{cases}.$$

Hence, the skier's net displacement vector is $\vec{D} = D_x\hat{i} + D_y\hat{j} = (-5.8\hat{i} + 2.2\hat{j})\text{km}$. On the way back to the lodge, his displacement is $\vec{B} = -\vec{D} = -(-5.8\hat{i} + 2.2\hat{j})\text{km} = (5.8\hat{i} - 2.2\hat{j})\text{km}$. Its magnitude is $B = \sqrt{B_x^2 + B_y^2} = \sqrt{(5.8)^2 + (-2.2)^2}\text{km} = 6.2\text{km}$ and its direction angle is $\theta = \tan^{-1}(-2.2/5.8) = -20.8^\circ$. Therefore, to return to the lodge, he must go 6.2 km in a direction about 21° south of east.

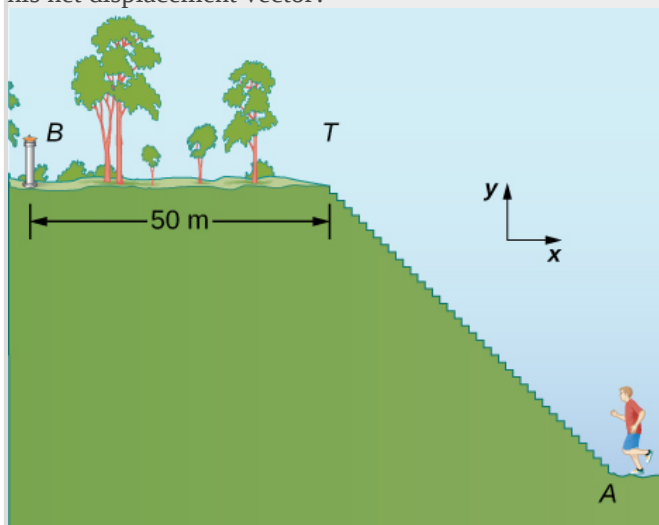
Significance

Notice that no figure is needed to solve this problem by the analytical method. Figures are required when using a graphical method; however, we can check if our solution makes sense by sketching it, which is a useful final step in solving any vector problem.

Example:

Displacement of a Jogger

A jogger runs up a flight of 200 identical steps to the top of a hill and then runs along the top of the hill 50.0 m before he stops at a drinking fountain ([link](#)). His displacement vector from point A at the bottom of the steps to point B at the fountain is $\vec{D}_{AB} = (-90.0\hat{i} + 30.0\hat{j})\text{m}$. What is the height and width of each step in the flight? What is the actual distance the jogger covers? If he makes a loop and returns to point A, what is his net displacement vector?



A jogger runs up a flight of steps.

Strategy

The displacement vector \vec{D}_{AB} is the vector sum of the jogger's displacement vector \vec{D}_{AT} along the stairs (from point A at the bottom of the stairs to point T at the top of the stairs) and his displacement vector \vec{D}_{TB} on the top of the hill (from point A at the top of the stairs to the fountain at point T). We must find the horizontal and the vertical components of \vec{D}_{AT} . If each step has width w and height h , the horizontal component of \vec{D}_{AT} must have a length of $200w$ and the vertical component must have a length of $200h$. The actual distance the jogger covers is the sum of the distance he runs up the stairs and the distance of 50.0 m that he runs along the top of the hill.

Solution

In the coordinate system indicated in [\[link\]](#), the jogger's displacement vector on the top of the hill is $\vec{D}_{TB} = (-50.0 \text{ m})\hat{i}$. His net displacement vector is

Equation:

$$\vec{D}_{AB} = \vec{D}_{AT} + \vec{D}_{TB}.$$

Therefore, his displacement vector \vec{D}_{TB} along the stairs is

Equation:

$$\begin{aligned}\vec{D}_{AT} &= \vec{D}_{AB} - \vec{D}_{TB} = (-90.0\hat{i} + 30.0\hat{j})\text{m} - (-50.0\text{m})\hat{i} = [(-90.0 + 50.0)\hat{i} + 30.0\hat{j}]\text{m} \\ &= (-40.0\hat{i} + 30.0\hat{j})\text{m}.\end{aligned}$$

Its scalar components are $D_{ATx} = -40.0 \text{ m}$ and $D_{ATy} = 30.0 \text{ m}$. Therefore, we must have

Equation:

$$200w = |-40.0|\text{m} \text{ and } 200h = 30.0 \text{ m}.$$

Hence, the step width is $w = 40.0 \text{ m}/200 = 0.2 \text{ m} = 20 \text{ cm}$, and the step height is $h = 30.0 \text{ m}/200 = 0.15 \text{ m} = 15 \text{ cm}$. The distance that the jogger covers along the stairs is

Equation:

$$D_{AT} = \sqrt{D_{ATx}^2 + D_{ATy}^2} = \sqrt{(-40.0)^2 + (30.0)^2} \text{ m} = 50.0 \text{ m}.$$

Thus, the actual distance he runs is $D_{AT} + D_{TB} = 50.0 \text{ m} + 50.0 \text{ m} = 100.0 \text{ m}$. When he makes a loop and comes back from the fountain to his initial position at point A, the total distance he covers is twice this distance, or 200.0 m. However, his net displacement vector is zero, because when his final position is the same as his initial position, the scalar components of his net displacement vector are zero ([\[link\]](#)).

In many physical situations, we often need to know the direction of a vector. For example, we may want to know the direction of a magnetic field vector at some point or the direction of motion of an object. We have already said direction is given by a unit vector, which is a dimensionless entity—that is, it has no physical units associated with it. When the vector in question lies along one of the axes in a Cartesian system of coordinates, the answer is simple, because then its unit vector of direction is either parallel or antiparallel to the direction of the unit vector of an axis. For example, the direction of vector $\vec{d} = -5 \text{ m}\hat{i}$ is unit vector $\hat{d} = -\hat{i}$. The general rule of finding the unit vector \hat{V} of direction for any vector \vec{V} is to divide it by its magnitude V :

Note:

Equation:

$$\hat{V} = \frac{\vec{V}}{V}.$$

We see from this expression that the unit vector of direction is indeed dimensionless because the numerator and the denominator in [\[link\]](#) have the same physical unit. In this way, [\[link\]](#) allows us to express the unit

vector of direction in terms of unit vectors of the axes. The following example illustrates this principle.

Example:

The Unit Vector of Direction

If the velocity vector of the military convoy in [\[link\]](#) is $\vec{v} = (4.000\hat{i} + 3.000\hat{j} + 0.100\hat{k})\text{km/h}$, what is the unit vector of its direction of motion?

Strategy

The unit vector of the convoy's direction of motion is the unit vector \hat{v} that is parallel to the velocity vector. The unit vector is obtained by dividing a vector by its magnitude, in accordance with [\[link\]](#).

Solution

The magnitude of the vector \vec{v} is

Equation:

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2} = \sqrt{4.000^2 + 3.000^2 + 0.100^2}\text{km/h} = 5.001\text{km/h}.$$

To obtain the unit vector \hat{v} , divide \vec{v} by its magnitude:

Equation:

$$\begin{aligned}\hat{v} &= \frac{\vec{v}}{v} = \frac{(4.000\hat{i} + 3.000\hat{j} + 0.100\hat{k})\text{km/h}}{5.001\text{km/h}} \\ &= \frac{(4.000\hat{i} + 3.000\hat{j} + 0.100\hat{k})}{5.001} \\ &= \frac{4.000}{5.001}\hat{i} + \frac{3.000}{5.001}\hat{j} + \frac{0.100}{5.001}\hat{k} \\ &= (79.98\hat{i} + 59.99\hat{j} + 2.00\hat{k}) \times 10^{-2}.\end{aligned}$$

Significance

Note that when using the analytical method with a calculator, it is advisable to carry out your calculations to at least three decimal places and then round off the final answer to the required number of significant figures, which is the way we performed calculations in this example. If you round off your partial answer too early, you risk your final answer having a huge numerical error, and it may be far off from the exact answer or from a value measured in an experiment.

Note:

Exercise:

Problem:

Check Your Understanding Verify that vector \hat{v} obtained in [\[link\]](#) is indeed a unit vector by computing its magnitude. If the convoy in [\[link\]](#) was moving across a desert flatland—that is, if the third component of its velocity was zero—what is the unit vector of its direction of motion? Which geographic direction does it represent?

Solution:

$$\hat{v} = 0.8\hat{i} + 0.6\hat{j}, 36.87^\circ \text{ north of east}$$

Summary

- Analytical methods of vector algebra allow us to find resultants of sums or differences of vectors without having to draw them. Analytical methods of vector addition are exact, contrary to graphical methods, which are approximate.
- Analytical methods of vector algebra are used routinely in mechanics, electricity, and magnetism. They are important mathematical tools of physics.

Problems

Exercise:

Problem:

For vectors $\vec{B} = -\hat{i} - 4\hat{j}$ and $\vec{A} = -3\hat{i} - 2\hat{j}$, calculate (a) $\vec{A} + \vec{B}$ and its magnitude and direction angle, and (b) $\vec{A} - \vec{B}$ and its magnitude and direction angle.

Solution:

$$\text{a. } \vec{A} + \vec{B} = -4\hat{i} - 6\hat{j}, \left| \vec{A} + \vec{B} \right| = 7.211, \theta = 236^\circ; \text{ b. } \vec{A} - \vec{B} = -2\hat{i} + 2\hat{j}, \left| \vec{A} - \vec{B} \right| = 2\sqrt{2}, \theta = 135^\circ$$

Exercise:

Problem:

A particle undergoes three consecutive displacements given by vectors

$$\vec{D}_1 = (3.0\hat{i} - 4.0\hat{j} - 2.0\hat{k})\text{mm}, \vec{D}_2 = (1.0\hat{i} - 7.0\hat{j} + 4.0\hat{k})\text{mm}, \text{ and}$$

$\vec{D}_3 = (-7.0\hat{i} + 4.0\hat{j} + 1.0\hat{k})\text{mm}$. (a) Find the resultant displacement vector of the particle. (b) What is the magnitude of the resultant displacement? (c) If all displacements were along one line, how far would the particle travel?

Exercise:

Problem:

Given two displacement vectors $\vec{A} = (3.00\hat{i} - 4.00\hat{j} + 4.00\hat{k})\text{m}$ and

$\vec{B} = (2.00\hat{i} + 3.00\hat{j} - 7.00\hat{k})\text{m}$, find the displacements and their magnitudes for (a) $\vec{C} = \vec{A} + \vec{B}$ and (b) $\vec{D} = 2\vec{A} - \vec{B}$.

Solution:

$$\text{a. } \vec{C} = (5.0\hat{i} - 1.0\hat{j} - 3.0\hat{k})\text{m}, C = 5.92 \text{ m};$$

$$\text{b. } \vec{D} = (4.0\hat{i} - 11.0\hat{j} + 15.0\hat{k})\text{m}, D = 19.03 \text{ m}$$

Exercise:

Problem:

A small plane flies 40.0 km in a direction 60° north of east and then flies 30.0 km in a direction 15° north of east. Use the analytical method to find the total distance the plane covers from the starting point, and the geographic direction of its displacement vector. What is its displacement vector?

Exercise:

Problem:

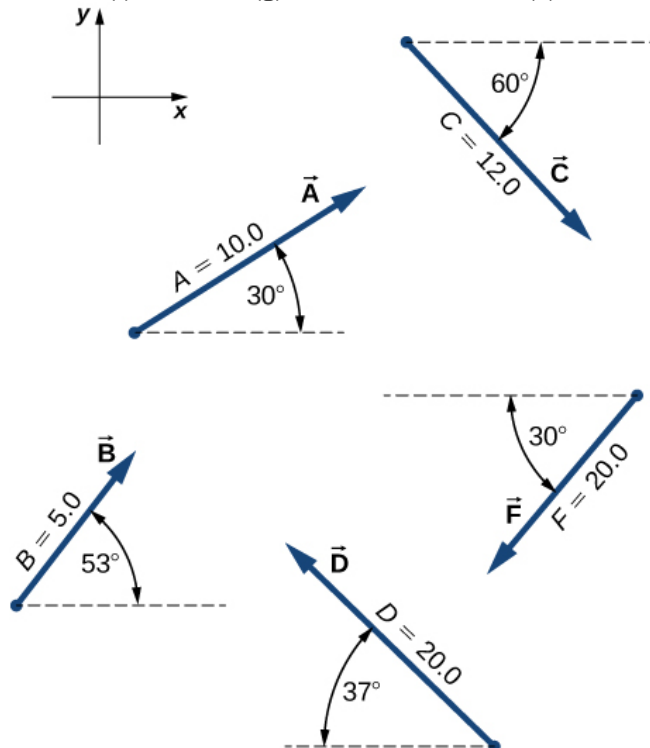
In an attempt to escape a desert island, a castaway builds a raft and sets out to sea. The wind shifts a great deal during the day, and she is blown along the following straight lines: 2.50 km and 45.0° north of west, then 4.70 km and 60.0° south of east, then 1.30 km and 25.0° south of west, then 5.10 km due east, then 1.70 km and 5.00° east of north, then 7.20 km and 55.0° south of west, and finally 2.80 km and 10.0° north of east. Use the analytical method to find the resultant vector of all her displacement vectors. What is its magnitude and direction?

Solution:

$$\vec{D} = (3.3\hat{i} - 6.6\hat{j})\text{km}, \hat{i} \text{ is to the east, } 7.34 \text{ km, } -63.5^\circ$$

Exercise:**Problem:**

Assuming the $+x$ -axis is horizontal to the right for the vectors given in the following figure, use the analytical method to find the following resultants: (a) $\vec{A} + \vec{B}$, (b) $\vec{C} + \vec{B}$, (c) $\vec{D} + \vec{F}$, (d) $\vec{A} - \vec{B}$, (e) $\vec{D} - \vec{F}$, (f) $\vec{A} + 2\vec{F}$, (g) $\vec{C} - 2\vec{D} + 3\vec{F}$, and (h) $\vec{A} - 4\vec{D} + 2\vec{F}$.

**Exercise:****Problem:**

Given the vectors in the preceding figure, find vector \vec{R} that solves equations (a) $\vec{D} + \vec{R} = \vec{F}$ and (b) $\vec{C} - 2\vec{D} + 5\vec{R} = 3\vec{F}$. Assume the $+x$ -axis is horizontal to the right.

Solution:

a. $\vec{R} = -1.35\hat{i} - 22.04\hat{j}$, b. $\vec{R} = -17.98\hat{i} + 0.89\hat{j}$

Exercise:

Problem:

A delivery man starts at the post office, drives 40 km north, then 20 km west, then 60 km northeast, and finally 50 km north to stop for lunch. Use the analytical method to determine the following: (a) Find his net displacement vector. (b) How far is the restaurant from the post office? (c) If he returns directly from the restaurant to the post office, what is his displacement vector on the return trip? (d) What is his compass heading on the return trip? Assume the +x-axis is to the east.

Exercise:

Problem:

An adventurous dog strays from home, runs three blocks east, two blocks north, and one block east, one block north, and two blocks west. Assuming that each block is about a 100 yd, use the analytical method to find the dog's net displacement vector, its magnitude, and its direction. Assume the +x-axis is to the east. How would your answer be affected if each block was about 100 m?

Solution:

$\vec{D} = (200\hat{i} + 300\hat{j})\text{yd}$, $D = 360.5\text{ yd}$, 56.3° north of east; The numerical answers would stay the same but the physical unit would be meters. The physical meaning and distances would be about the same because 1 yd is comparable with 1 m.

Exercise:

Problem:

If $\vec{D} = (6.00\hat{i} - 8.00\hat{j})\text{m}$, $\vec{B} = (-8.00\hat{i} + 3.00\hat{j})\text{m}$, and $\vec{A} = (26.0\hat{i} + 19.0\hat{j})\text{m}$, find the unknown constants a and b such that $a\vec{D} + b\vec{B} + \vec{A} = \vec{0}$.

Exercise:

Problem:

Given the displacement vector $\vec{D} = (3\hat{i} - 4\hat{j})\text{m}$, find the displacement vector \vec{R} so that $\vec{D} + \vec{R} = -4D\hat{j}$.

Solution:

$$\vec{R} = -3\hat{i} - 16\hat{j}$$

Exercise:

Problem:

Find the unit vector of direction for the following vector quantities: (a) Force $\vec{F} = (3.0\hat{i} - 2.0\hat{j})\text{N}$, (b) displacement $\vec{D} = (-3.0\hat{i} - 4.0\hat{j})\text{m}$, and (c) velocity $\vec{v} = (-5.00\hat{i} + 4.00\hat{j})\text{m/s}$.

Exercise:

Problem:

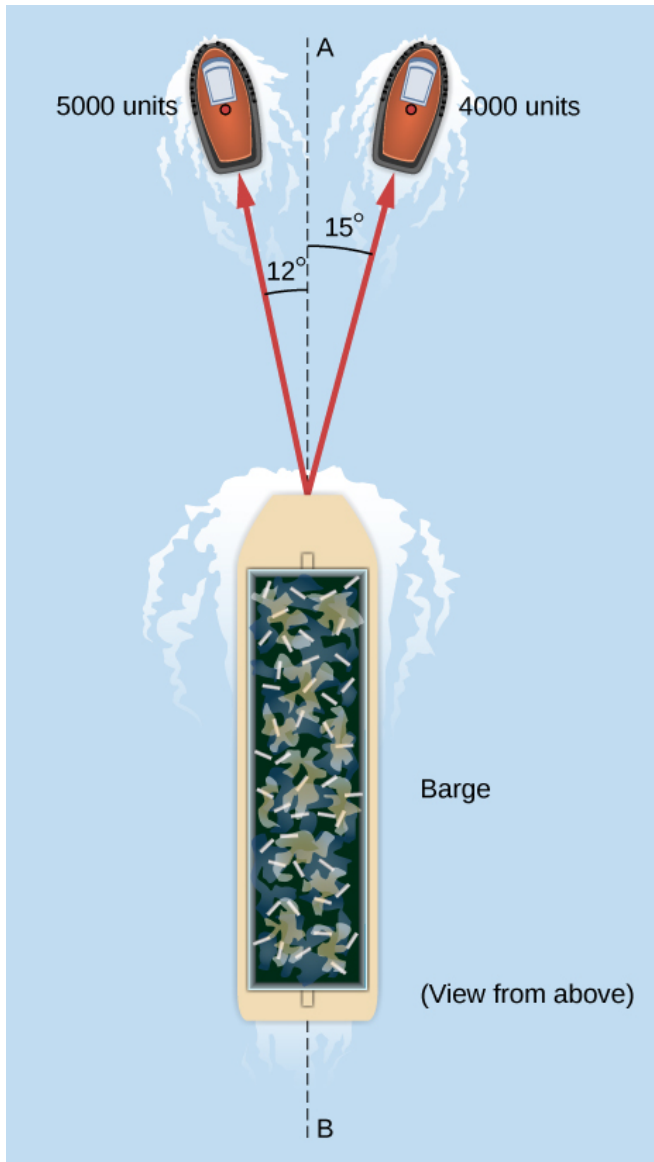
At one point in space, the direction of the electric field vector is given in the Cartesian system by the unit vector $\hat{\mathbf{E}} = 1/\sqrt{5}\hat{\mathbf{i}} - 2/\sqrt{5}\hat{\mathbf{j}}$. If the magnitude of the electric field vector is $E = 400.0 \text{ V/m}$, what are the scalar components E_x , E_y , and E_z of the electric field vector \vec{E} at this point? What is the direction angle θ_E of the electric field vector at this point?

Solution:

$$\vec{E} = E\hat{\mathbf{E}}, E_x = +178.9 \text{ V/m}, E_y = -357.8 \text{ V/m}, E_z = 0.0 \text{ V/m}, \theta_E = -\tan^{-1}(2)$$

Exercise:**Problem:**

A barge is pulled by the two tugboats shown in the following figure. One tugboat pulls on the barge with a force of magnitude 4000 units of force at 15° above the line AB (see the figure and the other tugboat pulls on the barge with a force of magnitude 5000 units of force at 12° below the line AB. Resolve the pulling forces to their scalar components and find the components of the resultant force pulling on the barge. What is the magnitude of the resultant pull? What is its direction relative to the line AB?



Exercise:

Problem:

In the control tower at a regional airport, an air traffic controller monitors two aircraft as their positions change with respect to the control tower. One plane is a cargo carrier Boeing 747 and the other plane is a Douglas DC-3. The Boeing is at an altitude of 2500 m, climbing at 10° above the horizontal, and moving 30° north of west. The DC-3 is at an altitude of 3000 m, climbing at 5° above the horizontal, and cruising directly west. (a) Find the position vectors of the planes relative to the control tower. (b) What is the distance between the planes at the moment the air traffic controller makes a note about their positions?

Solution:

a. $-34.290\hat{\mathbf{i}} + 7.089\hat{\mathbf{j}} + 2.500\hat{\mathbf{k}}$ km, $\vec{\mathbf{R}}_D = (-34.290\hat{\mathbf{i}} + 3.000\hat{\mathbf{k}})$ km; b.
 $|\vec{\mathbf{R}}_B - \vec{\mathbf{R}}_D| = 23.131$ km

Glossary

equal vectors

two vectors are equal if and only if all their corresponding components are equal; alternately, two parallel vectors of equal magnitudes

null vector

a vector with all its components equal to zero

Products of Vectors

By the end of this section, you will be able to:

- Explain the difference between the scalar product and the vector product of two vectors.
- Determine the scalar product of two vectors.
- Determine the vector product of two vectors.
- Describe how the products of vectors are used in physics.

A vector can be multiplied by another vector but may not be divided by another vector. There are two kinds of products of vectors used broadly in physics and engineering. One kind of multiplication is a *scalar multiplication of two vectors*. Taking a scalar product of two vectors results in a number (a scalar), as its name indicates. Scalar products are used to define work and energy relations. For example, the work that a force (a vector) performs on an object while causing its displacement (a vector) is defined as a scalar product of the force vector with the displacement vector. A quite different kind of multiplication is a *vector multiplication of vectors*. Taking a vector product of two vectors returns as a result a vector, as its name suggests. Vector products are used to define other derived vector quantities. For example, in describing rotations, a vector quantity called *torque* is defined as a vector product of an applied force (a vector) and its distance from pivot to force (a vector). It is important to distinguish between these two kinds of vector multiplications because the scalar product is a scalar quantity and a vector product is a vector quantity.

The Scalar Product of Two Vectors (the Dot Product)

Scalar multiplication of two vectors yields a scalar product.

Note:

Scalar Product (Dot Product)

The **scalar product** $\vec{A} \cdot \vec{B}$ of two vectors \vec{A} and \vec{B} is a number defined by the equation

Equation:

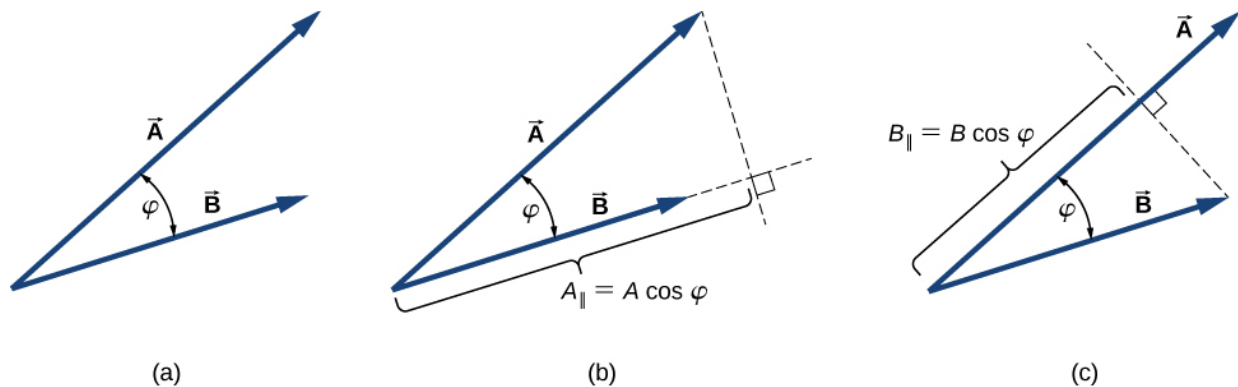
$$\vec{A} \cdot \vec{B} = AB \cos \varphi,$$

where φ is the angle between the vectors (shown in [link](#)). The scalar product is also called the **dot product** because of the dot notation that indicates it.

In the definition of the dot product, the direction of angle φ does not matter, and φ can be measured from either of the two vectors to the other because $\cos \varphi = \cos (-\varphi) = \cos (2\pi - \varphi)$. The dot product is a negative number when $90^\circ < \varphi \leq 180^\circ$ and is a positive number when $0^\circ \leq \varphi < 90^\circ$. Moreover, the dot product of two parallel vectors is $\vec{A} \cdot \vec{B} = AB \cos 0^\circ = AB$, and the dot product of two antiparallel vectors is $\vec{A} \cdot \vec{B} = AB \cos 180^\circ = -AB$. The scalar product of two *orthogonal vectors* vanishes: $\vec{A} \cdot \vec{B} = AB \cos 90^\circ = 0$. The scalar product of a vector with itself is the square of its magnitude:

Equation:

$$\vec{A}^2 \equiv \vec{A} \cdot \vec{A} = AA \cos 0^\circ = A^2.$$



The scalar product of two vectors. (a) The angle between the two vectors. (b) The orthogonal projection A_{\parallel} of vector \vec{A} onto the direction of vector \vec{B} . (c) The orthogonal projection B_{\parallel} of vector \vec{B} onto the direction of vector \vec{A} .

Example:
The Scalar Product

For the vectors shown in [\[link\]](#), find the scalar product $\vec{A} \cdot \vec{F}$.

Strategy

From [\[link\]](#), the magnitudes of vectors \vec{A} and \vec{F} are $A = 10.0$ and $F = 20.0$. Angle θ , between them, is the difference: $\theta = \varphi - \alpha = 110^\circ - 35^\circ = 75^\circ$. Substituting these values into [\[link\]](#) gives the scalar product.

Solution

A straightforward calculation gives us

Equation:

$$\vec{A} \cdot \vec{F} = AF \cos \theta = (10.0)(20.0) \cos 75^\circ = 51.76.$$

Note:

Exercise:

Problem:

Check Your Understanding For the vectors given in [\[link\]](#), find the scalar products $\vec{A} \cdot \vec{B}$ and $\vec{F} \cdot \vec{C}$.

Solution:

$$\vec{A} \cdot \vec{B} = -57.3, \vec{F} \cdot \vec{C} = 27.8$$

In the Cartesian coordinate system, scalar products of the unit vector of an axis with other unit vectors of axes always vanish because these unit vectors are orthogonal:

Equation:

$$\begin{aligned}\hat{\mathbf{i}} \cdot \hat{\mathbf{j}} &= \hat{\mathbf{i}} \cdot \hat{\mathbf{j}} \cos 90^\circ = (1)(1)(0) = 0, \\ \hat{\mathbf{i}} \cdot \hat{\mathbf{k}} &= \hat{\mathbf{i}} \cdot \hat{\mathbf{k}} \cos 90^\circ = (1)(1)(0) = 0, \\ \hat{\mathbf{k}} \cdot \hat{\mathbf{j}} &= \hat{\mathbf{k}} \cdot \hat{\mathbf{j}} \cos 90^\circ = (1)(1)(0) = 0.\end{aligned}$$

In these equations, we use the fact that the magnitudes of all unit vectors are one: $\hat{\mathbf{i}} = \hat{\mathbf{j}} = \hat{\mathbf{k}} = 1$. For unit vectors of the axes, [\[link\]](#) gives the following identities:

Equation:

$$\hat{\mathbf{i}} \cdot \hat{\mathbf{i}} = i^2 = \hat{\mathbf{j}} \cdot \hat{\mathbf{j}} = j^2 = \hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = k^2 = 1.$$

The scalar product $\vec{\mathbf{A}} \cdot \vec{\mathbf{B}}$ can also be interpreted as either the product of B with the projection A_{\parallel} of vector $\vec{\mathbf{A}}$ onto the direction of vector $\vec{\mathbf{B}}$ ([\[link\]](#)(b)) or the product of A with the projection B_{\parallel} of vector $\vec{\mathbf{B}}$ onto the direction of vector $\vec{\mathbf{A}}$ ([\[link\]](#)(c)):

Equation:

$$\begin{aligned}\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} &= AB \cos \varphi \\ &= B(A \cos \varphi) = BA_{\parallel} \\ &= A(B \cos \varphi) = AB_{\parallel}.\end{aligned}$$

For example, in the rectangular coordinate system in a plane, the scalar x -component of a vector is its dot product with the unit vector $\hat{\mathbf{i}}$, and the scalar y -component of a vector is its dot product with the unit vector $\hat{\mathbf{j}}$:

Equation:

$$\begin{aligned}\vec{\mathbf{A}} \cdot \hat{\mathbf{i}} &= \vec{\mathbf{A}} \cdot \hat{\mathbf{i}} \cos \theta_A = A \cos \theta_A = A_x \\ \vec{\mathbf{A}} \cdot \hat{\mathbf{j}} &= \vec{\mathbf{A}} \cdot \hat{\mathbf{j}} \cos (90^\circ - \theta_A) = A \sin \theta_A = A_y.\end{aligned}$$

Scalar multiplication of vectors is commutative,

Note:

Equation:

$$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = \vec{\mathbf{B}} \cdot \vec{\mathbf{A}},$$

and obeys the distributive law:

Note:

Equation:

$$\vec{A} \cdot (\vec{B} + \vec{C}) = \vec{A} \cdot \vec{B} + \vec{A} \cdot \vec{C}.$$

We can use the commutative and distributive laws to derive various relations for vectors, such as expressing the dot product of two vectors in terms of their scalar components.

Note:

Exercise:

Problem:

Check Your Understanding For vector $\vec{A} = A_x \hat{i} + A_y \hat{j} + A_z \hat{k}$ in a rectangular coordinate system, use [\[link\]](#) through [\[link\]](#) to show that $\vec{A} \cdot \hat{i} = A_x$, $\vec{A} \cdot \hat{j} = A_y$ and $\vec{A} \cdot \hat{k} = A_z$.

When the vectors in [\[link\]](#) are given in their vector component forms,

Equation:

$$\vec{A} = A_x \hat{i} + A_y \hat{j} + A_z \hat{k} \text{ and } \vec{B} = B_x \hat{i} + B_y \hat{j} + B_z \hat{k},$$

we can compute their scalar product as follows:

Equation:

$$\begin{aligned} \vec{A} \cdot \vec{B} &= (A_x \hat{i} + A_y \hat{j} + A_z \hat{k}) \cdot (B_x \hat{i} + B_y \hat{j} + B_z \hat{k}) \\ &= A_x B_x \hat{i} \cdot \hat{i} + A_x B_y \hat{i} \cdot \hat{j} + A_x B_z \hat{i} \cdot \hat{k} \\ &\quad + A_y B_x \hat{j} \cdot \hat{i} + A_y B_y \hat{j} \cdot \hat{j} + A_y B_z \hat{j} \cdot \hat{k} \\ &\quad + A_z B_x \hat{k} \cdot \hat{i} + A_z B_y \hat{k} \cdot \hat{j} + A_z B_z \hat{k} \cdot \hat{k}. \end{aligned}$$

Since scalar products of two different unit vectors of axes give zero, and scalar products of unit vectors with themselves give one (see [\[link\]](#) and [\[link\]](#)), there are only three nonzero terms in this expression. Thus, the scalar product simplifies to

Note:

Equation:

$$\vec{A} \cdot \vec{B} = A_x B_x + A_y B_y + A_z B_z.$$

We can use [\[link\]](#) for the scalar product in terms of scalar components of vectors to find the angle between two vectors. When we divide [\[link\]](#) by AB , we obtain the equation for $\cos \varphi$, into which we substitute [\[link\]](#):

Note:

Equation:

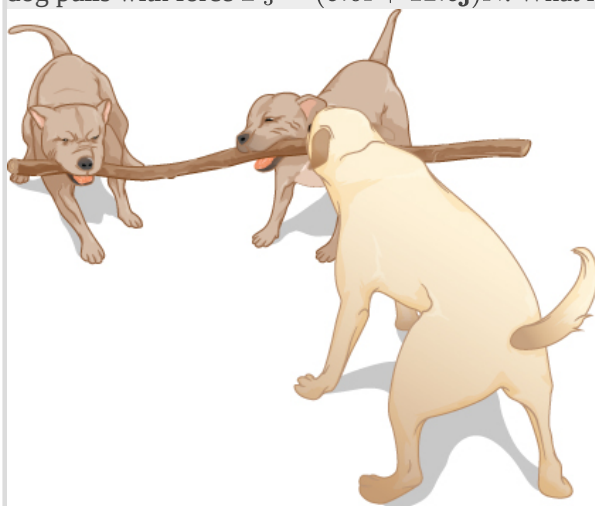
$$\cos \varphi = \frac{\vec{A} \cdot \vec{B}}{AB} = \frac{A_x B_x + A_y B_y + A_z B_z}{AB}.$$

Angle φ between vectors \vec{A} and \vec{B} is obtained by taking the inverse cosine of the expression in [\[link\]](#).

Example:

Angle between Two Forces

Three dogs are pulling on a stick in different directions, as shown in [\[link\]](#). The first dog pulls with force $\vec{F}_1 = (10.0\hat{i} - 20.4\hat{j} + 2.0\hat{k})\text{N}$, the second dog pulls with force $\vec{F}_2 = (-15.0\hat{i} - 6.2\hat{k})\text{N}$, and the third dog pulls with force $\vec{F}_3 = (5.0\hat{i} + 12.5\hat{j})\text{N}$. What is the angle between forces \vec{F}_1 and \vec{F}_2 ?



Three dogs are playing with a stick.

Strategy

The components of force vector \vec{F}_1 are $F_{1x} = 10.0\text{ N}$, $F_{1y} = -20.4\text{ N}$, and $F_{1z} = 2.0\text{ N}$, whereas those of force vector \vec{F}_2 are $F_{2x} = -15.0\text{ N}$, $F_{2y} = 0.0\text{ N}$, and $F_{2z} = -6.2\text{ N}$. Computing the scalar product of these vectors and their magnitudes, and substituting into [\[link\]](#) gives the angle of interest.

Solution

The magnitudes of forces \vec{F}_1 and \vec{F}_2 are

Equation:

$$F_1 = \sqrt{F_{1x}^2 + F_{1y}^2 + F_{1z}^2} = \sqrt{10.0^2 + 20.4^2 + 2.0^2} \text{ N} = 22.8 \text{ N}$$

and

Equation:

$$F_2 = \sqrt{F_{2x}^2 + F_{2y}^2 + F_{2z}^2} = \sqrt{15.0^2 + 6.2^2} \text{ N} = 16.2 \text{ N}.$$

Substituting the scalar components into [\[link\]](#) yields the scalar product

Equation:

$$\begin{aligned}\vec{\mathbf{F}}_1 \cdot \vec{\mathbf{F}}_2 &= F_{1x}F_{2x} + F_{1y}F_{2y} + F_{1z}F_{2z} \\ &= (10.0 \text{ N})(-15.0 \text{ N}) + (-20.4 \text{ N})(0.0 \text{ N}) + (2.0 \text{ N})(-6.2 \text{ N}) \\ &= -162.4 \text{ N}^2.\end{aligned}$$

Finally, substituting everything into [\[link\]](#) gives the angle

Equation:

$$\cos \varphi = \frac{\vec{\mathbf{F}}_1 \cdot \vec{\mathbf{F}}_2}{F_1 F_2} = \frac{-162.4 \text{ N}^2}{(22.8 \text{ N})(16.2 \text{ N})} = -0.439 \Rightarrow \varphi = \cos^{-1}(-0.439) = 116.0^\circ.$$

Significance

Notice that when vectors are given in terms of the unit vectors of axes, we can find the angle between them without knowing the specifics about the geographic directions the unit vectors represent. Here, for example, the +x-direction might be to the east and the +y-direction might be to the north. But, the angle between the forces in the problem is the same if the +x-direction is to the west and the +y-direction is to the south.

Note:

Exercise:

Problem: Check Your Understanding Find the angle between forces $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_3$ in [\[link\]](#).

Solution:

131.9°

Example:

The Work of a Force

When force $\vec{\mathbf{F}}$ pulls on an object and when it causes its displacement $\vec{\mathbf{D}}$, we say the force performs work. The amount of work the force does is the scalar product $\vec{\mathbf{F}} \cdot \vec{\mathbf{D}}$. If the stick in [\[link\]](#) moves momentarily and gets displaced by vector $\vec{\mathbf{D}} = (-7.9\hat{\mathbf{j}} - 4.2\hat{\mathbf{k}}) \text{ cm}$, how much work is done by the third dog in [\[link\]](#)?

Strategy

We compute the scalar product of displacement vector $\vec{\mathbf{D}}$ with force vector $\vec{\mathbf{F}}_3 = (5.0\hat{\mathbf{i}} + 12.5\hat{\mathbf{j}}) \text{ N}$, which is the pull from the third dog. Let's use W_3 to denote the work done by force $\vec{\mathbf{F}}_3$ on displacement $\vec{\mathbf{D}}$.

Solution

Calculating the work is a straightforward application of the dot product:

Equation:

$$\begin{aligned} W_3 &= \vec{\mathbf{F}}_3 \cdot \vec{\mathbf{D}} = F_{3x}D_x + F_{3y}D_y + F_{3z}D_z \\ &= (5.0 \text{ N})(0.0 \text{ cm}) + (12.5 \text{ N})(-7.9 \text{ cm}) + (0.0 \text{ N})(-4.2 \text{ cm}) \\ &= -98.7 \text{ N} \cdot \text{cm}. \end{aligned}$$

Significance

The SI unit of work is called the joule (J), where $1 \text{ J} = 1 \text{ N} \cdot \text{m}$. The unit $\text{cm} \cdot \text{N}$ can be written as $10^{-2} \text{m} \cdot \text{N} = 10^{-2} \text{J}$, so the answer can be expressed as $W_3 = -0.9875 \text{ J} \approx -1.0 \text{ J}$.

Note:

Exercise:

Problem:

Check Your Understanding How much work is done by the first dog and by the second dog in [\[link\]](#) on the displacement in [\[link\]](#)?

Solution:

$$W_1 = 1.5 \text{ J}, W_2 = 0.3 \text{ J}$$

The Vector Product of Two Vectors (the Cross Product)

Vector multiplication of two vectors yields a vector product.

Note:

Vector Product (Cross Product)

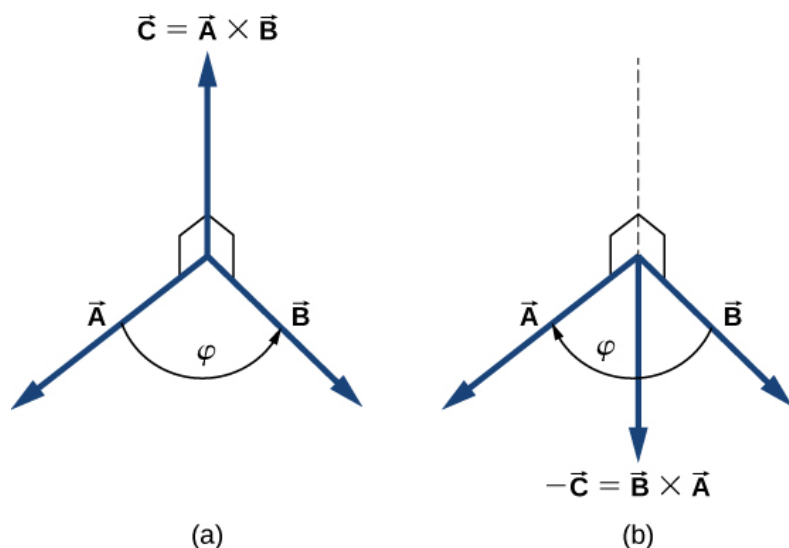
The **vector product** of two vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ is denoted by $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$ and is often referred to as a **cross product**. The vector product is a vector that has its direction perpendicular to both vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$. In other words, vector $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$ is perpendicular to the plane that contains vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$, as shown in [\[link\]](#). The magnitude of the vector product is defined as

Equation:

$$\vec{\mathbf{A}} \times \vec{\mathbf{B}} = AB \sin \varphi,$$

where angle φ , between the two vectors, is measured from vector $\vec{\mathbf{A}}$ (first vector in the product) to vector $\vec{\mathbf{B}}$ (second vector in the product), as indicated in [\[link\]](#), and is between 0° and 180° .

According to [\[link\]](#), the vector product vanishes for pairs of vectors that are either parallel ($\varphi = 0^\circ$) or antiparallel ($\varphi = 180^\circ$) because $\sin 0^\circ = \sin 180^\circ = 0$.



The vector product of two vectors is drawn in three-dimensional space. (a) The vector product $\vec{A} \times \vec{B}$ is a vector perpendicular to the plane that contains vectors \vec{A} and \vec{B} . Small squares drawn in perspective mark right angles between \vec{A} and \vec{C} , and between \vec{B} and \vec{C} so that if \vec{A} and \vec{B} lie on the floor, vector \vec{C} points vertically upward to the ceiling. (b) The vector product $\vec{B} \times \vec{A}$ is a vector antiparallel to vector $\vec{A} \times \vec{B}$.

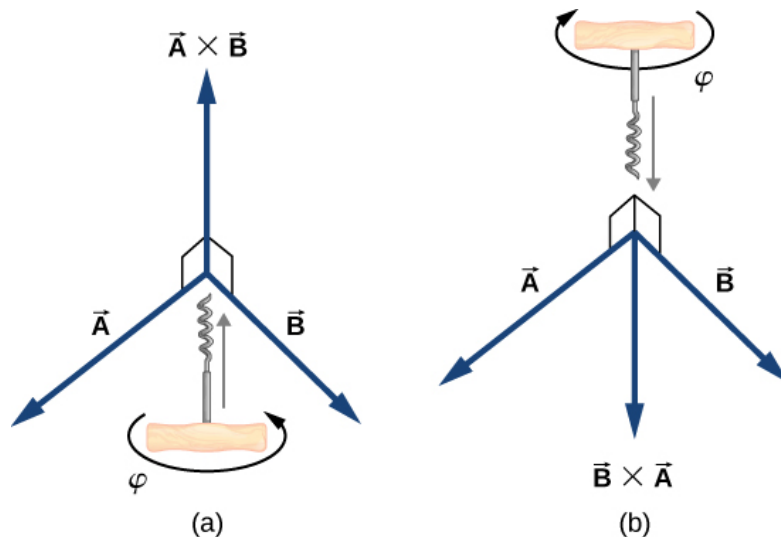
On the line perpendicular to the plane that contains vectors \vec{A} and \vec{B} there are two alternative directions—either up or down, as shown in [\[link\]](#)—and the direction of the vector product may be either one of them. In the standard right-handed orientation, where the angle between vectors is measured counterclockwise from the first vector, vector $\vec{A} \times \vec{B}$ points *upward*, as seen in [\[link\]](#)(a). If we reverse the order of multiplication, so that now \vec{B} comes first in the product, then vector $\vec{B} \times \vec{A}$ must point *downward*, as seen in [\[link\]](#)(b). This means that vectors $\vec{A} \times \vec{B}$ and $\vec{B} \times \vec{A}$ are *antiparallel* to each other and that vector multiplication is *not* commutative but *anticommutative*. The **anticommutative property** means the vector product reverses the sign when the order of multiplication is reversed:

Note:
Equation:

$$\vec{A} \times \vec{B} = -\vec{B} \times \vec{A}.$$

The **corkscrew right-hand rule** is a common mnemonic used to determine the direction of the vector product. As shown in [\[link\]](#), a corkscrew is placed in a direction perpendicular to the plane that contains

vectors \vec{A} and \vec{B} , and its handle is turned in the direction from the first to the second vector in the product. The direction of the cross product is given by the progression of the corkscrew.



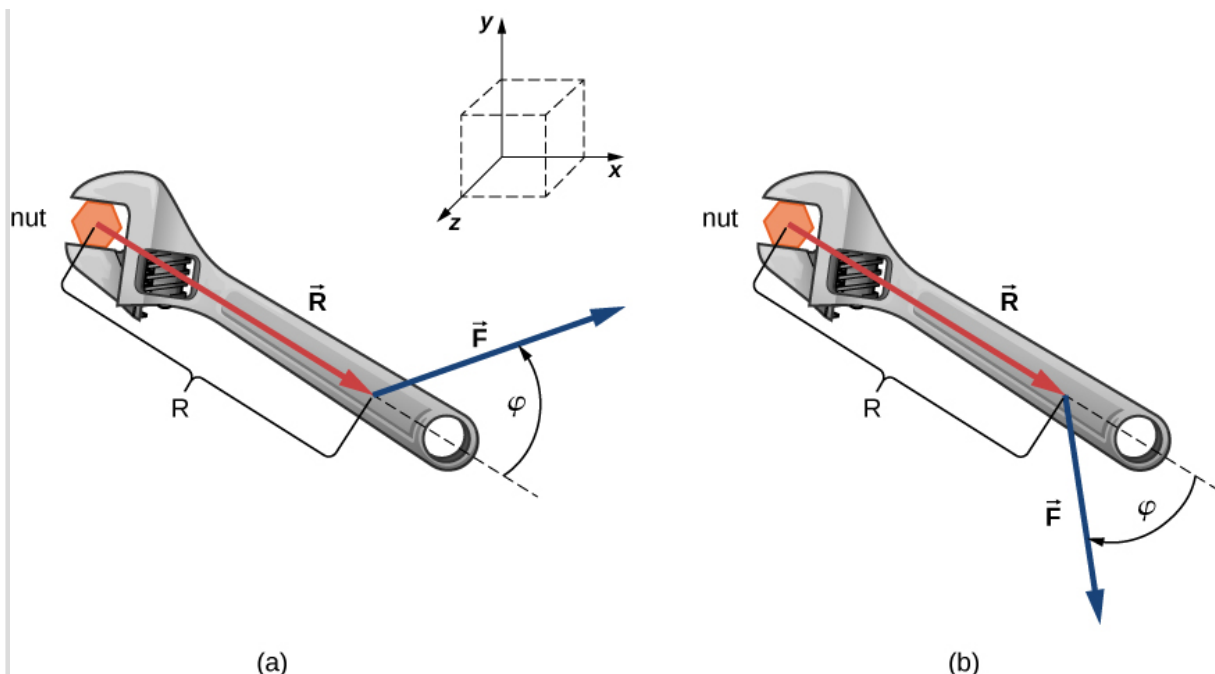
The corkscrew right-hand rule can be used to determine the direction of the cross product $\vec{A} \times \vec{B}$. Place a corkscrew in the direction perpendicular to the plane that contains vectors \vec{A} and \vec{B} , and turn it in the direction from the first to the second vector in the product. The direction of the cross product is given by the progression of the corkscrew. (a) Upward movement means the cross-product vector points up. (b) Downward movement means the cross-product vector points downward.

Example:

The Torque of a Force

The mechanical advantage that a familiar tool called a *wrench* provides ([link](#)) depends on magnitude F of the applied force, on its direction with respect to the wrench handle, and on how far from the nut this force is applied. The distance R from the nut to the point where force vector \vec{F} is attached is represented by the radial vector \vec{R} . The physical vector quantity that makes the nut turn is called *torque* (denoted by $\vec{\tau}$), and it is the vector product of the distance between the pivot to force with the force: $\vec{\tau} = \vec{R} \times \vec{F}$.

To loosen a rusty nut, a 20.00-N force is applied to the wrench handle at angle $\varphi = 40^\circ$ and at a distance of 0.25 m from the nut, as shown in [link](#)(a). Find the magnitude and direction of the torque applied to the nut. What would the magnitude and direction of the torque be if the force were applied at angle $\varphi = 45^\circ$, as shown in [link](#)(b)? For what value of angle φ does the torque have the largest magnitude?



A wrench provides grip and mechanical advantage in applying torque to turn a nut. (a) Turn counterclockwise to loosen the nut. (b) Turn clockwise to tighten the nut.

Strategy

We adopt the frame of reference shown in [\[link\]](#), where vectors $\vec{\mathbf{R}}$ and $\vec{\mathbf{F}}$ lie in the xy -plane and the origin is at the position of the nut. The radial direction along vector $\vec{\mathbf{R}}$ (pointing away from the origin) is the reference direction for measuring the angle φ because $\vec{\mathbf{R}}$ is the first vector in the vector product $\vec{\boldsymbol{\tau}} = \vec{\mathbf{R}} \times \vec{\mathbf{F}}$. Vector $\vec{\boldsymbol{\tau}}$ must lie along the z -axis because this is the axis that is perpendicular to the xy -plane, where both $\vec{\mathbf{R}}$ and $\vec{\mathbf{F}}$ lie. To compute the magnitude τ , we use [\[link\]](#). To find the direction of $\vec{\boldsymbol{\tau}}$, we use the corkscrew right-hand rule ([\[link\]](#)).

Solution

For the situation in (a), the corkscrew rule gives the direction of $\vec{\mathbf{R}} \times \vec{\mathbf{F}}$ in the positive direction of the z -axis. Physically, it means the torque vector $\vec{\boldsymbol{\tau}}$ points out of the page, perpendicular to the wrench handle. We identify $F = 20.00 \text{ N}$ and $R = 0.25 \text{ m}$, and compute the magnitude using [\[link\]](#):

Equation:

$$\tau = \vec{\mathbf{R}} \times \vec{\mathbf{F}} = RF \sin \varphi = (0.25 \text{ m})(20.00 \text{ N}) \sin 40^\circ = 3.21 \text{ N} \cdot \text{m}.$$

For the situation in (b), the corkscrew rule gives the direction of $\vec{\mathbf{R}} \times \vec{\mathbf{F}}$ in the negative direction of the z -axis. Physically, it means the vector $\vec{\boldsymbol{\tau}}$ points into the page, perpendicular to the wrench handle. The magnitude of this torque is

Equation:

$$\tau = \vec{\mathbf{R}} \times \vec{\mathbf{F}} = RF \sin \varphi = (0.25 \text{ m})(20.00 \text{ N}) \sin 45^\circ = 3.53 \text{ N} \cdot \text{m}.$$

The torque has the largest value when $\sin \varphi = 1$, which happens when $\varphi = 90^\circ$. Physically, it means the wrench is most effective—giving us the best mechanical advantage—when we apply the force

perpendicular to the wrench handle. For the situation in this example, this best-torque value is $\tau_{\text{best}} = RF = (0.25 \text{ m})(20.00 \text{ N}) = 5.00 \text{ N} \cdot \text{m}$.

Significance

When solving mechanics problems, we often do not need to use the corkscrew rule at all, as we'll see now in the following equivalent solution. Notice that once we have identified that vector $\vec{\mathbf{R}} \times \vec{\mathbf{F}}$ lies along the z-axis, we can write this vector in terms of the unit vector $\hat{\mathbf{k}}$ of the z-axis:

Equation:

$$\vec{\mathbf{R}} \times \vec{\mathbf{F}} = RF \sin \varphi \hat{\mathbf{k}}.$$

In this equation, the number that multiplies $\hat{\mathbf{k}}$ is the scalar z-component of the vector $\vec{\mathbf{R}} \times \vec{\mathbf{F}}$. In the computation of this component, care must be taken that the angle φ is measured *counterclockwise* from $\vec{\mathbf{R}}$ (first vector) to $\vec{\mathbf{F}}$ (second vector). Following this principle for the angles, we obtain $RF \sin(+40^\circ) = +3.2 \text{ N} \cdot \text{m}$ for the situation in (a), and we obtain $RF \sin(-45^\circ) = -3.5 \text{ N} \cdot \text{m}$ for the situation in (b). In the latter case, the angle is negative because the graph in [\[link\]](#) indicates the angle is measured clockwise; but, the same result is obtained when this angle is measured counterclockwise because $+(360^\circ - 45^\circ) = +315^\circ$ and $\sin(+315^\circ) = \sin(-45^\circ)$. In this way, we obtain the solution without reference to the corkscrew rule. For the situation in (a), the solution is $\vec{\mathbf{R}} \times \vec{\mathbf{F}} = +3.2 \text{ N} \cdot \text{m} \hat{\mathbf{k}}$; for the situation in (b), the solution is $\vec{\mathbf{R}} \times \vec{\mathbf{F}} = -3.5 \text{ N} \cdot \text{m} \hat{\mathbf{k}}$.

Note:

Exercise:

Problem:

Check Your Understanding For the vectors given in [\[link\]](#), find the vector products $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$ and $\vec{\mathbf{C}} \times \vec{\mathbf{F}}$.

Solution:

$\vec{\mathbf{A}} \times \vec{\mathbf{B}} = -40.1 \hat{\mathbf{k}}$ or, equivalently, $\vec{\mathbf{A}} \times \vec{\mathbf{B}} = 40.1$, and the direction is into the page;

$\vec{\mathbf{C}} \times \vec{\mathbf{F}} = +157.6 \hat{\mathbf{k}}$ or, equivalently, $\vec{\mathbf{C}} \times \vec{\mathbf{F}} = 157.6$, and the direction is out of the page.

Similar to the dot product ([\[link\]](#)), the cross product has the following distributive property:

Note:

Equation:

$$\vec{\mathbf{A}} \times (\vec{\mathbf{B}} + \vec{\mathbf{C}}) = \vec{\mathbf{A}} \times \vec{\mathbf{B}} + \vec{\mathbf{A}} \times \vec{\mathbf{C}}.$$

The distributive property is applied frequently when vectors are expressed in their component forms, in terms of unit vectors of Cartesian axes.

When we apply the definition of the cross product, [\[link\]](#), to unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ that define the positive x-, y-, and z-directions in space, we find that

Equation:

$$\hat{\mathbf{i}} \times \hat{\mathbf{i}} = \hat{\mathbf{j}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}} \times \hat{\mathbf{k}} = 0.$$

All other cross products of these three unit vectors must be vectors of unit magnitudes because $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ are orthogonal. For example, for the pair $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$, the magnitude is $\hat{\mathbf{i}} \times \hat{\mathbf{j}} = ij \sin 90^\circ = (1)(1)(1) = 1$.

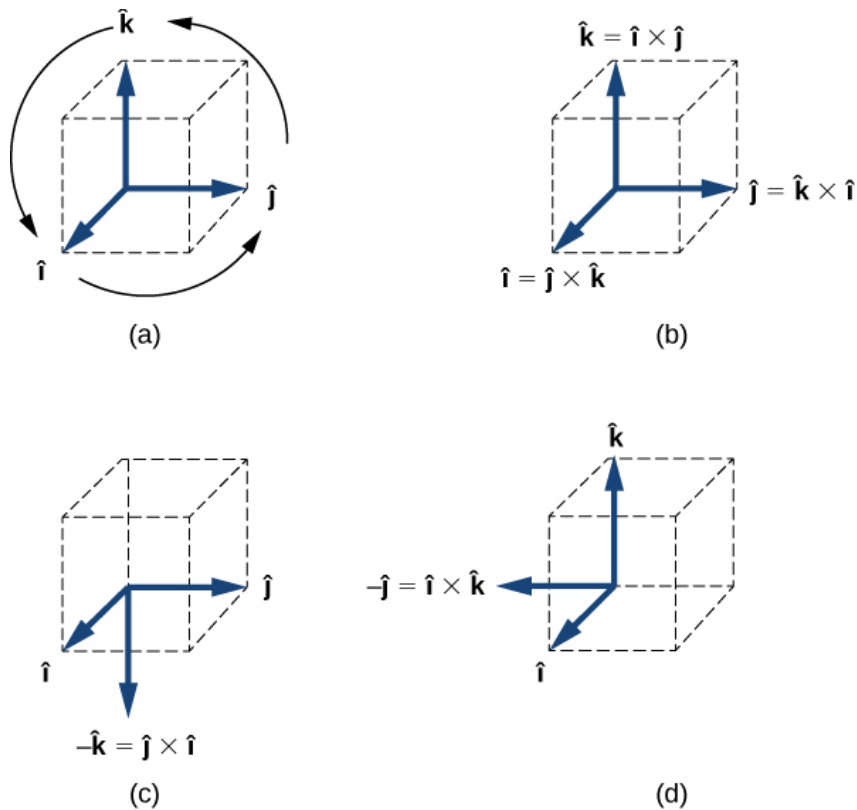
The direction of the vector product $\hat{\mathbf{i}} \times \hat{\mathbf{j}}$ must be orthogonal to the xy-plane, which means it must be along the z-axis. The only unit vectors along the z-axis are $-\hat{\mathbf{k}}$ or $+\hat{\mathbf{k}}$. By the corkscrew rule, the direction of vector $\hat{\mathbf{i}} \times \hat{\mathbf{j}}$ must be parallel to the positive z-axis. Therefore, the result of the multiplication $\hat{\mathbf{i}} \times \hat{\mathbf{j}}$ is identical to $+\hat{\mathbf{k}}$. We can repeat similar reasoning for the remaining pairs of unit vectors. The results of these multiplications are

Note:

Equation:

$$\begin{aligned}\hat{\mathbf{i}} \times \hat{\mathbf{j}} &= +\hat{\mathbf{k}}, \\ \hat{\mathbf{j}} \times \hat{\mathbf{k}} &= +\hat{\mathbf{i}}, \\ \hat{\mathbf{k}} \times \hat{\mathbf{i}} &= +\hat{\mathbf{j}}.\end{aligned}$$

Notice that in [\[link\]](#), the three unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ appear in the *cyclic order* shown in a diagram in [\[link\]](#) (a). The cyclic order means that in the product formula, $\hat{\mathbf{i}}$ follows $\hat{\mathbf{k}}$ and comes before $\hat{\mathbf{j}}$, or $\hat{\mathbf{k}}$ follows $\hat{\mathbf{j}}$ and comes before $\hat{\mathbf{i}}$, or $\hat{\mathbf{j}}$ follows $\hat{\mathbf{i}}$ and comes before $\hat{\mathbf{k}}$. The cross product of two different unit vectors is always a third unit vector. When two unit vectors in the cross product appear in the cyclic order, the result of such a multiplication is the remaining unit vector, as illustrated in [\[link\]](#)(b). When unit vectors in the cross product appear in a different order, the result is a unit vector that is antiparallel to the remaining unit vector (i.e., the result is with the minus sign, as shown by the examples in [\[link\]](#)(c) and [\[link\]](#)(d). In practice, when the task is to find cross products of vectors that are given in vector component form, this rule for the cross-multiplication of unit vectors is very useful.



(a) The diagram of the cyclic order of the unit vectors of the axes. (b) The only cross products where the unit vectors appear in the cyclic order. These products have the positive sign. (c, d) Two examples of cross products where the unit vectors do not appear in the cyclic order. These products have the negative sign.

Suppose we want to find the cross product $\vec{A} \times \vec{B}$ for vectors $\vec{A} = A_x \hat{i} + A_y \hat{j} + A_z \hat{k}$ and $\vec{B} = B_x \hat{i} + B_y \hat{j} + B_z \hat{k}$. We can use the distributive property ([\[link\]](#)), the anticommutative property ([\[link\]](#)), and the results in [\[link\]](#) and [\[link\]](#) for unit vectors to perform the following algebra:

Equation:

$$\begin{aligned}
 \vec{A} \times \vec{B} &= (A_x \hat{i} + A_y \hat{j} + A_z \hat{k}) \times (B_x \hat{i} + B_y \hat{j} + B_z \hat{k}) \\
 &= A_x \hat{i} \times (B_x \hat{i} + B_y \hat{j} + B_z \hat{k}) + A_y \hat{j} \times (B_x \hat{i} + B_y \hat{j} + B_z \hat{k}) + A_z \hat{k} \times (B_x \hat{i} + B_y \hat{j} + B_z \hat{k}) \\
 &= A_x B_x \hat{i} \times \hat{i} + A_x B_y \hat{i} \times \hat{j} + A_x B_z \hat{i} \times \hat{k} \\
 &\quad + A_y B_x \hat{j} \times \hat{i} + A_y B_y \hat{j} \times \hat{j} + A_y B_z \hat{j} \times \hat{k} \\
 &\quad + A_z B_x \hat{k} \times \hat{i} + A_z B_y \hat{k} \times \hat{j} + A_z B_z \hat{k} \times \hat{k} \\
 &= A_x B_x (0) + A_x B_y (+\hat{k}) + A_x B_z (-\hat{j}) \\
 &\quad + A_y B_x (-\hat{k}) + A_y B_y (0) + A_y B_z (+\hat{i}) \\
 &\quad + A_z B_x (+\hat{j}) + A_z B_y (-\hat{i}) + A_z B_z (0).
 \end{aligned}$$

When performing algebraic operations involving the cross product, be very careful about keeping the correct order of multiplication because the cross product is anticommutative. The last two steps that we still have to do to complete our task are, first, grouping the terms that contain a common unit vector and, second, factoring. In this way we obtain the following very useful expression for the computation of the cross product:

Note:

Equation:

$$\vec{C} = \vec{A} \times \vec{B} = (A_y B_z - A_z B_y)\hat{i} + (A_z B_x - A_x B_z)\hat{j} + (A_x B_y - A_y B_x)\hat{k}.$$

In this expression, the scalar components of the cross-product vector are

Equation:

$$C_x = A_y B_z - A_z B_y,$$

$$C_y = A_z B_x - A_x B_z,$$

$$C_z = A_x B_y - A_y B_x.$$

When finding the cross product, in practice, we can use either [\[link\]](#) or [\[link\]](#), depending on which one of them seems to be less complex computationally. They both lead to the same final result. One way to make sure if the final result is correct is to use them both.

Example:

A Particle in a Magnetic Field

When moving in a magnetic field, some particles may experience a magnetic force. Without going into details—a detailed study of magnetic phenomena comes in later chapters—let’s acknowledge that the magnetic field \vec{B} is a vector, the magnetic force \vec{F} is a vector, and the velocity \vec{u} of the particle is a vector. The magnetic force vector is proportional to the vector product of the velocity vector with the magnetic field vector, which we express as $\vec{F} = \zeta \vec{u} \times \vec{B}$. In this equation, a constant ζ takes care of the consistency in physical units, so we can omit physical units on vectors \vec{u} and \vec{B} . In this example, let’s assume the constant ζ is positive.

A particle moving in space with velocity vector $\vec{u} = -5.0\hat{i} - 2.0\hat{j} + 3.5\hat{k}$ enters a region with a magnetic field and experiences a magnetic force. Find the magnetic force \vec{F} on this particle at the entry point to the region where the magnetic field vector is (a) $\vec{B} = 7.2\hat{i} - \hat{j} - 2.4\hat{k}$ and (b) $\vec{B} = 4.5\hat{k}$. In each case, find magnitude F of the magnetic force and angle θ the force vector \vec{F} makes with the given magnetic field vector \vec{B} .

Strategy

First, we want to find the vector product $\vec{u} \times \vec{B}$, because then we can determine the magnetic force using $\vec{F} = \zeta \vec{u} \times \vec{B}$. Magnitude F can be found either by using components, $F = \sqrt{F_x^2 + F_y^2 + F_z^2}$, or by computing the magnitude $|\vec{u} \times \vec{B}|$ directly using [\[link\]](#). In the latter approach, we would have to find the angle between vectors \vec{u} and \vec{B} . When we have \vec{F} , the general method for finding the direction angle θ

involves the computation of the scalar product $\vec{\mathbf{F}} \cdot \vec{\mathbf{B}}$ and substitution into [\[link\]](#). To compute the vector product we can either use [\[link\]](#) or compute the product directly, whichever way is simpler.

Solution

The components of the velocity vector are $u_x = -5.0$, $u_y = -2.0$, and $u_z = 3.5$.

(a) The components of the magnetic field vector are $B_x = 7.2$, $B_y = -1.0$, and $B_z = -2.4$. Substituting them into [\[link\]](#) gives the scalar components of vector $\vec{\mathbf{F}} = \zeta \vec{\mathbf{u}} \times \vec{\mathbf{B}}$:

Equation:

$$\begin{aligned} F_x &= \zeta(u_y B_z - u_z B_y) = \zeta[(-2.0)(-2.4) - (3.5)(-1.0)] = 8.3\zeta \\ F_y &= \zeta(u_z B_x - u_x B_z) = \zeta[(3.5)(7.2) - (-5.0)(-2.4)] = 13.2\zeta \\ F_z &= \zeta(u_x B_y - u_y B_x) = \zeta[(-5.0)(-1.0) - (-2.0)(7.2)] = 19.4\zeta \end{aligned}$$

Thus, the magnetic force is $\vec{\mathbf{F}} = \zeta(8.3\hat{\mathbf{i}} + 13.2\hat{\mathbf{j}} + 19.4\hat{\mathbf{k}})$ and its magnitude is

Equation:

$$F = \sqrt{F_x^2 + F_y^2 + F_z^2} = \zeta \sqrt{(8.3)^2 + (13.2)^2 + (19.4)^2} = 24.9\zeta.$$

To compute angle θ , we may need to find the magnitude of the magnetic field vector,

Equation:

$$B = \sqrt{B_x^2 + B_y^2 + B_z^2} = \sqrt{(7.2)^2 + (-1.0)^2 + (-2.4)^2} = 7.6,$$

and the scalar product $\vec{\mathbf{F}} \cdot \vec{\mathbf{B}}$:

Equation:

$$\vec{\mathbf{F}} \cdot \vec{\mathbf{B}} = F_x B_x + F_y B_y + F_z B_z = (8.3\zeta)(7.2) + (13.2\zeta)(-1.0) + (19.4\zeta)(-2.4) = 0.$$

Now, substituting into [\[link\]](#) gives angle θ :

Equation:

$$\cos \theta = \frac{\vec{\mathbf{F}} \cdot \vec{\mathbf{B}}}{FB} = \frac{0}{(24.9\zeta)(7.6)} = 0 \Rightarrow \theta = 90^\circ.$$

Hence, the magnetic force vector is perpendicular to the magnetic field vector. (We could have saved some time if we had computed the scalar product earlier.)

(b) Because vector $\vec{\mathbf{B}} = 4.5\hat{\mathbf{k}}$ has only one component, we can perform the algebra quickly and find the vector product directly:

Equation:

$$\begin{aligned} \vec{\mathbf{F}} &= \zeta \vec{\mathbf{u}} \times \vec{\mathbf{B}} = \zeta(-5.0\hat{\mathbf{i}} - 2.0\hat{\mathbf{j}} + 3.5\hat{\mathbf{k}}) \times (4.5\hat{\mathbf{k}}) \\ &= \zeta[(-5.0)(4.5)\hat{\mathbf{i}} \times \hat{\mathbf{k}} + (-2.0)(4.5)\hat{\mathbf{j}} \times \hat{\mathbf{k}} + (3.5)(4.5)\hat{\mathbf{k}} \times \hat{\mathbf{k}}] \\ &= \zeta[-22.5(-\hat{\mathbf{j}}) - 9.0(+\hat{\mathbf{i}}) + 0] = \zeta(-9.0\hat{\mathbf{i}} + 22.5\hat{\mathbf{j}}). \end{aligned}$$

The magnitude of the magnetic force is

Equation:

$$F = \sqrt{F_x^2 + F_y^2 + F_z^2} = \zeta \sqrt{(-9.0)^2 + (22.5)^2 + (0.0)^2} = 24.2\zeta.$$

Because the scalar product is

Equation:

$$\vec{\mathbf{F}} \cdot \vec{\mathbf{B}} = F_x B_x + F_y B_y + F_z B_z = (-9.0\zeta)(0) + (22.5\zeta)(0) + (0)(4.5) = 0,$$

the magnetic force vector $\vec{\mathbf{F}}$ is perpendicular to the magnetic field vector $\vec{\mathbf{B}}$.

Significance

Even without actually computing the scalar product, we can predict that the magnetic force vector must always be perpendicular to the magnetic field vector because of the way this vector is constructed. Namely, the magnetic force vector is the vector product $\vec{\mathbf{F}} = \zeta \vec{\mathbf{u}} \times \vec{\mathbf{B}}$ and, by the definition of the vector product (see [\[link\]](#)), vector $\vec{\mathbf{F}}$ must be perpendicular to both vectors $\vec{\mathbf{u}}$ and $\vec{\mathbf{B}}$.

Note:

Exercise:

Problem:

Check Your Understanding Given two vectors $\vec{\mathbf{A}} = -\hat{\mathbf{i}} + \hat{\mathbf{j}}$ and $\vec{\mathbf{B}} = 3\hat{\mathbf{i}} - \hat{\mathbf{j}}$, find (a) $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$, (b) $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$, (c) the angle between $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$, and (d) the angle between $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$ and vector $\vec{\mathbf{C}} = \hat{\mathbf{i}} + \hat{\mathbf{k}}$.

Solution:

a. $-2\hat{\mathbf{k}}$, b. 2, c. 153.4° , d. 135°

In conclusion to this section, we want to stress that “dot product” and “cross product” are entirely different mathematical objects that have different meanings. The dot product is a scalar; the cross product is a vector. Later chapters use the terms *dot product* and *scalar product* interchangeably. Similarly, the terms *cross product* and *vector product* are used interchangeably.

Summary

- There are two kinds of multiplication for vectors. One kind of multiplication is the scalar product, also known as the dot product. The other kind of multiplication is the vector product, also known as the cross product. The scalar product of vectors is a number (scalar). The vector product of vectors is a vector.
- Both kinds of multiplication have the distributive property, but only the scalar product has the commutative property. The vector product has the anticommutative property, which means that when we change the order in which two vectors are multiplied, the result acquires a minus sign.
- The scalar product of two vectors is obtained by multiplying their magnitudes with the cosine of the angle between them. The scalar product of orthogonal vectors vanishes; the scalar product of antiparallel vectors is negative.
- The vector product of two vectors is a vector perpendicular to both of them. Its magnitude is obtained by multiplying their magnitudes by the sine of the angle between them. The direction of the vector product can be determined by the corkscrew right-hand rule. The vector product of two either parallel or antiparallel vectors vanishes. The magnitude of the vector product is largest for orthogonal vectors.
- The scalar product of vectors is used to find angles between vectors and in the definitions of derived scalar physical quantities such as work or energy.

- The cross product of vectors is used in definitions of derived vector physical quantities such as torque or magnetic force, and in describing rotations.

Key Equations

Multiplication by a scalar (vector equation)	$\vec{\mathbf{B}} = \alpha \vec{\mathbf{A}}$
Multiplication by a scalar (scalar equation for magnitudes)	$B = \alpha A$
Resultant of two vectors	$\vec{\mathbf{D}}_{AD} = \vec{\mathbf{D}}_{AC} + \vec{\mathbf{D}}_{CD}$
Commutative law	$\vec{\mathbf{A}} + \vec{\mathbf{B}} = \vec{\mathbf{B}} + \vec{\mathbf{A}}$
Associative law	$(\vec{\mathbf{A}} + \vec{\mathbf{B}}) + \vec{\mathbf{C}} = \vec{\mathbf{A}} + (\vec{\mathbf{B}} + \vec{\mathbf{C}})$
Distributive law	$\alpha_1 \vec{\mathbf{A}} + \alpha_2 \vec{\mathbf{A}} = (\alpha_1 + \alpha_2) \vec{\mathbf{A}}$
The component form of a vector in two dimensions	$\vec{\mathbf{A}} = A_x \hat{\mathbf{i}} + A_y \hat{\mathbf{j}}$
Scalar components of a vector in two dimensions	$\begin{cases} A_x = x_e - x_b \\ A_y = y_e - y_b \end{cases}$
Magnitude of a vector in a plane	$A = \sqrt{A_x^2 + A_y^2}$
The direction angle of a vector in a plane	$\theta_A = \tan^{-1} \left(\frac{A_y}{A_x} \right)$
Scalar components of a vector in a plane	$\begin{cases} A_x = A \cos \theta_A \\ A_y = A \sin \theta_A \end{cases}$
Polar coordinates in a plane	$\begin{cases} x = r \cos \varphi \\ y = r \sin \varphi \end{cases}$
The component form of a vector in three dimensions	$\vec{\mathbf{A}} = A_x \hat{\mathbf{i}} + A_y \hat{\mathbf{j}} + A_z \hat{\mathbf{k}}$

The scalar z-component of a vector in three dimensions	$A_z = z_e - z_b$
Magnitude of a vector in three dimensions	$A = \sqrt{A_x^2 + A_y^2 + A_z^2}$
Distributive property	$\alpha(\vec{\mathbf{A}} + \vec{\mathbf{B}}) = \alpha\vec{\mathbf{A}} + \alpha\vec{\mathbf{B}}$
Antiparallel vector to $\vec{\mathbf{A}}$	$-\vec{\mathbf{A}} = -A_x\hat{\mathbf{i}} - A_y\hat{\mathbf{j}} - A_z\hat{\mathbf{k}}$
Equal vectors	$\vec{\mathbf{A}} = \vec{\mathbf{B}} \Leftrightarrow \begin{aligned} A_x &= B_x \\ A_y &= B_y \\ A_z &= B_z \end{aligned}$
Components of the resultant of N vectors	$F_{Rx} = \sum_{k=1}^N F_{kx} = F_{1x} + F_{2x} + \dots + F_{Nx}$ $F_{Ry} = \sum_{k=1}^N F_{ky} = F_{1y} + F_{2y} + \dots + F_{Ny}$ $F_{Rz} = \sum_{k=1}^N F_{kz} = F_{1z} + F_{2z} + \dots + F_{Nz}$
General unit vector	$\hat{\mathbf{V}} = \frac{\vec{\mathbf{V}}}{V}$
Definition of the scalar product	$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = AB \cos \varphi$
Commutative property of the scalar product	$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = \vec{\mathbf{B}} \cdot \vec{\mathbf{A}}$
Distributive property of the scalar product	$\vec{\mathbf{A}} \cdot (\vec{\mathbf{B}} + \vec{\mathbf{C}}) = \vec{\mathbf{A}} \cdot \vec{\mathbf{B}} + \vec{\mathbf{A}} \cdot \vec{\mathbf{C}}$
Scalar product in terms of scalar components of vectors	$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = A_x B_x + A_y B_y + A_z B_z$
Cosine of the angle between two vectors	$\cos \varphi = \frac{\vec{\mathbf{A}} \cdot \vec{\mathbf{B}}}{AB}$
Dot products of unit vectors	$\hat{\mathbf{i}} \cdot \hat{\mathbf{j}} = \hat{\mathbf{j}} \cdot \hat{\mathbf{k}} = \hat{\mathbf{k}} \cdot \hat{\mathbf{i}} = 0$
Magnitude of the vector product (definition)	$\vec{\mathbf{A}} \times \vec{\mathbf{B}} = AB \sin \varphi$

Anticommutative property of the vector product	$\vec{A} \times \vec{B} = -\vec{B} \times \vec{A}$
Distributive property of the vector product	$\vec{A} \times (\vec{B} + \vec{C}) = \vec{A} \times \vec{B} + \vec{A} \times \vec{C}$
Cross products of unit vectors	$\hat{i} \times \hat{j} = +\hat{k},$ $\hat{j} \times \hat{k} = +\hat{i},$ $\hat{k} \times \hat{i} = +\hat{j}.$
The cross product in terms of scalar components of vectors	$\vec{A} \times \vec{B} = (A_y B_z - A_z B_y)\hat{i} + (A_z B_x - A_x B_z)\hat{j} + (A_x B_y - A_y B_x)\hat{k}$

Conceptual Questions

Exercise:

Problem:

What is wrong with the following expressions? How can you correct them? (a) $C = \vec{A}\vec{B}$, (b) $\vec{C} = \vec{A}\vec{B}$, (c) $C = \vec{A} \times \vec{B}$, (d) $C = A\vec{B}$, (e) $C + 2\vec{A} = B$, (f) $\vec{C} = A \times \vec{B}$, (g) $\vec{A} \cdot \vec{B} = \vec{A} \times \vec{B}$, (h) $\vec{C} = 2\vec{A} \cdot \vec{B}$, (i) $C = \vec{A}/\vec{B}$, and (j) $C = \vec{A}/B$.

Solution:

a. $C = \vec{A} \cdot \vec{B}$, b. $\vec{C} = \vec{A} \times \vec{B}$ or $\vec{C} = \vec{A} - \vec{B}$, c. $\vec{C} = \vec{A} \times \vec{B}$, d. $\vec{C} = A\vec{B}$, e. $\vec{C} + 2\vec{A} = \vec{B}$, f. $\vec{C} = \vec{A} \times \vec{B}$, g. left side is a scalar and right side is a vector, h. $\vec{C} = 2\vec{A} \times \vec{B}$, i. $\vec{C} = \vec{A}/B$, j. $\vec{C} = \vec{A}/B$

Exercise:

Problem: If the cross product of two vectors vanishes, what can you say about their directions?

Exercise:

Problem: If the dot product of two vectors vanishes, what can you say about their directions?

Solution:

They are orthogonal.

Exercise:

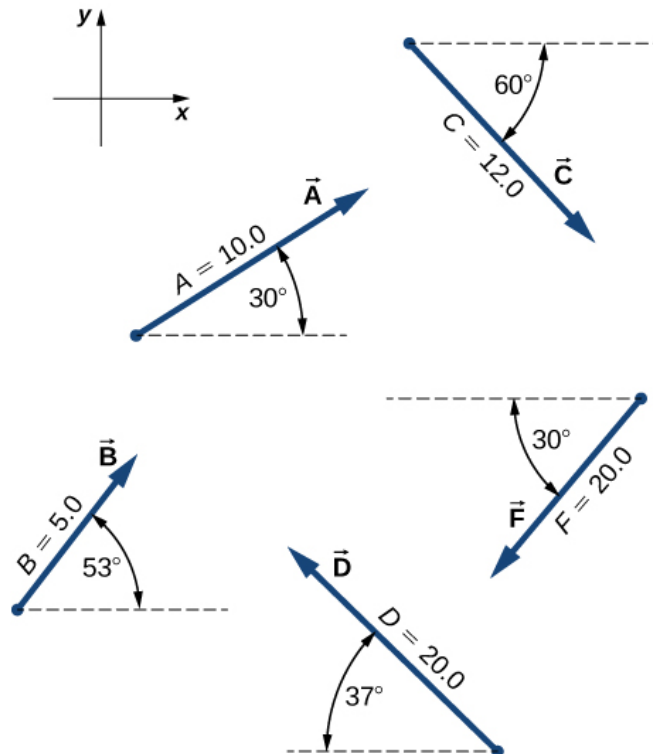
Problem:

What is the dot product of a vector with the cross product that this vector has with another vector?

Problems

Exercise:**Problem:**

Assuming the +x-axis is horizontal to the right for the vectors in the following figure, find the following scalar products: (a) $\vec{A} \cdot \vec{C}$, (b) $\vec{A} \cdot \vec{F}$, (c) $\vec{D} \cdot \vec{C}$, (d) $\vec{A} \cdot (\vec{F} + 2\vec{C})$, (e) $\hat{i} \cdot \vec{B}$, (f) $\hat{j} \cdot \vec{B}$, (g) $(3\hat{i} - \hat{j}) \cdot \vec{B}$, and (h) $\vec{B} \cdot \vec{B}$.

**Exercise:****Problem:**

Assuming the +x-axis is horizontal to the right for the vectors in the preceding figure, find (a) the component of vector \vec{A} along vector \vec{C} , (b) the component of vector \vec{C} along vector \vec{A} , (c) the component of vector \hat{i} along vector \vec{F} , and (d) the component of vector \vec{F} along vector \hat{i} .

Solution:

a. 0, b. 0, c. -0.866, d. -17.32

Exercise:**Problem:**

Find the angle between vectors for (a) $\vec{D} = (-3.0\hat{i} - 4.0\hat{j})\text{m}$ and $\vec{A} = (-3.0\hat{i} + 4.0\hat{j})\text{m}$ and (b) $\vec{D} = (2.0\hat{i} - 4.0\hat{j} + \hat{k})\text{m}$ and $\vec{B} = (-2.0\hat{i} + 3.0\hat{j} + 2.0\hat{k})\text{m}$.

Exercise:

Problem: Find the angles that vector $\vec{D} = (2.0\hat{i} - 4.0\hat{j} + \hat{k})\text{m}$ makes with the x -, y -, and z - axes.

Solution:

$$\theta_i = 64.12^\circ, \theta_j = 150.79^\circ, \theta_k = 77.39^\circ$$

Exercise:

Problem:

Show that the force vector $\vec{D} = (2.0\hat{i} - 4.0\hat{j} + \hat{k})\text{N}$ is orthogonal to the force vector $\vec{G} = (3.0\hat{i} + 4.0\hat{j} + 10.0\hat{k})\text{N}$.

Exercise:

Problem:

Assuming the $+x$ -axis is horizontal to the right for the vectors in the previous figure, find the following vector products: (a) $\vec{A} \times \vec{C}$, (b) $\vec{A} \times \vec{F}$, (c) $\vec{D} \times \vec{C}$, (d) $\vec{A} \times (\vec{F} + 2\vec{C})$, (e) $\hat{i} \times \vec{B}$, (f) $\hat{j} \times \vec{B}$, (g) $(3\hat{i} - \hat{j}) \times \vec{B}$, and (h) $\hat{B} \times \vec{B}$.

Solution:

a. $-120\hat{k}$, b. $0\hat{k}$, c. $-94\hat{k}$, d. $-240\hat{k}$, e. $4.0\hat{k}$, f. $-3.0\hat{k}$, g. $15\hat{k}$, h. 0

Exercise:

Problem:

Find the cross product $\vec{A} \times \vec{C}$ for (a) $\vec{A} = 2.0\hat{i} - 4.0\hat{j} + \hat{k}$ and $\vec{C} = 3.0\hat{i} + 4.0\hat{j} + 10.0\hat{k}$, (b) $\vec{A} = 3.0\hat{i} + 4.0\hat{j} + 10.0\hat{k}$ and $\vec{C} = 2.0\hat{i} - 4.0\hat{j} + \hat{k}$, (c) $\vec{A} = -3.0\hat{i} - 4.0\hat{j}$ and $\vec{C} = -3.0\hat{i} + 4.0\hat{j}$, and (d) $\vec{C} = -2.0\hat{i} + 3.0\hat{j} + 2.0\hat{k}$ and $\vec{A} = -9.0\hat{j}$.

Exercise:

Problem:

For the vectors in the earlier figure, find (a) $(\vec{A} \times \vec{F}) \cdot \vec{D}$, (b) $(\vec{A} \times \vec{F}) \cdot (\vec{D} \times \vec{B})$, and (c) $(\vec{A} \cdot \vec{F})(\vec{D} \times \vec{B})$.

Solution:

a. 0, b. 0, c. $+20,000\hat{k}$

Exercise:

Problem:

(a) If $\vec{A} \times \vec{F} = \vec{B} \times \vec{F}$, can we conclude $\vec{A} = \vec{B}$? (b) If $\vec{A} \cdot \vec{F} = \vec{B} \cdot \vec{F}$, can we conclude $\vec{A} = \vec{B}$? (c) If $F\vec{A} = \vec{B}F$, can we conclude $\vec{A} = \vec{B}$? Why or why not?

Additional Problems

Exercise:**Problem:**

You fly 32.0 km in a straight line in still air in the direction 35.0° south of west. (a) Find the distances you would have to fly due south and then due west to arrive at the same point. (b) Find the distances you would have to fly first in a direction 45.0° south of west and then in a direction 45.0° west of north. Note these are the components of the displacement along a different set of axes—namely, the one rotated by 45° with respect to the axes in (a).

Solution:

a. 18.4 km and 26.2 km, b. 31.5 km and 5.56 km

Exercise:**Problem:**

Rectangular coordinates of a point are given by $(2, y)$ and its polar coordinates are given by $(r, \pi/6)$. Find y and r .

Exercise:**Problem:**

If the polar coordinates of a point are (r, φ) and its rectangular coordinates are (x, y) , determine the polar coordinates of the following points: (a) $(-x, y)$, (b) $(-2x, -2y)$, and (c) $(3x, -3y)$.

Solution:

a. $(r, \pi - \varphi)$, b. $(2r, \varphi + 2\pi)$, (c) $(3r, -\varphi)$

Exercise:**Problem:**

Vectors \vec{A} and \vec{B} have identical magnitudes of 5.0 units. Find the angle between them if $\vec{A} + \vec{B} = 5\sqrt{2}\vec{j}$.

Exercise:**Problem:**

Starting at the island of Moi in an unknown archipelago, a fishing boat makes a round trip with two stops at the islands of Noi and Poi. It sails from Moi for 4.76 nautical miles (nmi) in a direction 37° north of east to Noi. From Noi, it sails 69° west of north to Poi. On its return leg from Poi, it sails 28° east of south. What distance does the boat sail between Noi and Poi? What distance does it sail between Moi and Poi? Express your answer both in nautical miles and in kilometers. Note: 1 nmi = 1852 m.

Solution:

$d_{PM} = 6.2 \text{ nmi} = 11.4 \text{ km}$, $d_{NP} = 7.2 \text{ nmi} = 13.3 \text{ km}$

Exercise:

Problem:

An air traffic controller notices two signals from two planes on the radar monitor. One plane is at altitude 800 m and in a 19.2-km horizontal distance to the tower in a direction 25° south of west. The second plane is at altitude 1100 m and its horizontal distance is 17.6 km and 20° south of west. What is the distance between these planes?

Exercise:**Problem:**

Show that when $\vec{A} + \vec{B} = \vec{C}$, then $C^2 = A^2 + B^2 + 2AB \cos \varphi$, where φ is the angle between vectors \vec{A} and \vec{B} .

Solution:

proof

Exercise:**Problem:**

Four force vectors each have the same magnitude f . What is the largest magnitude the resultant force vector may have when these forces are added? What is the smallest magnitude of the resultant? Make a graph of both situations.

Exercise:**Problem:**

A skater glides along a circular path of radius 5.00 m in clockwise direction. When he coasts around one-half of the circle, starting from the west point, find (a) the magnitude of his displacement vector and (b) how far he actually skated. (c) What is the magnitude of his displacement vector when he skates all the way around the circle and comes back to the west point?

Solution:

a. 10.00 m, b. 5π m, c. 0

Exercise:**Problem:**

A stubborn dog is being walked on a leash by its owner. At one point, the dog encounters an interesting scent at some spot on the ground and wants to explore it in detail, but the owner gets impatient and pulls on the leash with force $\vec{F} = (98.0\hat{i} + 132.0\hat{j} + 32.0\hat{k})\text{N}$ along the leash. (a) What is the magnitude of the pulling force? (b) What angle does the leash make with the vertical?

Exercise:**Problem:**

If the velocity vector of a polar bear is $\vec{u} = (-18.0\hat{i} - 13.0\hat{j})\text{km/h}$, how fast and in what geographic direction is it heading? Here, \hat{i} and \hat{j} are directions to geographic east and north, respectively.

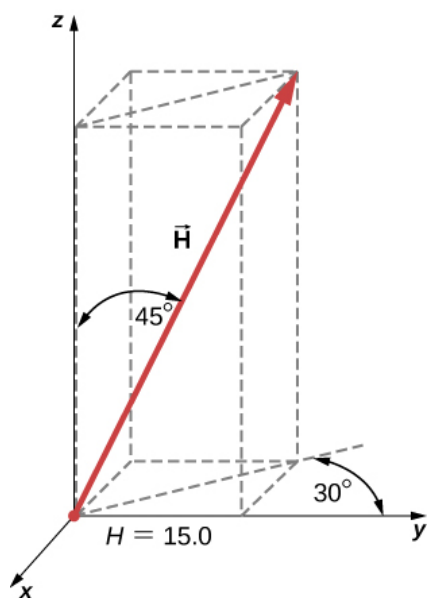
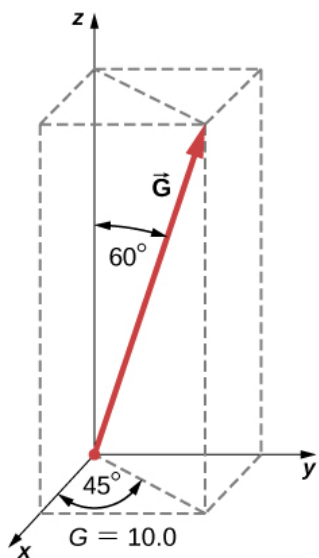
Solution:

22.2 km/h, 35.8° south of west

Exercise:

Problem:

Find the scalar components of three-dimensional vectors \vec{G} and \vec{H} in the following figure and write the vectors in vector component form in terms of the unit vectors of the axes.



Exercise:

Problem:

A diver explores a shallow reef off the coast of Belize. She initially swims 90.0 m north, makes a turn to the east and continues for 200.0 m, then follows a big grouper for 80.0 m in the direction 30° north of east. In the meantime, a local current displaces her by 150.0 m south. Assuming the current is no longer present, in what direction and how far should she now swim to come back to the point where she started?

Solution:

270 m, 4.2° north of west

Exercise:**Problem:**

A force vector \vec{A} has x - and y -components, respectively, of -8.80 units of force and 15.00 units of force. The x - and y -components of force vector \vec{B} are, respectively, 13.20 units of force and -6.60 units of force. Find the components of force vector \vec{C} that satisfies the vector equation $\vec{A} - \vec{B} + 3\vec{C} = 0$.

Exercise:**Problem:**

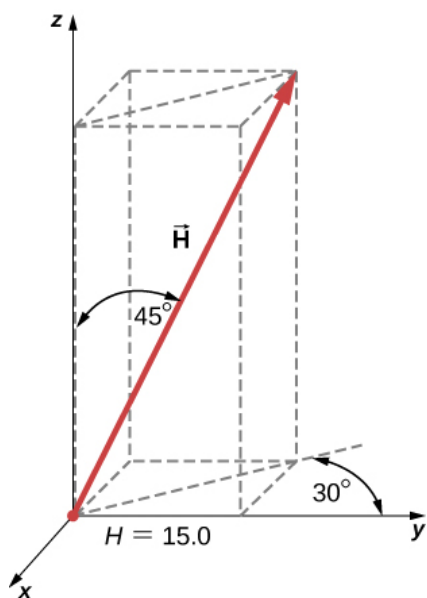
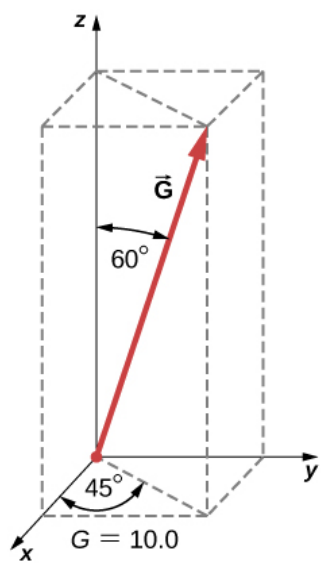
Vectors \vec{A} and \vec{B} are two orthogonal vectors in the xy -plane and they have identical magnitudes. If $\vec{A} = 3.0\hat{i} + 4.0\hat{j}$, find \vec{B} .

Solution:

$$\vec{B} = -4.0\hat{i} + 3.0\hat{j} \text{ or } \vec{B} = 4.0\hat{i} - 3.0\hat{j}$$

Exercise:**Problem:**

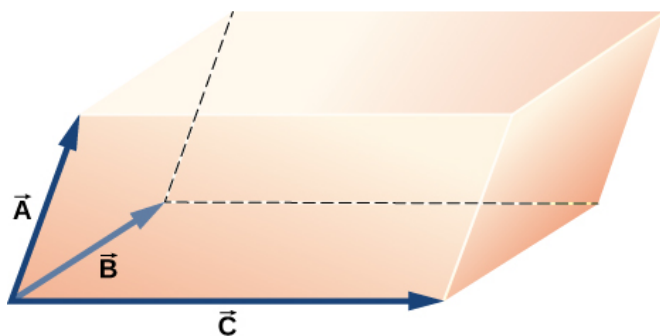
For the three-dimensional vectors in the following figure, find (a) $\vec{G} \times \vec{H}$, (b) $\vec{G} \times \vec{H}$, and (c) $\vec{G} \cdot \vec{H}$.



Exercise:

Problem:

Show that $(\vec{B} \times \vec{C}) \cdot \vec{A}$ is the volume of the parallelepiped, with edges formed by the three vectors in the following figure.



Solution:

proof

Challenge Problems

Exercise:

Problem:

Vector \vec{B} is 5.0 cm long and vector \vec{A} is 4.0 cm long. Find the angle between these two vectors when $\vec{A} + \vec{B} = 3.0$ cm and $\vec{A} - \vec{B} = 3.0$ cm.

Exercise:

Problem:

What is the component of the force vector $\vec{G} = (3.0\hat{i} + 4.0\hat{j} + 10.0\hat{k})\text{N}$ along the force vector $\vec{H} = (1.0\hat{i} + 4.0\hat{j})\text{N}$?

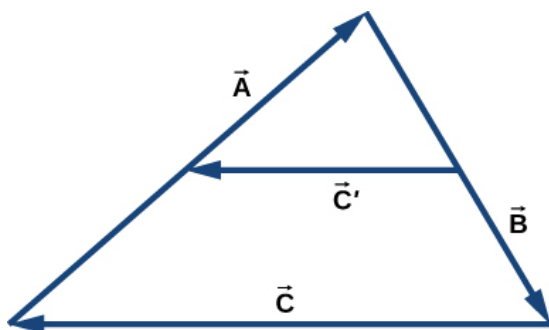
Solution:

$$G_H = 19 \text{ N} / \sqrt{17} \approx 4.6 \text{ N}$$

Exercise:

Problem:

The following figure shows a triangle formed by the three vectors \vec{A} , \vec{B} , and \vec{C} . If vector \vec{C}' is drawn between the midpoints of vectors \vec{A} and \vec{B} , show that $\vec{C}' = \vec{C}/2$.



Exercise:**Problem:**

Distances between points in a plane do not change when a coordinate system is rotated. In other words, the magnitude of a vector is *invariant* under rotations of the coordinate system. Suppose a coordinate system S is rotated about its origin by angle φ to become a new coordinate system S' , as shown in the following figure. A point in a plane has coordinates (x, y) in S and coordinates (x', y') in S' .

(a) Show that, during the transformation of rotation, the coordinates in S' are expressed in terms of the coordinates in S by the following relations:

Equation:

$$\begin{cases} x' = x \cos \varphi + y \sin \varphi \\ y' = -x \sin \varphi + y \cos \varphi \end{cases}$$

(b) Show that the distance of point P to the origin is invariant under rotations of the coordinate system. Here, you have to show that

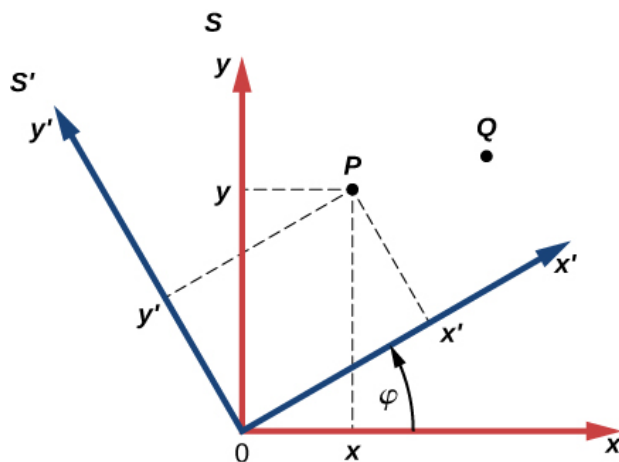
Equation:

$$\sqrt{x^2 + y^2} = \sqrt{x'^2 + y'^2}.$$

(c) Show that the distance between points P and Q is invariant under rotations of the coordinate system. Here, you have to show that

Equation:

$$\sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2} = \sqrt{(x'_P - x'_Q)^2 + (y'_P - y'_Q)^2}.$$

**Solution:**

proof

Glossary

anticommutative property

change in the order of operation introduces the minus sign

corkscrew right-hand rule

a rule used to determine the direction of the vector product

cross product

the result of the vector multiplication of vectors is a vector called a cross product; also called a vector product

dot product

the result of the scalar multiplication of two vectors is a scalar called a dot product; also called a scalar product

scalar product

the result of the scalar multiplication of two vectors is a scalar called a scalar product; also called a dot product

vector product

the result of the vector multiplication of vectors is a vector called a vector product; also called a cross product

Introduction

class="introduction"

A JR Central L0 series five-car maglev (magnetic levitation) train undergoing a test run on the Yamanashi Test Track. The maglev train's motion can be described using kinematics, the subject of this chapter. (credit: modification of work by “Maryland GovPics”/Flickr)



Our universe is full of objects in motion. From the stars, planets, and galaxies; to the motion of people and animals; down to the microscopic scale of atoms and molecules—everything in our universe is in motion. We can describe motion using the two disciplines of kinematics and dynamics. We study dynamics, which is concerned with the causes of motion, in [Newton's Laws of Motion](#); but, there is much to be learned about motion without referring to what causes it, and this is the study of kinematics. Kinematics involves describing motion through properties such as position, time, velocity, and acceleration.

A full treatment of **kinematics** considers motion in two and three dimensions. For now, we discuss motion in one dimension, which provides us with the tools necessary to study multidimensional motion. A good example of an object undergoing one-dimensional motion is the maglev (magnetic levitation) train depicted at the beginning of this chapter. As it travels, say, from Tokyo to Kyoto, it is at different positions along the track at various times in its journey, and therefore has displacements, or changes in position. It also has a variety of velocities along its path and it undergoes accelerations (changes in velocity). With the skills learned in this chapter we can calculate these quantities and average velocity. All these quantities can be described using kinematics, without knowing the train's mass or the forces involved.

Position, Displacement, and Average Velocity

By the end of this section, you will be able to:

- Define position, displacement, and distance traveled.
- Calculate the total displacement given the position as a function of time.
- Determine the total distance traveled.
- Calculate the average velocity given the displacement and elapsed time.

When you're in motion, the basic questions to ask are: Where are you? Where are you going? How fast are you getting there? The answers to these questions require that you specify your position, your displacement, and your average velocity—the terms we define in this section.

Position

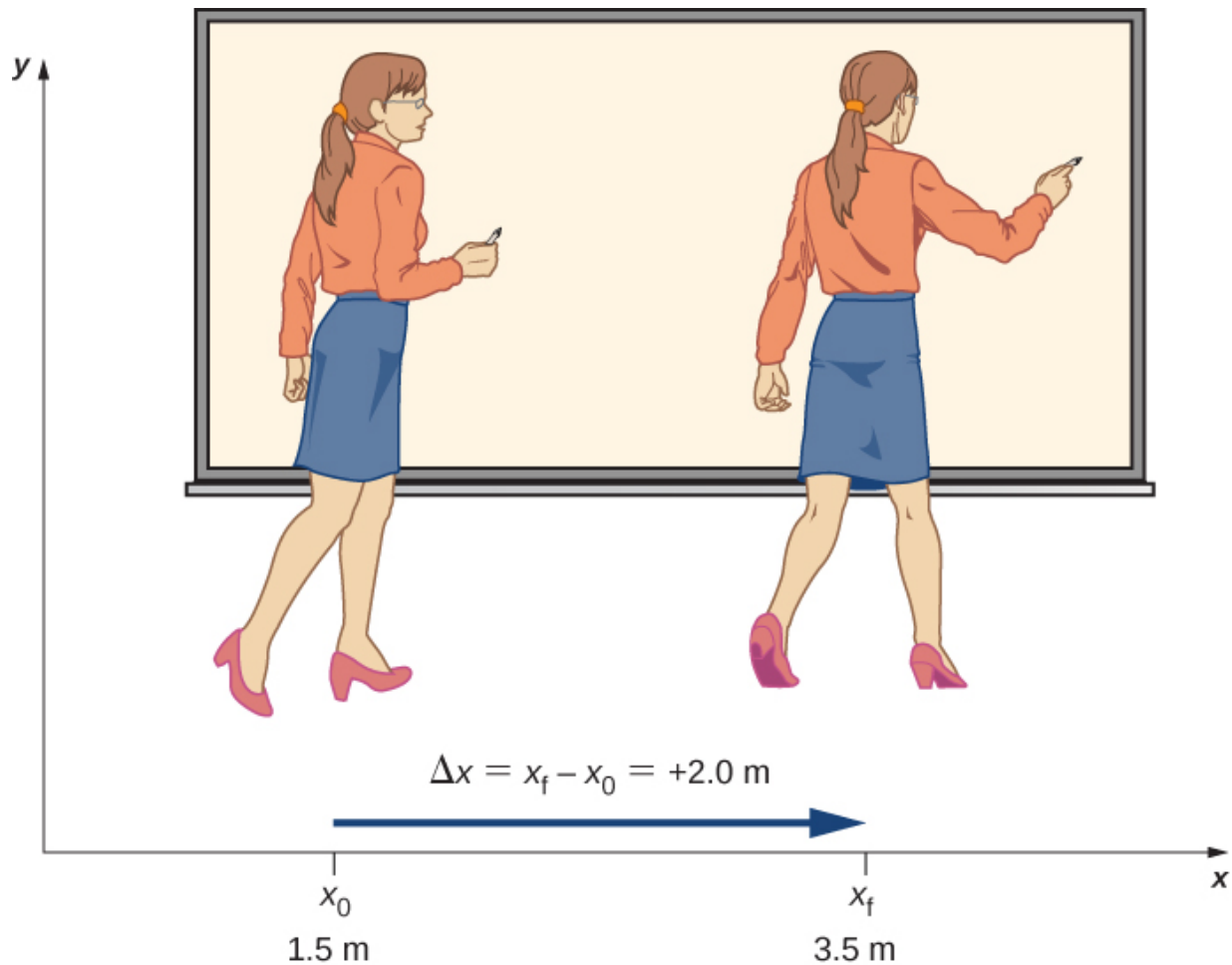
To describe the motion of an object, you must first be able to describe its **position** (x): *where it is at any particular time*. More precisely, we need to specify its position relative to a convenient frame of reference. A frame of reference is an arbitrary set of axes from which the position and motion of an object are described. Earth is often used as a frame of reference, and we often describe the position of an object as it relates to stationary objects on Earth. For example, a rocket launch could be described in terms of the position of the rocket with respect to Earth as a whole, whereas a cyclist's position could be described in terms of where she is in relation to the buildings she passes [\[link\]](#). In other cases, we use reference frames that are not stationary but are in motion relative to Earth. To describe the position of a person in an airplane, for example, we use the airplane, not Earth, as the reference frame. To describe the position of an object undergoing one-dimensional motion, we often use the variable x . Later in the chapter, during the discussion of free fall, we use the variable y .



These cyclists in Vietnam can be described by their position relative to buildings or a canal. Their motion can be described by their change in position, or displacement, in a frame of reference. (credit: modification of work by Suzan Black)

Displacement

If an object moves relative to a frame of reference—for example, if a professor moves to the right relative to a whiteboard [\[link\]](#)—then the object's position changes. This change in position is called **displacement**. The word *displacement* implies that an object has moved, or has been displaced. Although position is the numerical value of x along a straight line where an object might be located, displacement gives the *change* in position along this line. Since displacement indicates direction, it is a vector and can be either positive or negative, depending on the choice of positive direction. Also, an analysis of motion can have many displacements embedded in it. If right is positive and an object moves 2 m to the right, then 4 m to the left, the individual displacements are 2 m and -4 m, respectively.



A professor paces left and right while lecturing. Her position relative to Earth is given by x . The +2.0-m displacement of the professor relative to Earth is represented by an arrow pointing to the right.

Note:

Displacement

Displacement Δx is the change in position of an object:

Equation:

$$\Delta x = x_f - x_0,$$

where Δx is displacement, x_f is the final position, and x_0 is the initial position.

We use the uppercase Greek letter delta (Δ) to mean “change in” whatever quantity follows it; thus, Δx means *change in position* (final position less initial position). We always solve for displacement by subtracting initial position x_0 from final position x_f . Note that the SI unit for displacement is the meter, but sometimes we use kilometers or other units of length. Keep in mind that when units other than meters are used in a problem, you may need to convert them to meters to complete the calculation (see [Appendix B](#)).

Objects in motion can also have a series of displacements. In the previous example of the pacing professor, the individual displacements are 2 m and -4 m, giving a total displacement of -2 m. We define **total displacement** Δx_{Total} , as *the sum of the individual displacements*, and express this mathematically with the equation

Note:

Equation:

$$\Delta x_{\text{Total}} = \sum \Delta x_i,$$

where Δx_i are the individual displacements. In the earlier example,

Equation:

$$\Delta x_1 = x_1 - x_0 = 2 - 0 = 2 \text{ m.}$$

Similarly,

Equation:

$$\Delta x_2 = x_2 - x_1 = -2 - (2) = -4 \text{ m.}$$

Thus,

Equation:

$$\Delta x_{\text{Total}} = \Delta x_1 + \Delta x_2 = 2 - 4 = -2 \text{ m.}$$

The total displacement is $2 - 4 = -2$ m along the x -axis. It is also useful to calculate the magnitude of the displacement, or its size. The magnitude of the displacement is always positive. This is the absolute value of the displacement, because displacement is a vector and cannot have a negative value of magnitude. In our example, the magnitude of the total displacement is 2 m, whereas the magnitudes of the individual displacements are 2 m and 4 m.

The magnitude of the total displacement should not be confused with the distance traveled. Distance traveled x_{Total} , is the total length of the path traveled between two positions. In the previous problem, the **distance traveled** is the sum of the magnitudes of the individual displacements:

Equation:

$$x_{\text{Total}} = |\Delta x_1| + |\Delta x_2| = 2 + 4 = 6 \text{ m.}$$

Average Velocity

To calculate the other physical quantities in kinematics we must introduce the time variable. The time variable allows us not only to state where the object is (its position) during its motion, but also how fast it is moving. How fast an object is moving is given by the rate at which the position changes with time.

For each position x_i , we assign a particular time t_i . If the details of the motion at each instant are not important, the rate is usually expressed as the **average velocity** \bar{v} . This vector quantity is simply the total displacement between two points divided by the time taken to travel between them. The time taken to travel between two points is called the **elapsed time** Δt .

Note:

Average Velocity

If x_1 and x_2 are the positions of an object at times t_1 and t_2 , respectively, then

Equation:

$$\begin{aligned} \text{Average velocity} = \bar{v} &= \frac{\text{Displacement between two points}}{\text{Elapsed time between two points}} \\ \bar{v} &= \frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1}. \end{aligned}$$

It is important to note that the average velocity is a vector and can be negative, depending on positions x_1 and x_2 .

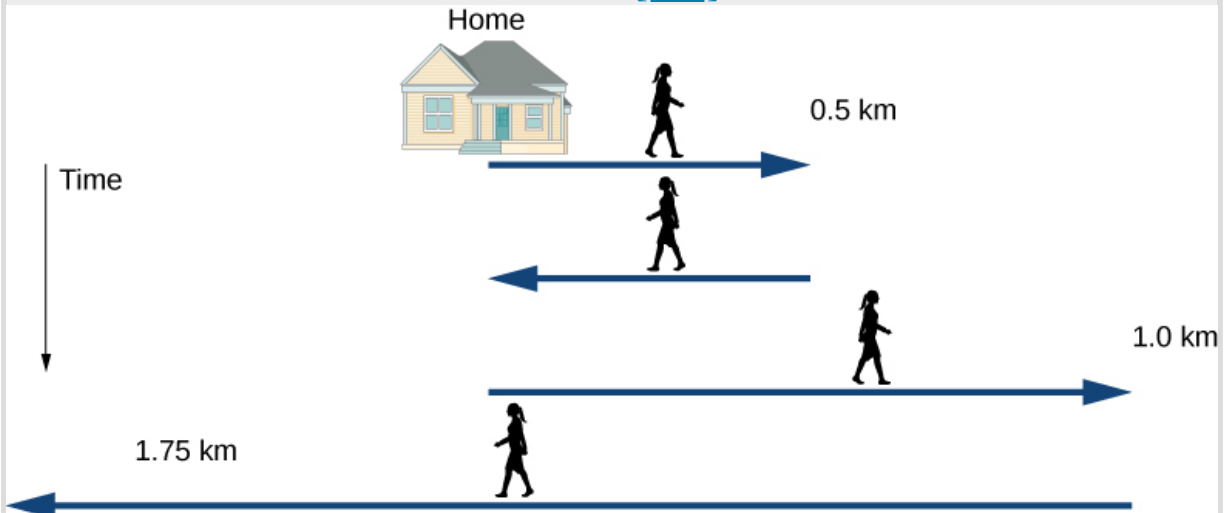
Example:

Delivering Flyers

Jill sets out from her home to deliver flyers for her yard sale, traveling due east along her street lined with houses. At 0.5 km and 9 minutes later she runs out of flyers and has to retrace her steps back to her house to get more. This takes an additional 9 minutes. After picking up more flyers, she sets out again on the same path, continuing where she left off, and ends up 1.0 km from her house. This third leg of her trip takes 15 minutes. At this point she turns back toward her house, heading west. After 1.75 km and 25 minutes she stops to rest.

- What is Jill's total displacement to the point where she stops to rest?
- What is the magnitude of the final displacement?
- What is the average velocity during her entire trip?
- What is the total distance traveled?
- Make a graph of position versus time.

A sketch of Jill's movements is shown in [\[link\]](#).



Timeline of Jill's movements.

Strategy

The problem contains data on the various legs of Jill's trip, so it would be useful to make a table of the physical quantities. We are given position and time in the wording of the problem so we can calculate the displacements and the elapsed time. We take east to be the positive direction. From this information we can find the total displacement and average velocity. Jill's home is the starting point x_0 . The following table gives Jill's time and position in the first two columns, and the displacements are calculated in the third column.

Time t_i (min)	Position x_i (km)	Displacement Δx_i (km)
$t_0 = 0$	$x_0 = 0$	$\Delta x_0 = 0$
$t_1 = 9$	$x_1 = 0.5$	$\Delta x_1 = x_1 - x_0 = 0.5$
$t_2 = 18$	$x_2 = 0$	$\Delta x_2 = x_2 - x_1 = -0.5$
$t_3 = 33$	$x_3 = 1.0$	$\Delta x_3 = x_3 - x_2 = 1.0$
$t_4 = 58$	$x_4 = -0.75$	$\Delta x_4 = x_4 - x_3 = -1.75$

Solution

- a. From the above table, the total displacement is

Equation:

$$\sum \Delta x_i = 0.5 - 0.5 + 1.0 - 1.75 \text{ km} = -0.75 \text{ km}.$$

- b. The magnitude of the total displacement is $|-0.75| \text{ km} = 0.75 \text{ km}$.

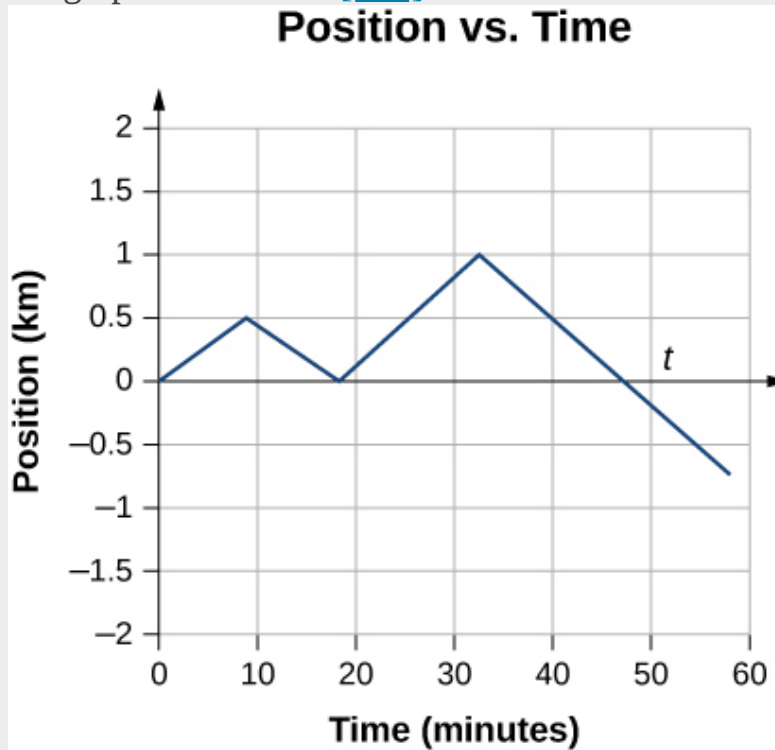
c.

$$\text{Average velocity} = \frac{\text{Total displacement}}{\text{Elapsed time}} = \bar{v} = \frac{-0.75 \text{ km}}{58 \text{ min}} = -0.013 \text{ km/min}$$

- d. The total distance traveled (sum of magnitudes of individual displacements) is

$$x_{\text{Total}} = \sum |\Delta x_i| = 0.5 + 0.5 + 1.0 + 1.75 \text{ km} = 3.75 \text{ km}.$$

- e. We can graph Jill's position versus time as a useful aid to see the motion; the graph is shown in [\[link\]](#).



This graph depicts Jill's position versus time.
The average velocity is the slope of a line connecting the initial and final points.

Significance

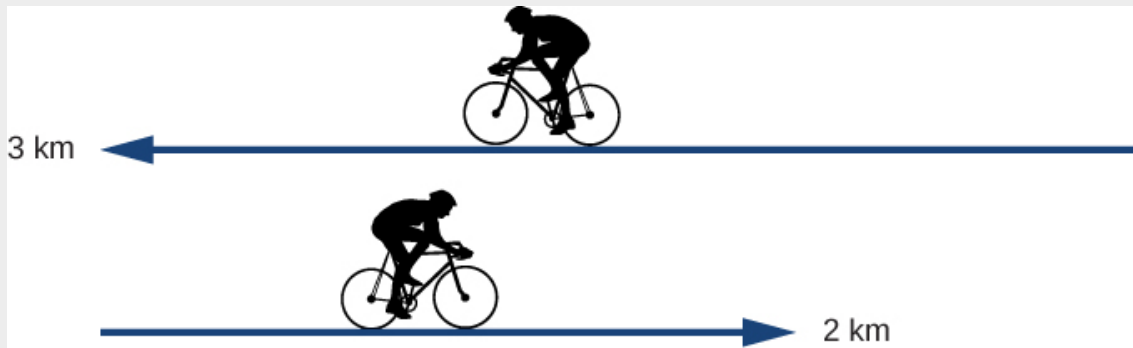
Jill's total displacement is -0.75 km , which means at the end of her trip she ends up 0.75 km due west of her home. The average velocity means if someone was to walk due west at 0.013 km/min starting at the same time Jill left her home, they both would arrive at the final stopping point at the same time. Note that if Jill were to end her trip at her house, her total displacement would be zero, as well as her average velocity. The total distance traveled during the 58 minutes of elapsed time for her trip is 3.75 km .

Note:

Exercise:

Problem:

Check Your Understanding A cyclist rides 3 km west and then turns around and rides 2 km east. (a) What is his displacement? (b) What is the distance traveled? (c) What is the magnitude of his displacement?



Solution:

(a) The rider's displacement is $\Delta x = x_f - x_0 = -1$ km. (The displacement is negative because we take east to be positive and west to be negative.) (b) The distance traveled is $3 \text{ km} + 2 \text{ km} = 5 \text{ km}$. (c) The magnitude of the displacement is 1 km.

Summary

- Kinematics is the description of motion without considering its causes. In this chapter, it is limited to motion along a straight line, called one-dimensional motion.
- Displacement is the change in position of an object. The SI unit for displacement is the meter. Displacement has direction as well as magnitude.
- Distance traveled is the total length of the path traveled between two positions.
- Time is measured in terms of change. The time between two position points x_1 and x_2 is $\Delta t = t_2 - t_1$. Elapsed time for an event is $\Delta t = t_f - t_0$, where t_f is the final time and t_0 is the initial time. The initial time is often taken to be zero.

- Average velocity \bar{v} is defined as displacement divided by elapsed time. If x_1, t_1 and x_2, t_2 are two position time points, the average velocity between these points is

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1}.$$

Conceptual Questions

Exercise:

Problem:

Give an example in which there are clear distinctions among distance traveled, displacement, and magnitude of displacement. Identify each quantity in your example specifically.

Solution:

You drive your car into town and return to drive past your house to a friend's house.

Exercise:

Problem:

Under what circumstances does distance traveled equal magnitude of displacement? What is the only case in which magnitude of displacement and distance are exactly the same?

Exercise:

Problem:

Bacteria move back and forth using their flagella (structures that look like little tails). Speeds of up to $50 \mu\text{m/s}$ ($50 \times 10^{-6} \text{ m/s}$) have been observed. The total distance traveled by a bacterium is large for its size, whereas its displacement is small. Why is this?

Solution:

If the bacteria are moving back and forth, then the displacements are canceling each other and the final displacement is small.

Exercise:

Problem:

Give an example of a device used to measure time and identify what change in that device indicates a change in time.

Exercise:

Problem:

Does a car's odometer measure distance traveled or displacement?

Solution:

Distance traveled

Exercise:

Problem:

During a given time interval the average velocity of an object is zero. What can you conclude about its displacement over the time interval?

Problems

Exercise:

Problem:

Consider a coordinate system in which the positive x axis is directed upward vertically. What are the positions of a particle (a) 5.0 m directly above the origin and (b) 2.0 m below the origin?

Exercise:

Problem:

A car is 2.0 km west of a traffic light at $t = 0$ and 5.0 km east of the light at $t = 6.0$ min. Assume the origin of the coordinate system is the light and the positive x direction is eastward. (a) What are the car's position vectors at these two times? (b) What is the car's displacement between 0 min and 6.0 min?

Solution:

a. $\vec{x}_1 = (-2.0 \text{ m})\hat{i}$, $\vec{x}_2 = (5.0 \text{ m})\hat{i}$; b. 7.0 m east

Exercise:**Problem:**

The Shanghai maglev train connects Longyang Road to Pudong International Airport, a distance of 30 km. The journey takes 8 minutes on average. What is the maglev train's average velocity?

Exercise:**Problem:**

The position of a particle moving along the x -axis is given by $x(t) = 4.0 - 2.0t$ m. (a) At what time does the particle cross the origin? (b) What is the displacement of the particle between $t = 3.0$ s and $t = 6.0$ s?

Solution:

a. $t = 2.0$ s; b. $x(6.0) - x(3.0) = -8.0 - (-2.0) = -6.0$ m

Exercise:**Problem:**

A cyclist rides 8.0 km east for 20 minutes, then he turns and heads west for 8 minutes and 3.2 km. Finally, he rides east for 16 km, which takes 40 minutes. (a) What is the final displacement of the cyclist? (b) What is his average velocity?

Exercise:

Problem:

On February 15, 2013, a superbolide meteor (brighter than the Sun) entered Earth's atmosphere over Chelyabinsk, Russia, and exploded at an altitude of 23.5 km. Eyewitnesses could feel the intense heat from the fireball, and the blast wave from the explosion blew out windows in buildings. The blast wave took approximately 2 minutes 30 seconds to reach ground level. The blast wave traveled at 10° above the horizon. (a) What was the average velocity of the blast wave? b) Compare this with the speed of sound, which is 343 m/s at sea level.

Solution:

a. 150.0 s, $\bar{v} = 156.7 \text{ m/s}$; b. 163% the speed of sound at sea level or about Mach 2.

Glossary

average velocity

the displacement divided by the time over which displacement occurs under constant acceleration

displacement

the change in position of an object

distance traveled

the total length of the path traveled between two positions

elapsed time

the difference between the ending time and the beginning time

kinematics

the description of motion through properties such as position, time, velocity, and acceleration

position

the location of an object at a particular time

total displacement

the sum of individual displacements over a given time period

Instantaneous Velocity and Speed

By the end of this section, you will be able to:

- Explain the difference between average velocity and instantaneous velocity.
- Describe the difference between velocity and speed.
- Calculate the instantaneous velocity given the mathematical equation for the velocity.
- Calculate the speed given the instantaneous velocity.

We have now seen how to calculate the average velocity between two positions. However, since objects in the real world move continuously through space and time, we would like to find the velocity of an object at any single point. We can find the velocity of the object anywhere along its path by using some fundamental principles of calculus. This section gives us better insight into the physics of motion and will be useful in later chapters.

Instantaneous Velocity

The quantity that tells us how fast an object is moving anywhere along its path is the **instantaneous velocity**, usually called simply *velocity*. It is the average velocity between two points on the path in the limit that the time (and therefore the displacement) between the two points approaches zero. To illustrate this idea mathematically, we need to express position x as a continuous function of t denoted by $x(t)$. The expression for the average velocity between two points using this notation is $\bar{v} = \frac{x(t_2) - x(t_1)}{t_2 - t_1}$. To find the instantaneous velocity at any position, we let $t_1 = t$ and $t_2 = t + \Delta t$. After inserting these expressions into the equation for the average velocity and taking the limit as $\Delta t \rightarrow 0$, we find the expression for the instantaneous velocity:

Equation:

$$v(t) = \lim_{\Delta t \rightarrow 0} \frac{x(t + \Delta t) - x(t)}{\Delta t} = \frac{dx(t)}{dt}.$$

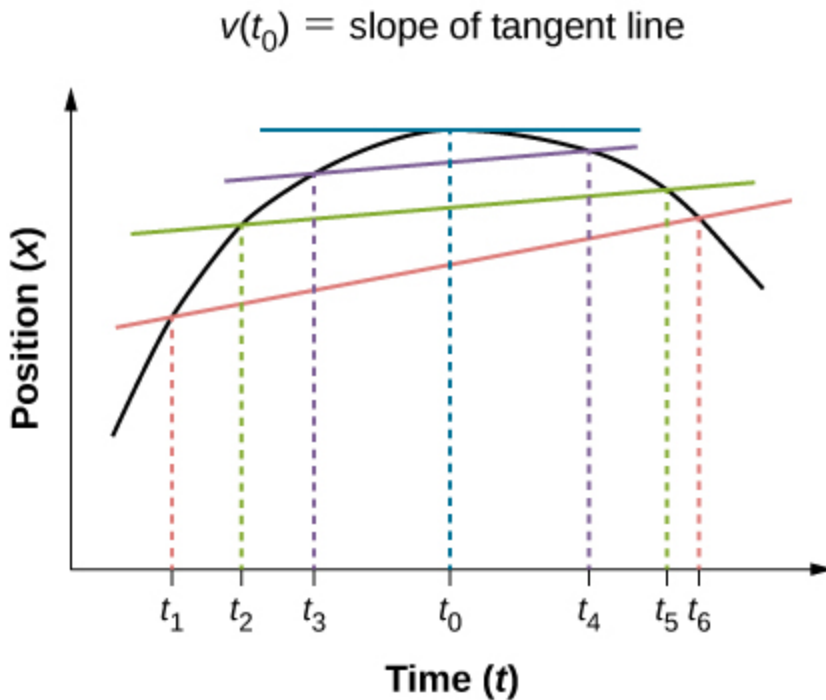
Note:**Instantaneous Velocity**

The instantaneous velocity of an object is the limit of the average velocity as the elapsed time approaches zero, or the derivative of x with respect to t :

Equation:

$$v(t) = \frac{d}{dt}x(t).$$

Like average velocity, instantaneous velocity is a vector with dimension of length per time. The instantaneous velocity at a specific time point t_0 is the rate of change of the position function, which is the slope of the position function $x(t)$ at t_0 . [\[link\]](#) shows how the average velocity $\bar{v} = \frac{\Delta x}{\Delta t}$ between two times approaches the instantaneous velocity at t_0 . The instantaneous velocity is shown at time t_0 , which happens to be at the maximum of the position function. The slope of the position graph is zero at this point, and thus the instantaneous velocity is zero. At other times, t_1 , t_2 , and so on, the instantaneous velocity is not zero because the slope of the position graph would be positive or negative. If the position function had a minimum, the slope of the position graph would also be zero, giving an instantaneous velocity of zero there as well. Thus, the zeros of the velocity function give the minimum and maximum of the position function.



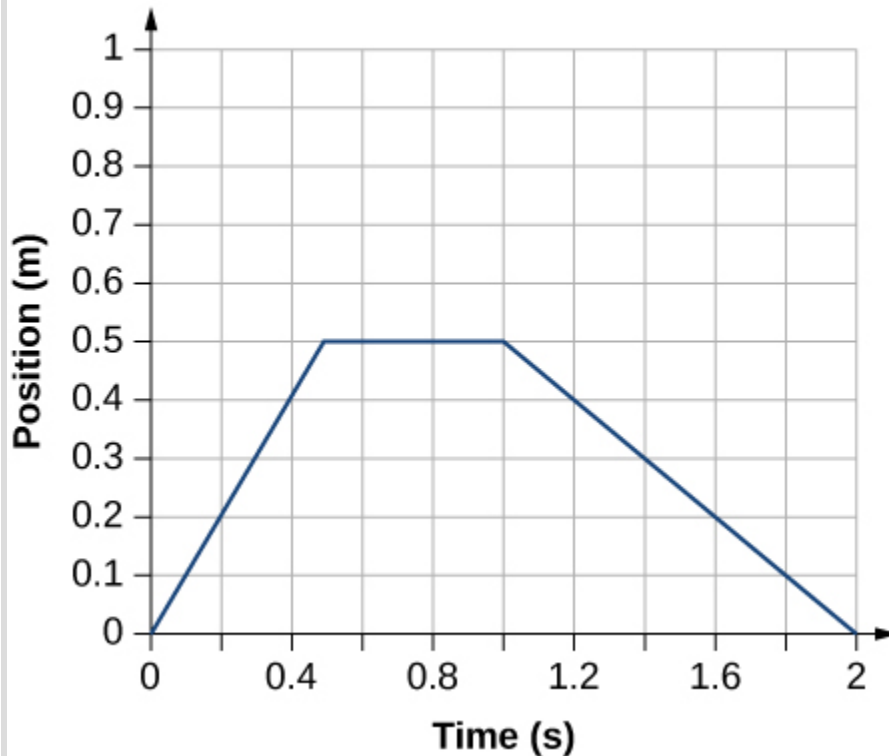
In a graph of position versus time, the instantaneous velocity is the slope of the tangent line at a given point. The average velocities $\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_i}{t_f - t_i}$ between times $\Delta t = t_6 - t_1$, $\Delta t = t_5 - t_2$, and $\Delta t = t_4 - t_3$ are shown. When $\Delta t \rightarrow 0$, the average velocity approaches the instantaneous velocity at $t = t_0$.

Example:

Finding Velocity from a Position-Versus-Time Graph

Given the position-versus-time graph of [\[link\]](#), find the velocity-versus-time graph.

Position vs. Time



The object starts out in the positive direction, stops for a short time, and then reverses direction, heading back toward the origin. Notice that the object comes to rest instantaneously, which would require an infinite force. Thus, the graph is an approximation of motion in the real world. (The concept of force is discussed in [Newton's Laws of Motion](#).)

Strategy

The graph contains three straight lines during three time intervals. We find the velocity during each time interval by taking the slope of the line using the grid.

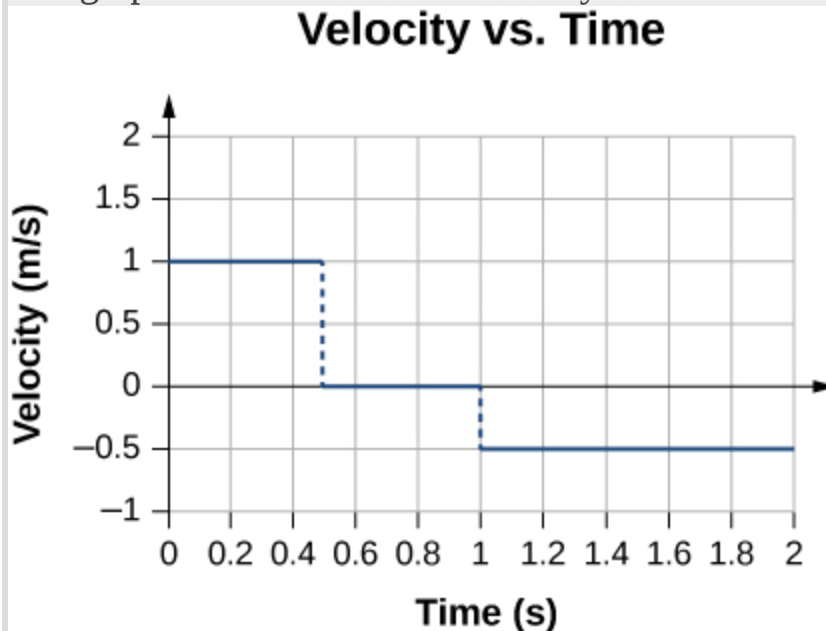
Solution

$$\text{Time interval } 0 \text{ s to } 0.5 \text{ s: } \bar{v} = \frac{\Delta x}{\Delta t} = \frac{0.5 \text{ m} - 0.0 \text{ m}}{0.5 \text{ s} - 0.0 \text{ s}} = 1.0 \text{ m/s}$$

$$\text{Time interval } 0.5 \text{ s to } 1.0 \text{ s: } \bar{v} = \frac{\Delta x}{\Delta t} = \frac{0.5 \text{ m} - 0.5 \text{ m}}{1.0 \text{ s} - 0.5 \text{ s}} = 0.0 \text{ m/s}$$

Time interval 1.0 s to 2.0 s: $\bar{v} = \frac{\Delta x}{\Delta t} = \frac{0.0 \text{ m} - 0.5 \text{ m}}{2.0 \text{ s} - 1.0 \text{ s}} = -0.5 \text{ m/s}$

The graph of these values of velocity versus time is shown in [\[link\]](#).



The velocity is positive for the first part of the trip, zero when the object is stopped, and negative when the object reverses direction.

Significance

During the time interval between 0 s and 0.5 s, the object's position is moving away from the origin and the position-versus-time curve has a positive slope. At any point along the curve during this time interval, we can find the instantaneous velocity by taking its slope, which is +1 m/s, as shown in [\[link\]](#). In the subsequent time interval, between 0.5 s and 1.0 s, the position doesn't change and we see the slope is zero. From 1.0 s to 2.0 s, the object is moving back toward the origin and the slope is -0.5 m/s. The object has reversed direction and has a negative velocity.

Speed

In everyday language, most people use the terms *speed* and *velocity* interchangeably. In physics, however, they do not have the same meaning and are distinct concepts. One major difference is that speed has no direction; that is, speed is a scalar.

We can calculate the **average speed** by finding the total distance traveled divided by the elapsed time:

Note:

Equation:

$$\text{Average speed} = \bar{s} = \frac{\text{Total distance}}{\text{Elapsed time}}.$$

Average speed is not necessarily the same as the magnitude of the average velocity, which is found by dividing the magnitude of the total displacement by the elapsed time. For example, if a trip starts and ends at the same location, the total displacement is zero, and therefore the average velocity is zero. The average speed, however, is not zero, because the total distance traveled is greater than zero. If we take a road trip of 300 km and need to be at our destination at a certain time, then we would be interested in our average speed.

However, we can calculate the **instantaneous speed** from the magnitude of the instantaneous velocity:

Note:

Equation:

$$\text{Instantaneous speed} = |v(t)|.$$

If a particle is moving along the x -axis at $+7.0$ m/s and another particle is moving along the same axis at -7.0 m/s, they have different velocities, but both have the same speed of 7.0 m/s. Some typical speeds are shown in the following table.

Speed	m/s	mi/h
Continental drift	10^{-7}	2×10^{-7}
Brisk walk	1.7	3.9
Cyclist	4.4	10
Sprint runner	12.2	27
Rural speed limit	24.6	56
Official land speed record	341.1	763
Speed of sound at sea level	343	768
Space shuttle on reentry	7800	17,500
Escape velocity of Earth*	11,200	25,000
Orbital speed of Earth around the Sun	29,783	66,623
Speed of light in a vacuum	299,792,458	670,616,629

Speeds of Various Objects*Escape velocity is the velocity at which an object must be launched so that it overcomes Earth's gravity and is not pulled back toward Earth.

Calculating Instantaneous Velocity

When calculating instantaneous velocity, we need to specify the explicit form of the position function $x(t)$. If each term in the $x(t)$ equation has the form of At^n where A is a constant and n is an integer, this can be differentiated using the power rule to be:

Note:

Equation:

$$\frac{d(At^n)}{dt} = Ant^{n-1}.$$

Note that if there are additional terms added together, this power rule of differentiation can be done multiple times and the solution is the sum of those terms. The following example illustrates the use of [\[link\]](#).

Example:

Instantaneous Velocity Versus Average Velocity

The position of a particle is given by $x(t) = 3.0t + 0.5t^3$ m.

- Using [\[link\]](#) and [\[link\]](#), find the instantaneous velocity at $t = 2.0$ s.
- Calculate the average velocity between 1.0 s and 3.0 s.

Strategy

[\[link\]](#) gives the instantaneous velocity of the particle as the derivative of the position function. Looking at the form of the position function given, we see that it is a polynomial in t . Therefore, we can use [\[link\]](#), the power rule from calculus, to find the solution. We use [\[link\]](#) to calculate the average velocity of the particle.

Solution

a. $v(t) = \frac{dx(t)}{dt} = 3.0 + 1.5t^2 \text{ m/s}.$

Substituting $t = 2.0 \text{ s}$ into this equation gives

$$v(2.0 \text{ s}) = [3.0 + 1.5(2.0)^2] \text{ m/s} = 9.0 \text{ m/s}.$$

- b. To determine the average velocity of the particle between 1.0 s and 3.0 s , we calculate the values of $x(1.0 \text{ s})$ and $x(3.0 \text{ s})$:

Equation:

$$x(1.0 \text{ s}) = [(3.0)(1.0) + 0.5(1.0)^3] \text{ m} = 3.5 \text{ m}$$

Equation:

$$x(3.0 \text{ s}) = [(3.0)(3.0) + 0.5(3.0)^3] \text{ m} = 22.5 \text{ m}.$$

Then the average velocity is

Equation:

$$\bar{v} = \frac{x(3.0 \text{ s}) - x(1.0 \text{ s})}{t(3.0 \text{ s}) - t(1.0 \text{ s})} = \frac{22.5 - 3.5 \text{ m}}{3.0 - 1.0 \text{ s}} = 9.5 \text{ m/s}.$$

Significance

In the limit that the time interval used to calculate \bar{v} goes to zero, the value obtained for \bar{v} converges to the value of v .

Example:

Instantaneous Velocity Versus Speed

Consider the motion of a particle in which the position is

$$x(t) = 3.0t - 3t^2 \text{ m}.$$

- What is the instantaneous velocity at $t = 0.25 \text{ s}$, $t = 0.50 \text{ s}$, and $t = 1.0 \text{ s}$?
- What is the speed of the particle at these times?

Strategy

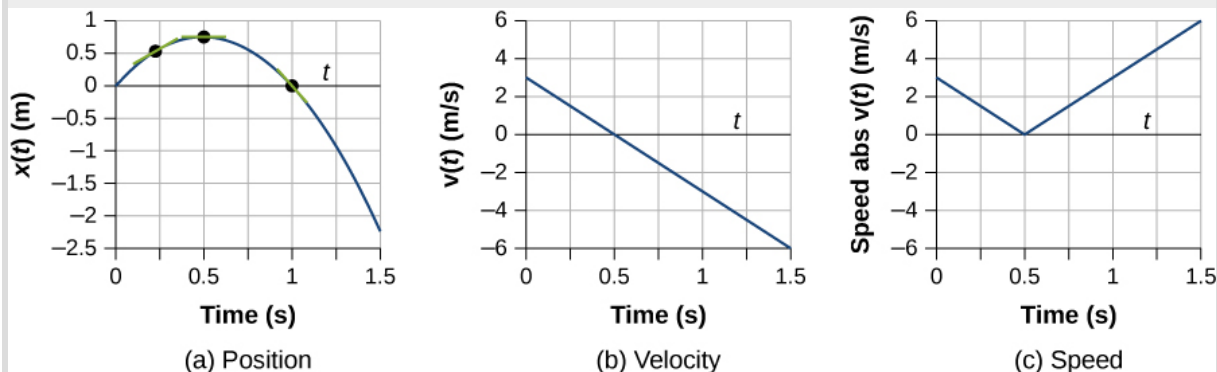
The instantaneous velocity is the derivative of the position function and the speed is the magnitude of the instantaneous velocity. We use [\[link\]](#) and [\[link\]](#) to solve for instantaneous velocity.

Solution

a. $v(t) = \frac{dx(t)}{dt} = 3.0 - 6.0t \text{ m/s}$
 $v(0.25 \text{ s}) = 1.50 \text{ m/s}$, $v(0.5 \text{ s}) = 0 \text{ m/s}$, $v(1.0 \text{ s}) = -3.0 \text{ m/s}$
b. Speed = $|v(t)| = 1.50 \text{ m/s}$, 0.0 m/s , and 3.0 m/s

Significance

The velocity of the particle gives us direction information, indicating the particle is moving to the left (west) or right (east). The speed gives the magnitude of the velocity. By graphing the position, velocity, and speed as functions of time, we can understand these concepts visually [\[link\]](#). In (a), the graph shows the particle moving in the positive direction until $t = 0.5 \text{ s}$, when it reverses direction. The reversal of direction can also be seen in (b) at 0.5 s where the velocity is zero and then turns negative. At 1.0 s it is back at the origin where it started. The particle's velocity at 1.0 s in (b) is negative, because it is traveling in the negative direction. But in (c), however, its speed is positive and remains positive throughout the travel time. We can also interpret velocity as the slope of the position-versus-time graph. The slope of $x(t)$ is decreasing toward zero, becoming zero at 0.5 s and increasingly negative thereafter. This analysis of comparing the graphs of position, velocity, and speed helps catch errors in calculations. The graphs must be consistent with each other and help interpret the calculations.



(a) Position: $x(t)$ versus time. (b) Velocity: $v(t)$ versus time. The slope

of the position graph is the velocity. A rough comparison of the slopes of the tangent lines in (a) at 0.25 s, 0.5 s, and 1.0 s with the values for velocity at the corresponding times indicates they are the same values.

(c) Speed: $|v(t)|$ versus time. Speed is always a positive number.

Note:

Exercise:

Problem:

Check Your Understanding The position of an object as a function of time is $x(t) = -3t^2$ m. (a) What is the velocity of the object as a function of time? (b) Is the velocity ever positive? (c) What are the velocity and speed at $t = 1.0$ s?

Solution:

(a) Taking the derivative of $x(t)$ gives $v(t) = -6t$ m/s. (b) No, because time can never be negative. (c) The velocity is $v(1.0 \text{ s}) = -6$ m/s and the speed is $|v(1.0 \text{ s})| = 6$ m/s.

Summary

- Instantaneous velocity is a continuous function of time and gives the velocity at any point in time during a particle's motion. We can calculate the instantaneous velocity at a specific time by taking the derivative of the position function, which gives us the functional form of instantaneous velocity $v(t)$.
- Instantaneous velocity is a vector and can be negative.
- Instantaneous speed is found by taking the absolute value of instantaneous velocity, and it is always positive.

- Average speed is total distance traveled divided by elapsed time.
- The slope of a position-versus-time graph at a specific time gives instantaneous velocity at that time.

Conceptual Questions

Exercise:

Problem:

There is a distinction between average speed and the magnitude of average velocity. Give an example that illustrates the difference between these two quantities.

Solution:

Average speed is the total distance traveled divided by the elapsed time. If you go for a walk, leaving and returning to your home, your average speed is a positive number. Since $\text{Average velocity} = \text{Displacement} / \text{Elapsed time}$, your average velocity is zero.

Exercise:

Problem: Does the speedometer of a car measure speed or velocity?

Exercise:

Problem:

If you divide the total distance traveled on a car trip (as determined by the odometer) by the elapsed time of the trip, are you calculating average speed or magnitude of average velocity? Under what circumstances are these two quantities the same?

Solution:

Average speed. They are the same if the car doesn't reverse direction.

Exercise:

Problem:

How are instantaneous velocity and instantaneous speed related to one another? How do they differ?

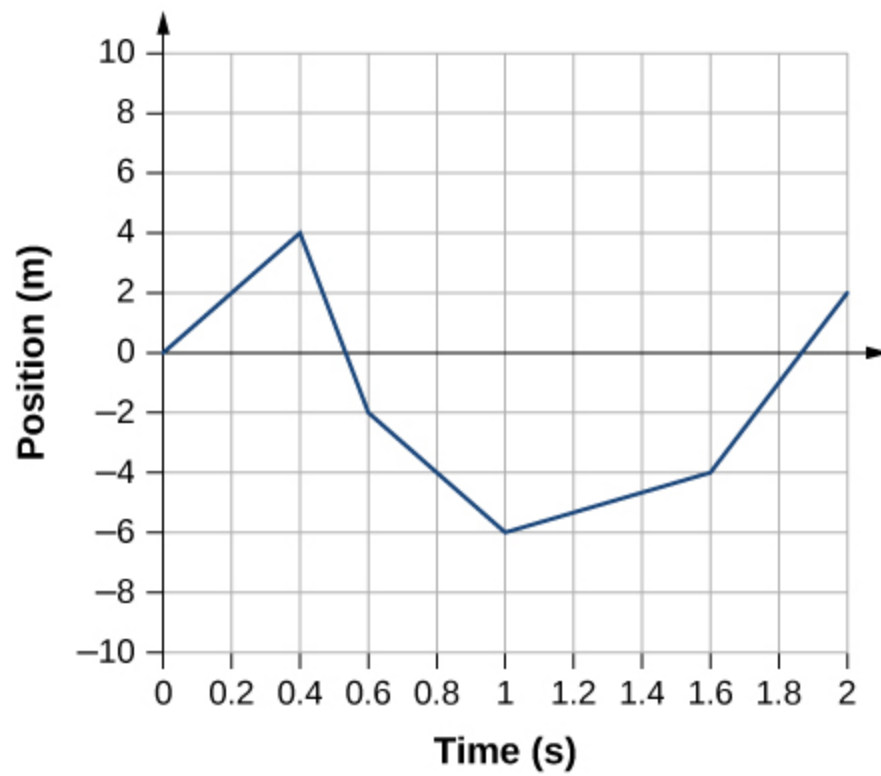
Problems**Exercise:****Problem:**

A woodchuck runs 20 m to the right in 5 s, then turns and runs 10 m to the left in 3 s. (a) What is the average velocity of the woodchuck? (b) What is its average speed?

Exercise:**Problem:**

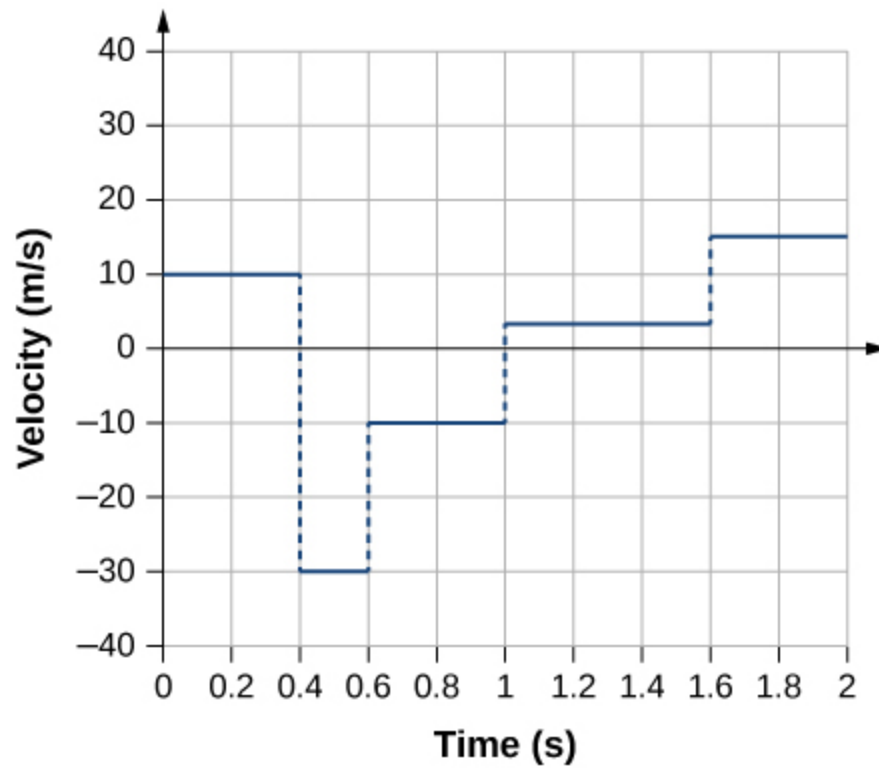
Sketch the velocity-versus-time graph from the following position-versus-time graph.

Position vs. Time



Solution:

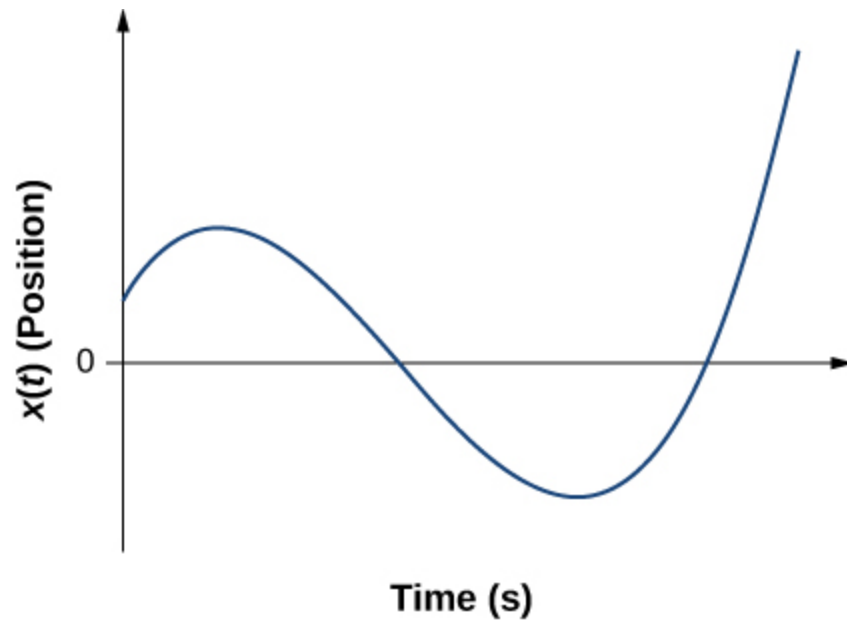
Velocity vs. Time



Exercise:

Problem:

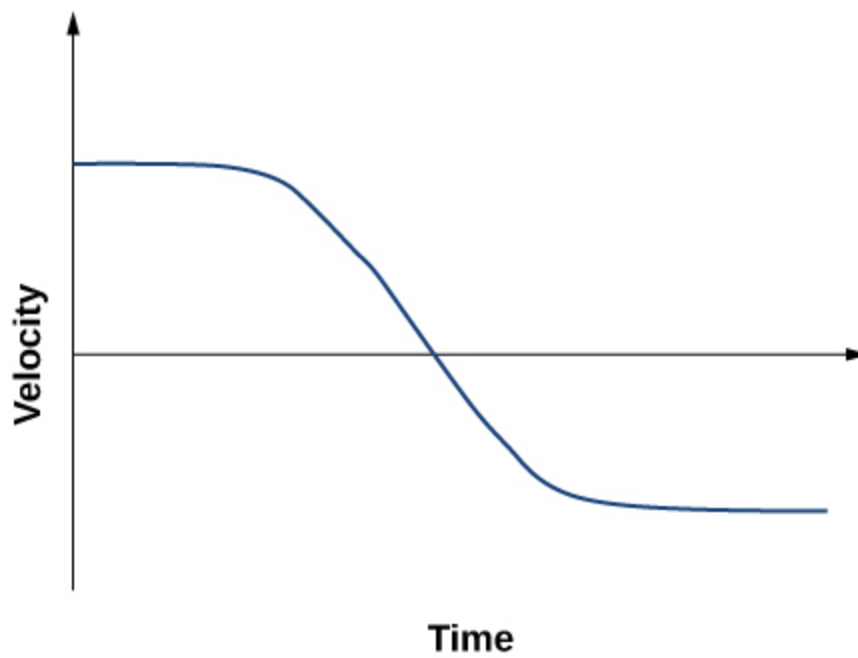
Sketch the velocity-versus-time graph from the following position-versus-time graph.



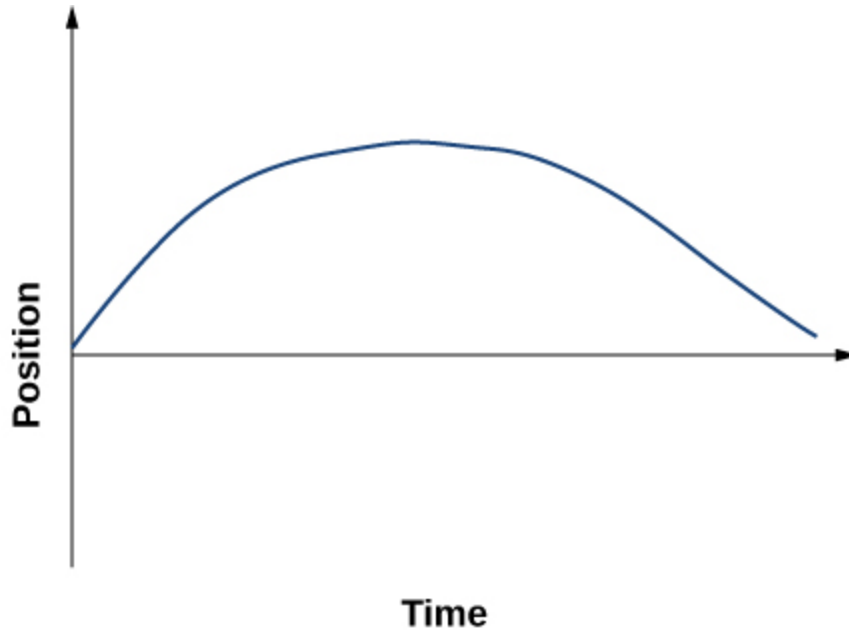
Exercise:

Problem:

Given the following velocity-versus-time graph, sketch the position-versus-time graph.



Solution:



Exercise:

Problem:

An object has a position function $x(t) = 5t$ m. (a) What is the velocity as a function of time? (b) Graph the position function and the velocity function.

Exercise:

Problem:

A particle moves along the x -axis according to $x(t) = 10t - 2t^2$ m. (a) What is the instantaneous velocity at $t = 2$ s and $t = 3$ s? (b) What is the instantaneous speed at these times? (c) What is the average velocity between $t = 2$ s and $t = 3$ s?

Solution:

a. $v(t) = (10 - 4t)$ m/s; $v(2 \text{ s}) = 2 \text{ m/s}$, $v(3 \text{ s}) = -2 \text{ m/s}$; b. $|v(2 \text{ s})| = 2 \text{ m/s}$, $|v(3 \text{ s})| = 2 \text{ m/s}$; (c) $\bar{v} = 0 \text{ m/s}$

Exercise:

Problem:

Unreasonable results. A particle moves along the x -axis according to $x(t) = 3t^3 + 5t$. At what time is the velocity of the particle equal to zero? Is this reasonable?

Glossary

instantaneous velocity

the velocity at a specific instant or time point

instantaneous speed

the absolute value of the instantaneous velocity

average speed

the total distance traveled divided by elapsed time

Average and Instantaneous Acceleration

By the end of this section, you will be able to:

- Calculate the average acceleration between two points in time.
- Calculate the instantaneous acceleration given the functional form of velocity.
- Explain the vector nature of instantaneous acceleration and velocity.
- Explain the difference between average acceleration and instantaneous acceleration.
- Find instantaneous acceleration at a specified time on a graph of velocity versus time.

The importance of understanding acceleration spans our day-to-day experience, as well as the vast reaches of outer space and the tiny world of subatomic physics. In everyday conversation, to *accelerate* means to speed up; applying the brake pedal causes a vehicle to slow down. We are familiar with the acceleration of our car, for example. The greater the acceleration, the greater the change in velocity over a given time. Acceleration is widely seen in experimental physics. In linear particle accelerator experiments, for example, subatomic particles are accelerated to very high velocities in collision experiments, which tell us information about the structure of the subatomic world as well as the origin of the universe. In space, cosmic rays are subatomic particles that have been accelerated to very high energies in supernovas (exploding massive stars) and active galactic nuclei. It is important to understand the processes that accelerate cosmic rays because these rays contain highly penetrating radiation that can damage electronics flown on spacecraft, for example.

Average Acceleration

The formal definition of acceleration is consistent with these notions just described, but is more inclusive.

Note:

Average Acceleration

Average acceleration is the rate at which velocity changes:

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0},$$

where \bar{a} is **average acceleration**, v is velocity, and t is time. (The bar over the a means *average* acceleration.)

Because acceleration is velocity in meters per second divided by time in seconds, the SI units for acceleration are often abbreviated m/s^2 —that is, meters per second squared or meters per second per second. This literally means by how many meters per second the velocity changes every second. Recall that velocity is a vector—it has both magnitude and direction—which means that a change in velocity can be a change in magnitude (or speed), but it can also be a change in direction. For example, if a runner traveling at 10 km/h due east slows to a stop, reverses direction, and continues her run at 10 km/h due west, her velocity has changed as a result of the change in direction, although the *magnitude* of the velocity is the same in both directions. Thus, acceleration occurs when velocity changes in magnitude (an increase or decrease in speed) or in direction, or both.

Note:

Acceleration as a Vector

Acceleration is a vector in the same direction as the *change* in velocity, Δv . Since velocity is a vector, it can change in magnitude or in direction, or both. Acceleration is, therefore, a change in speed or direction, or both.

Keep in mind that although acceleration is in the direction of the change in velocity, it is not always in the direction of motion. When an object slows down, its acceleration is opposite to the direction of its motion. Although

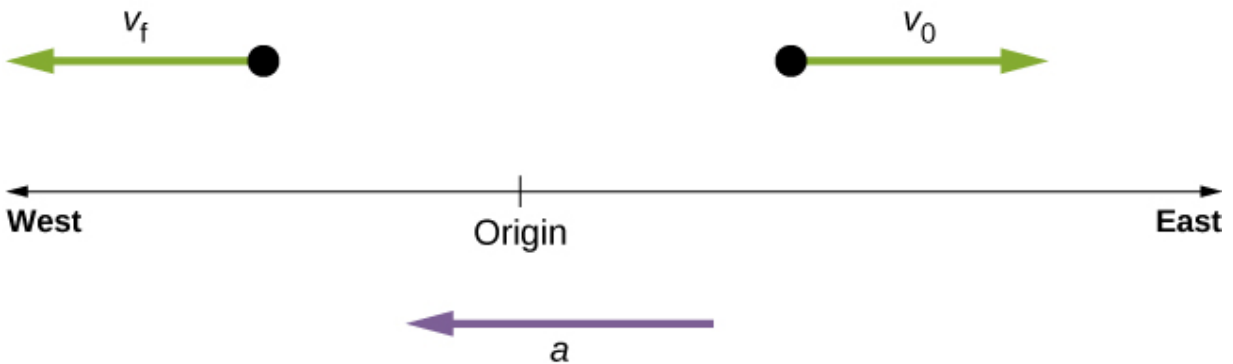
this is commonly referred to as *deceleration* [\[link\]](#), we say the train is accelerating in a direction opposite to its direction of motion.



A subway train in Sao Paulo, Brazil, accelerates opposite to the motion as it comes into a station. It is accelerating in a direction opposite to its direction of motion. (credit: modification of work by Yusuke Kawasaki)

The term *deceleration* can cause confusion in our analysis because it is not a vector and it does not point to a specific direction with respect to a coordinate system, so we do not use it. Acceleration is a vector, so we must choose the appropriate sign for it in our chosen coordinate system. In the case of the train in [\[link\]](#), acceleration is *in the negative direction in the chosen coordinate system*, so we say the train is undergoing negative acceleration.

If an object in motion has a velocity in the positive direction with respect to a chosen origin and it acquires a constant negative acceleration, the object eventually comes to a rest and reverses direction. If we wait long enough, the object passes through the origin going in the opposite direction. This is illustrated in [\[link\]](#).



An object in motion with a velocity vector toward the east under negative acceleration comes to a rest and reverses direction. It passes the origin going in the opposite direction after a long enough time.

Example:

Calculating Average Acceleration: A Racehorse Leaves the Gate

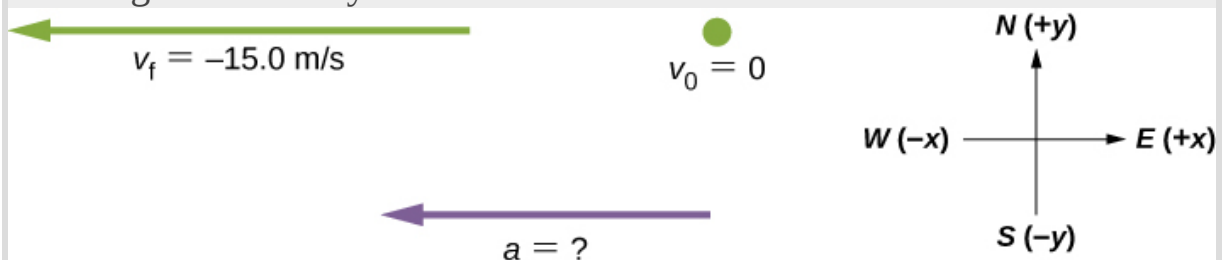
A racehorse coming out of the gate accelerates from rest to a velocity of 15.0 m/s due west in 1.80 s. What is its average acceleration?



Racehorses accelerating out of the gate. (credit: modification of work by Jon Sullivan)

Strategy

First we draw a sketch and assign a coordinate system to the problem [\[link\]](#). This is a simple problem, but it always helps to visualize it. Notice that we assign east as positive and west as negative. Thus, in this case, we have negative velocity.



Identify the coordinate system, the given information, and what you want to determine.

We can solve this problem by identifying Δv and Δt from the given information, and then calculating the average acceleration directly from the

equation $\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}$.

Solution

First, identify the knowns: $v_0 = 0$, $v_f = -15.0 \text{ m/s}$ (the negative sign indicates direction toward the west), $\Delta t = 1.80 \text{ s}$.

Second, find the change in velocity. Since the horse is going from zero to -15.0 m/s , its change in velocity equals its final velocity:

Equation:

$$\Delta v = v_f - v_0 = v_f = -15.0 \text{ m/s}.$$

Last, substitute the known values (Δv and Δt) and solve for the unknown \bar{a} :

Equation:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-15.0 \text{ m/s}}{1.80 \text{ s}} = -8.33 \text{ m/s}^2.$$

Significance

The negative sign for acceleration indicates that acceleration is toward the west. An acceleration of 8.33 m/s^2 due west means the horse increases its velocity by 8.33 m/s due west each second; that is, $8.33 \text{ meters per second per second}$, which we write as 8.33 m/s^2 . This is truly an average acceleration, because the ride is not smooth. We see later that an acceleration of this magnitude would require the rider to hang on with a force nearly equal to his weight.

Note:

Exercise:

Problem:

Check Your Understanding Protons in a linear accelerator are accelerated from rest to $2.0 \times 10^7 \text{ m/s}$ in 10^{-4} s . What is the average acceleration of the protons?

Solution:

Inserting the knowns, we have

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{2.0 \times 10^7 \text{ m/s} - 0}{10^{-4} \text{ s} - 0} = 2.0 \times 10^{11} \text{ m/s}^2.$$

Instantaneous Acceleration

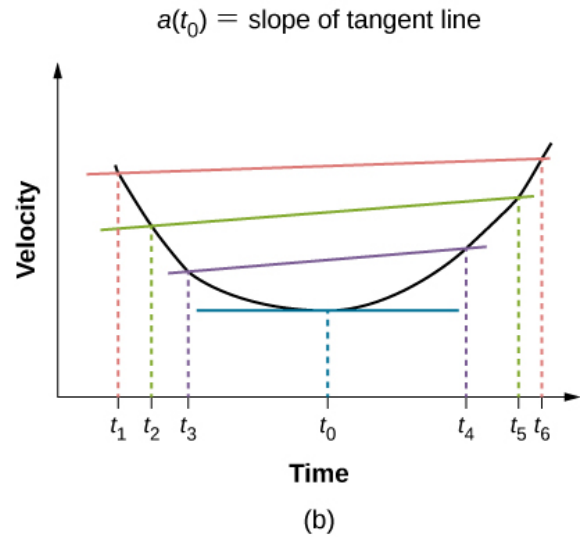
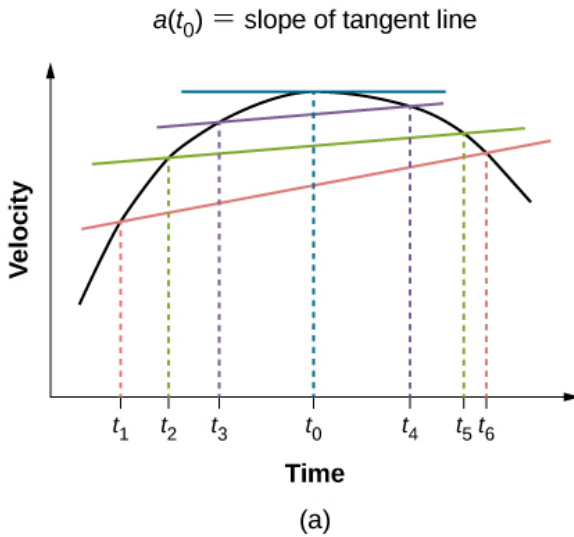
Instantaneous acceleration a , or *acceleration at a specific instant in time*, is obtained using the same process discussed for instantaneous velocity. That is, we calculate the average acceleration between two points in time separated by Δt and let Δt approach zero. The result is the derivative of the velocity function $v(t)$, which is **instantaneous acceleration** and is expressed mathematically as

Note:

Equation:

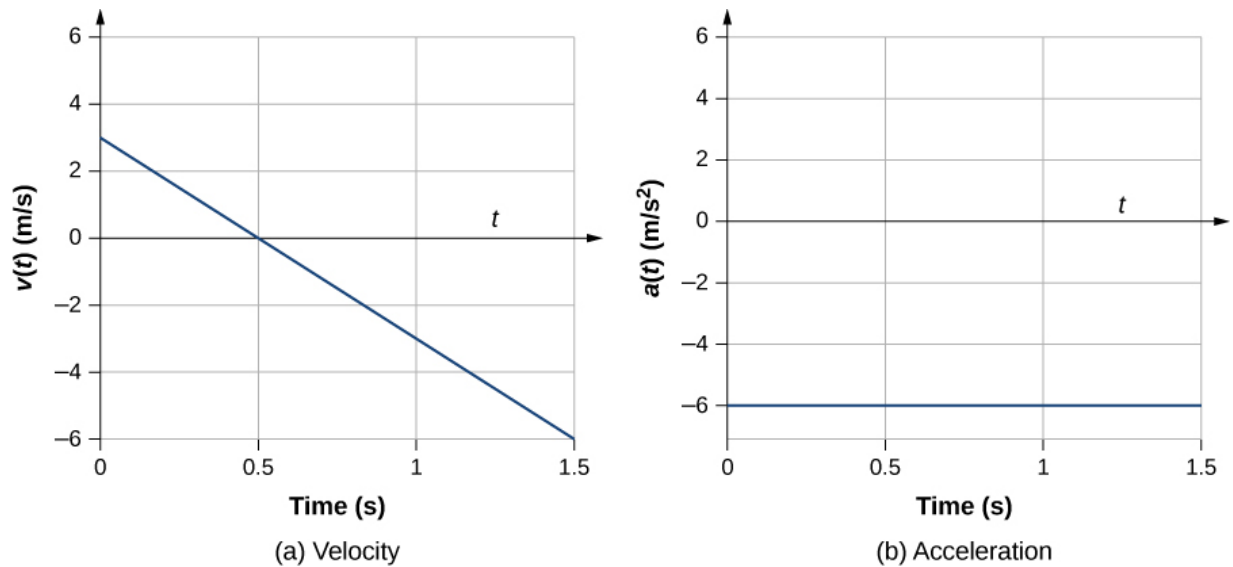
$$a(t) = \frac{d}{dt}v(t).$$

Thus, similar to velocity being the derivative of the position function, instantaneous acceleration is the derivative of the velocity function. We can show this graphically in the same way as instantaneous velocity. In [\[link\]](#), instantaneous acceleration at time t_0 is the slope of the tangent line to the velocity-versus-time graph at time t_0 . We see that average acceleration $\bar{a} = \frac{\Delta v}{\Delta t}$ approaches instantaneous acceleration as Δt approaches zero. Also in part (a) of the figure, we see that velocity has a maximum when its slope is zero. This time corresponds to the zero of the acceleration function. In part (b), instantaneous acceleration at the minimum velocity is shown, which is also zero, since the slope of the curve is zero there, too. Thus, for a given velocity function, the zeros of the acceleration function give either the minimum or the maximum velocity.



In a graph of velocity versus time, instantaneous acceleration is the slope of the tangent line. (a) Shown is average acceleration $\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_i}{t_f - t_i}$ between times $\Delta t = t_6 - t_1$, $\Delta t = t_5 - t_2$, and $\Delta t = t_4 - t_3$. When $\Delta t \rightarrow 0$, the average acceleration approaches instantaneous acceleration at time t_0 . In view (a), instantaneous acceleration is shown for the point on the velocity curve at maximum velocity. At this point, instantaneous acceleration is the slope of the tangent line, which is zero. At any other time, the slope of the tangent line—and thus instantaneous acceleration—would not be zero. (b) Same as (a) but shown for instantaneous acceleration at minimum velocity.

To illustrate this concept, let's look at two examples. First, a simple example is shown using [\[link\]](#)(b), the velocity-versus-time graph of [\[link\]](#), to find acceleration graphically. This graph is depicted in [\[link\]](#)(a), which is a straight line. The corresponding graph of acceleration versus time is found from the slope of velocity and is shown in [\[link\]](#)(b). In this example, the velocity function is a straight line with a constant slope, thus acceleration is a constant. In the next example, the velocity function has a more complicated functional dependence on time.



(a, b) The velocity-versus-time graph is linear and has a negative constant slope (a) that is equal to acceleration, shown in (b).

If we know the functional form of velocity, $v(t)$, we can calculate instantaneous acceleration $a(t)$ at any time point in the motion using [\[link\]](#).

Example:

Calculating Instantaneous Acceleration

A particle is in motion and is accelerating. The functional form of the velocity is $v(t) = 20t - 5t^2$ m/s.

- Find the functional form of the acceleration.
- Find the instantaneous velocity at $t = 1, 2, 3$, and 5 s.
- Find the instantaneous acceleration at $t = 1, 2, 3$, and 5 s.
- Interpret the results of (c) in terms of the directions of the acceleration and velocity vectors.

Strategy

We find the functional form of acceleration by taking the derivative of the velocity function. Then, we calculate the values of instantaneous velocity

and acceleration from the given functions for each. For part (d), we need to compare the directions of velocity and acceleration at each time.

Solution

a. $a(t) = \frac{dv(t)}{dt} = 20 - 10t \text{ m/s}^2$

b. $v(1 \text{ s}) = 15 \text{ m/s}$, $v(2 \text{ s}) = 20 \text{ m/s}$, $v(3 \text{ s}) = 15 \text{ m/s}$,
 $v(5 \text{ s}) = -25 \text{ m/s}$

c. $a(1 \text{ s}) = 10 \text{ m/s}^2$, $a(2 \text{ s}) = 0 \text{ m/s}^2$, $a(3 \text{ s}) = -10 \text{ m/s}^2$,
 $a(5 \text{ s}) = -30 \text{ m/s}^2$

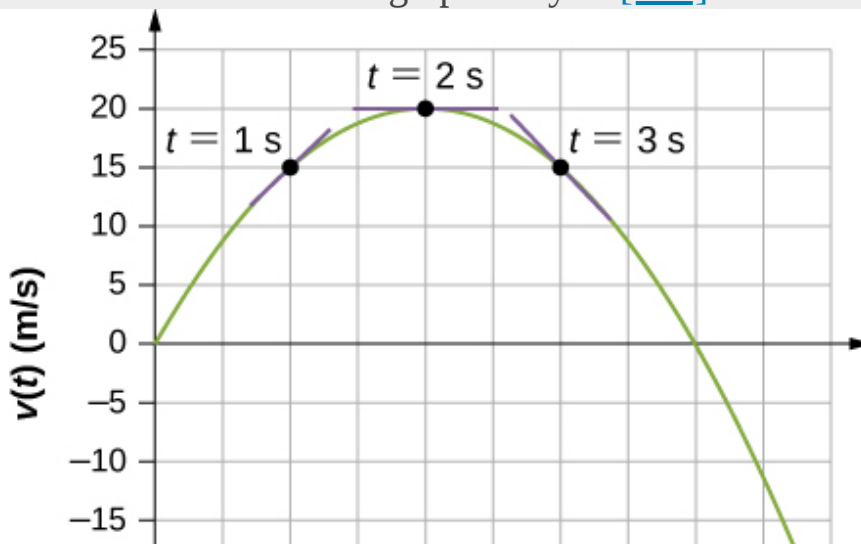
d. At $t = 1 \text{ s}$, velocity $v(1 \text{ s}) = 15 \text{ m/s}$ is positive and acceleration is positive, so both velocity and acceleration are in the same direction. The particle is moving faster.

At $t = 2 \text{ s}$, velocity has increased to $v(2 \text{ s}) = 20 \text{ m/s}$, where it is maximum, which corresponds to the time when the acceleration is zero. We see that the maximum velocity occurs when the slope of the velocity function is zero, which is just the zero of the acceleration function.

At $t = 3 \text{ s}$, velocity is $v(3 \text{ s}) = 15 \text{ m/s}$ and acceleration is negative. The particle has reduced its velocity and the acceleration vector is negative. The particle is slowing down.

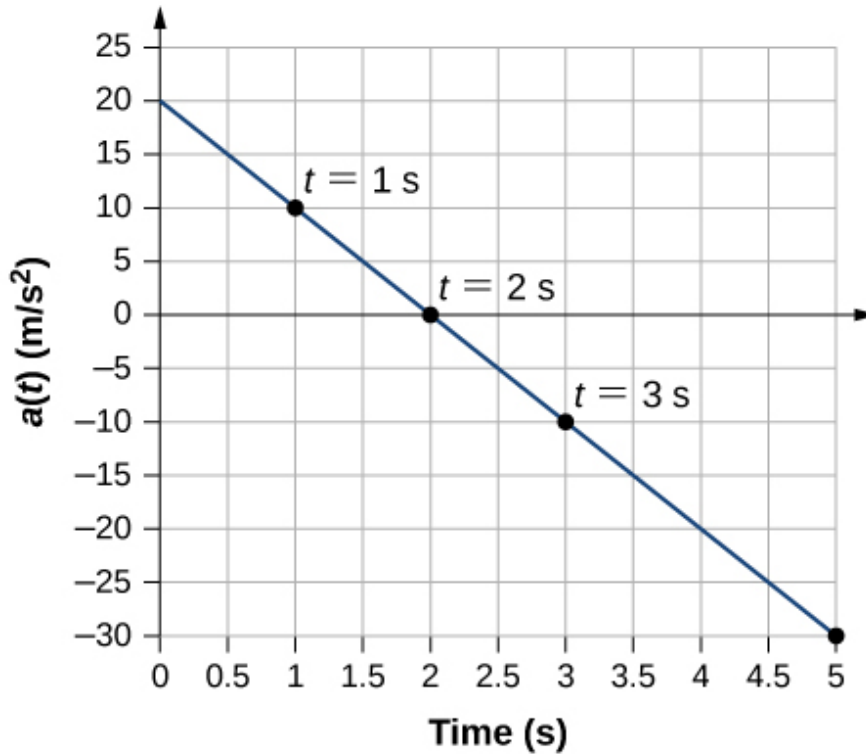
At $t = 5 \text{ s}$, velocity is $v(5 \text{ s}) = -25 \text{ m/s}$ and acceleration is increasingly negative. Between the times $t = 3 \text{ s}$ and $t = 5 \text{ s}$ the particle has decreased its velocity to zero and then become negative, thus reversing its direction. The particle is now speeding up again, but in the opposite direction.

We can see these results graphically in [\[link\]](#).





(a) Velocity



(b) Acceleration

(a) Velocity versus time. Tangent lines are indicated at times 1, 2, and 3 s. The slopes of the tangent lines are the accelerations. At $t = 3$ s, velocity is positive. At $t = 5$ s, velocity is negative, indicating the particle has reversed direction. (b) Acceleration versus time. Comparing the values of accelerations given by the black dots with the corresponding slopes of the tangent lines (slopes of lines through black dots) in (a), we see they are identical.

Significance

By doing both a numerical and graphical analysis of velocity and acceleration of the particle, we can learn much about its motion. The numerical analysis complements the graphical analysis in giving a total view of the motion. The zero of the acceleration function corresponds to the maximum of the velocity in this example. Also in this example, when acceleration is positive and in the same direction as velocity, velocity increases. As acceleration tends toward zero, eventually becoming negative, the velocity reaches a maximum, after which it starts decreasing. If we wait long enough, velocity also becomes negative, indicating a reversal of direction. A real-world example of this type of motion is a car with a velocity that is increasing to a maximum, after which it starts slowing down, comes to a stop, then reverses direction.

Note:

Exercise:

Problem:

Check Your Understanding An airplane lands on a runway traveling east. Describe its acceleration.

Solution:

If we take east to be positive, then the airplane has negative acceleration because it is accelerating toward the west. It is also acceleration opposite to the motion; its acceleration is opposite in direction to its velocity.

Getting a Feel for Acceleration

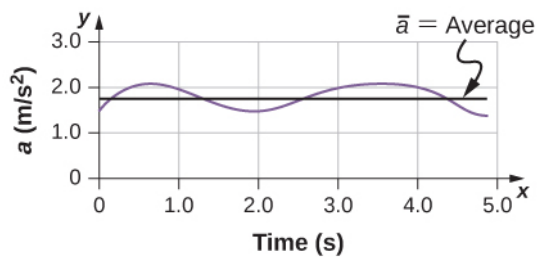
You are probably used to experiencing acceleration when you step into an elevator, or step on the gas pedal in your car. However, acceleration is happening to many other objects in our universe with which we don't have direct contact. [\[link\]](#) presents the acceleration of various objects. We can see the magnitudes of the accelerations extend over many orders of magnitude.

Acceleration	Value (m/s²)
High-speed train	0.25
Elevator	2
Cheetah	5
Object in a free fall without air resistance near the surface of Earth	9.8
Space shuttle maximum during launch	29
Parachutist peak during normal opening of parachute	59
F16 aircraft pulling out of a dive	79
Explosive seat ejection from aircraft	147
Sprint missile	982
Fastest rocket sled peak acceleration	1540
Jumping flea	3200

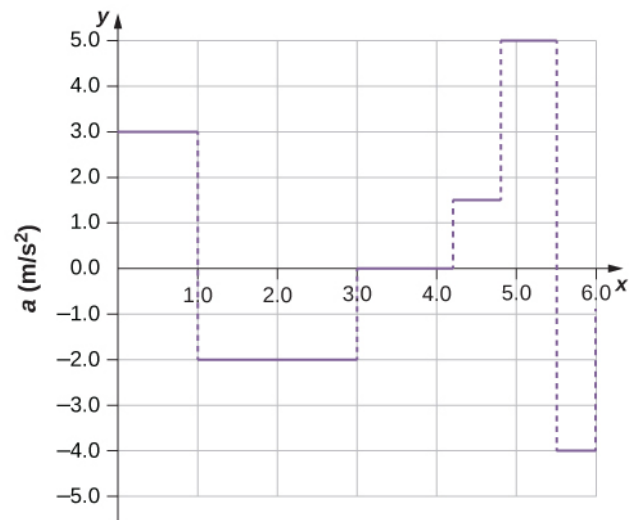
Acceleration	Value (m/s ²)
Baseball struck by a bat	30,000
Closing jaws of a trap-jaw ant	1,000,000
Proton in the large Hadron collider	1.9×10^9

Typical Values of Acceleration(credit: Wikipedia: Orders of Magnitude (acceleration))

In this table, we see that typical accelerations vary widely with different objects and have nothing to do with object size or how massive it is. Acceleration can also vary widely with time during the motion of an object. A drag racer has a large acceleration just after its start, but then it tapers off as the vehicle reaches a constant velocity. Its average acceleration can be quite different from its instantaneous acceleration at a particular time during its motion. [\[link\]](#) compares graphically average acceleration with instantaneous acceleration for two very different motions.



(a)



(b)

Graphs of instantaneous acceleration versus time for two different one-dimensional motions. (a) Acceleration varies only slightly and is always in the same direction, since it is positive. The average over the interval is nearly the same as the acceleration at any given time. (b) Acceleration varies greatly, perhaps representing a package on a post office conveyor belt that is accelerated forward and backward as it bumps along. It is necessary to consider small time intervals (such as from 0–1.0 s) with constant or nearly constant acceleration in such a situation.

Note:

Learn about position, velocity, and acceleration graphs. Move the little man back and forth with a mouse and plot his motion. Set the position, velocity, or acceleration and let the simulation move the man for you. Visit [this link](#) to use the moving man simulation.

Summary

- Acceleration is the rate at which velocity changes. Acceleration is a vector; it has both a magnitude and direction. The SI unit for acceleration is meters per second squared.
- Acceleration can be caused by a change in the magnitude or the direction of the velocity, or both.
- Instantaneous acceleration $a(t)$ is a continuous function of time and gives the acceleration at any specific time during the motion. It is calculated from the derivative of the velocity function. Instantaneous acceleration is the slope of the velocity-versus-time graph.
- Negative acceleration (sometimes called deceleration) is acceleration in the negative direction in the chosen coordinate system.

Conceptual Questions

Exercise:

Problem:

Is it possible for speed to be constant while acceleration is not zero?

Solution:

No, in one dimension constant speed requires zero acceleration.

Exercise:

Problem:

Is it possible for velocity to be constant while acceleration is not zero?
Explain.

Exercise:

Problem:

Give an example in which velocity is zero yet acceleration is not.

Solution:

A ball is thrown into the air and its velocity is zero at the apex of the throw, but acceleration is not zero.

Exercise:

Problem:

If a subway train is moving to the left (has a negative velocity) and then comes to a stop, what is the direction of its acceleration? Is the acceleration positive or negative?

Exercise:

Problem:

Plus and minus signs are used in one-dimensional motion to indicate direction. What is the sign of an acceleration that reduces the magnitude of a negative velocity? Of a positive velocity?

Solution:

Plus, minus

Exercise:**Problem:**

A cheetah can accelerate from rest to a speed of 30.0 m/s in 7.00 s.
What is its acceleration?

Solution:

$$a = 4.29 \text{ m/s}^2$$

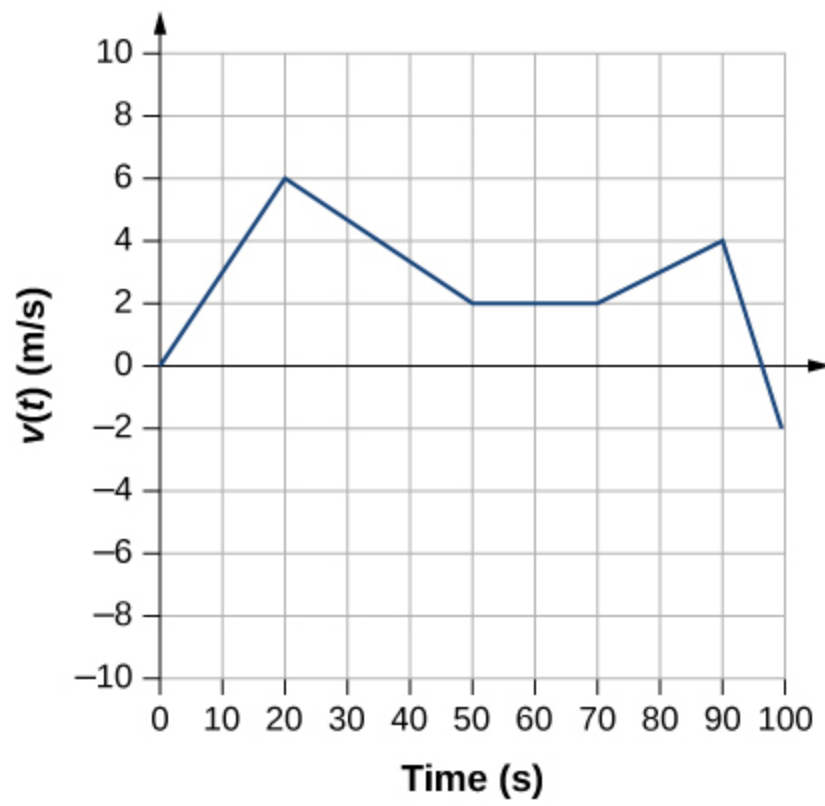
Exercise:**Problem:**

Dr. John Paul Stapp was a U.S. Air Force officer who studied the effects of extreme acceleration on the human body. On December 10, 1954, Stapp rode a rocket sled, accelerating from rest to a top speed of 282 m/s (1015 km/h) in 5.00 s and was brought jarringly back to rest in only 1.40 s. Calculate his (a) acceleration in his direction of motion and (b) acceleration opposite to his direction of motion. Express each in multiples of g (9.80 m/s^2) by taking its ratio to the acceleration of gravity.

Exercise:**Problem:**

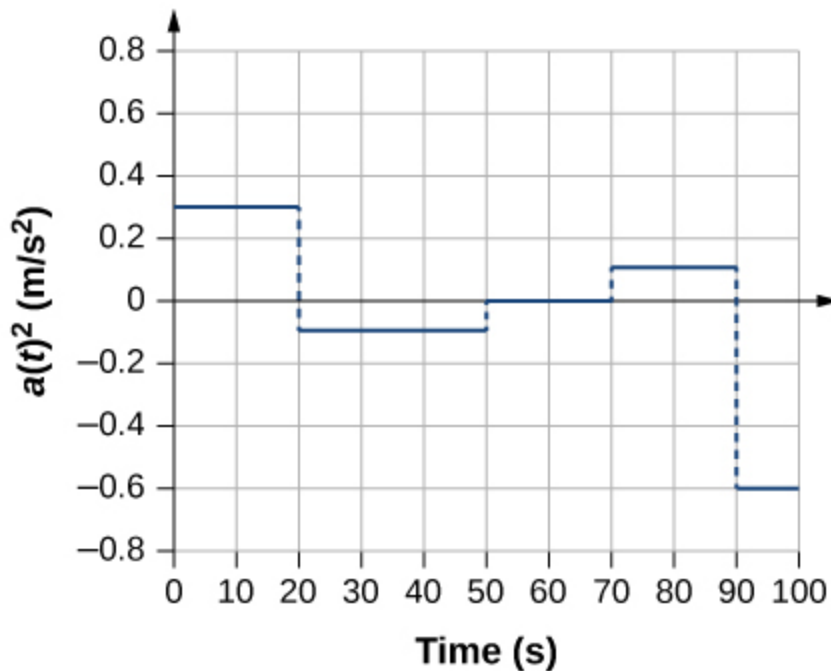
Sketch the acceleration-versus-time graph from the following velocity-versus-time graph.

Velocity vs. Time



Solution:

Acceleration vs. Time



Exercise:

Problem:

A commuter backs her car out of her garage with an acceleration of 1.40 m/s^2 . (a) How long does it take her to reach a speed of 2.00 m/s ? (b) If she then brakes to a stop in 0.800 s , what is her acceleration?

Exercise:

Problem:

Assume an intercontinental ballistic missile goes from rest to a suborbital speed of 6.50 km/s in 60.0 s (the actual speed and time are classified). What is its average acceleration in meters per second and in multiples of g (9.80 m/s^2)?

Solution:

$$a = 11.1g$$

Exercise:

Problem:

An airplane, starting from rest, moves down the runway at constant acceleration for 18 s and then takes off at a speed of 60 m/s. What is the average acceleration of the plane?

Glossary

average acceleration

the rate of change in velocity; the change in velocity over time

instantaneous acceleration

acceleration at a specific point in time

Motion with Constant Acceleration

By the end of this section, you will be able to:

- Identify which equations of motion are to be used to solve for unknowns.
- Use appropriate equations of motion to solve a two-body pursuit problem.

You might guess that the greater the acceleration of, say, a car moving away from a stop sign, the greater the car's displacement in a given time. But, we have not developed a specific equation that relates acceleration and displacement. In this section, we look at some convenient equations for kinematic relationships, starting from the definitions of displacement, velocity, and acceleration. We first investigate a single object in motion, called single-body motion. Then we investigate the motion of two objects, called **two-body pursuit problems**.

Notation

First, let us make some simplifications in notation. Taking the initial time to be zero, as if time is measured with a stopwatch, is a great simplification. Since elapsed time is $\Delta t = t_f - t_0$, taking $t_0 = 0$ means that $\Delta t = t_f$, the final time on the stopwatch. When initial time is taken to be zero, we use the subscript 0 to denote initial values of position and velocity. That is, x_0 is *the initial position* and v_0 is *the initial velocity*. We put no subscripts on the final values. That is, t is *the final time*, x is *the final position*, and v is *the final velocity*. This gives a simpler expression for elapsed time, $\Delta t = t$. It also simplifies the expression for x displacement, which is now $\Delta x = x - x_0$. Also, it simplifies the expression for change in velocity, which is now $\Delta v = v - v_0$. To summarize, using the simplified notation, with the initial time taken to be zero,

Equation:

$$\Delta t = t$$

$$\Delta x = x - x_0$$

$$\Delta v = v - v_0,$$

where the subscript 0 denotes an initial value and the absence of a subscript denotes a final value in whatever motion is under consideration.

We now make the important assumption that *acceleration is constant*. This assumption allows us to avoid using calculus to find instantaneous acceleration. Since acceleration is constant, the average and instantaneous accelerations are equal—that is,

Equation:

$$\bar{a} = a = \text{constant}.$$

Thus, we can use the symbol a for acceleration at all times. Assuming acceleration to be constant does not seriously limit the situations we can study nor does it degrade the accuracy of our treatment. For one thing, acceleration *is* constant in a great number of situations. Furthermore, in many other situations we can describe motion accurately by assuming a constant acceleration equal to the average acceleration for that motion. Lastly, for motion during which acceleration changes drastically, such as a car accelerating to top speed and then braking to a stop, motion can be considered in separate parts, each of which has its own constant acceleration.

Displacement and Position from Velocity

To get our first two equations, we start with the definition of average velocity:

Equation:

$$\bar{v} = \frac{\Delta x}{\Delta t}.$$

Substituting the simplified notation for Δx and Δt yields

Equation:

$$\bar{v} = \frac{x - x_0}{t}.$$

Solving for x gives us

Note:

Equation:

$$x = x_0 + \bar{v}t,$$

where the average velocity is

Note:

Equation:

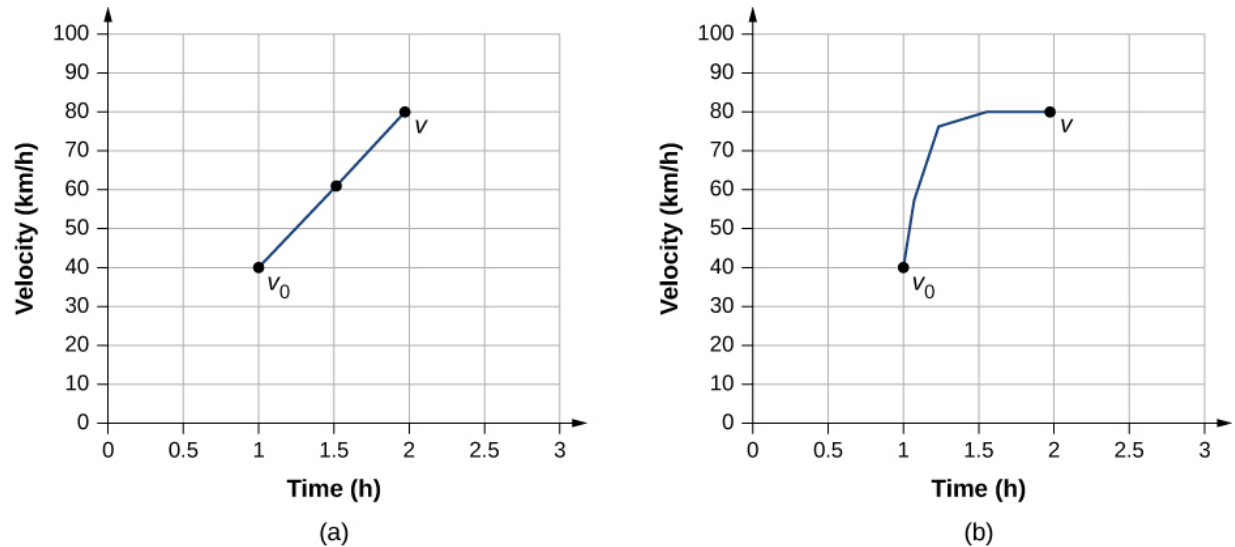
$$\bar{v} = \frac{v_0 + v}{2}.$$

The equation $\bar{v} = \frac{v_0 + v}{2}$ reflects the fact that when acceleration is constant, \bar{v} is just the simple average of the initial and final velocities. [\[link\]](#) illustrates this concept graphically. In part (a) of the figure, acceleration is constant, with velocity increasing at a constant rate. The average velocity during the 1-h interval from 40 km/h to 80 km/h is 60 km/h:

Equation:

$$\bar{v} = \frac{v_0 + v}{2} = \frac{40 \text{ km/h} + 80 \text{ km/h}}{2} = 60 \text{ km/h}.$$

In part (b), acceleration is not constant. During the 1-h interval, velocity is closer to 80 km/h than 40 km/h. Thus, the average velocity is greater than in part (a).



- (a) Velocity-versus-time graph with constant acceleration showing the initial and final velocities v_0 and v . The average velocity is $\frac{1}{2}(v_0 + v) = 60$ km/h. (b) Velocity-versus-time graph with an acceleration that changes with time. The average velocity is not given by $\frac{1}{2}(v_0 + v)$, but is greater than 60 km/h.

Solving for Final Velocity from Acceleration and Time

We can derive another useful equation by manipulating the definition of acceleration:

Equation:

$$a = \frac{\Delta v}{\Delta t}.$$

Substituting the simplified notation for Δv and Δt gives us

Equation:

$$a = \frac{v - v_0}{t} \quad (\text{constant } a).$$

Solving for v yields

Note:

Equation:

$$v = v_0 + at \quad (\text{constant } a).$$

Example:

Calculating Final Velocity

An airplane lands with an initial velocity of 70.0 m/s and then accelerates opposite to the motion at 1.50 m/s² for 40.0 s. What is its final velocity?

Strategy

First, we identify the knowns: $v_0 = 70 \text{ m/s}$, $a = -1.50 \text{ m/s}^2$, $t = 40 \text{ s}$.
Second, we identify the unknown; in this case, it is final velocity v_f .
Last, we determine which equation to use. To do this we figure out which kinematic equation gives the unknown in terms of the knowns. We calculate the final velocity using [\[link\]](#), $v = v_0 + at$.

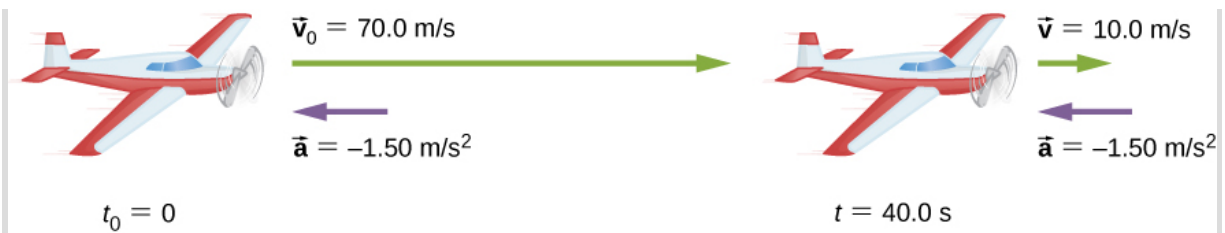
Solution

Substitute the known values and solve:

Equation:

$$v = v_0 + at = 70.0 \text{ m/s} + (-1.50 \text{ m/s}^2)(40.0 \text{ s}) = 10.0 \text{ m/s}.$$

[\[link\]](#) is a sketch that shows the acceleration and velocity vectors.



The airplane lands with an initial velocity of 70.0 m/s and slows to a final velocity of 10.0 m/s before heading for the terminal. Note the acceleration is negative because its direction is opposite to its velocity, which is positive.

Significance

The final velocity is much less than the initial velocity, as desired when slowing down, but is still positive (see figure). With jet engines, reverse thrust can be maintained long enough to stop the plane and start moving it backward, which is indicated by a negative final velocity, but is not the case here.

In addition to being useful in problem solving, the equation $v = v_0 + at$ gives us insight into the relationships among velocity, acceleration, and time. We can see, for example, that

- Final velocity depends on how large the acceleration is and how long it lasts
- If the acceleration is zero, then the final velocity equals the initial velocity ($v = v_0$), as expected (in other words, velocity is constant)
- If a is negative, then the final velocity is less than the initial velocity

All these observations fit our intuition. Note that it is always useful to examine basic equations in light of our intuition and experience to check that they do indeed describe nature accurately.

Solving for Final Position with Constant Acceleration

We can combine the previous equations to find a third equation that allows us to calculate the final position of an object experiencing constant acceleration. We start with

Equation:

$$v = v_0 + at.$$

Adding v_0 to each side of this equation and dividing by 2 gives

Equation:

$$\frac{v_0 + v}{2} = v_0 + \frac{1}{2}at.$$

Since $\frac{v_0 + v}{2} = \bar{v}$ for constant acceleration, we have

Equation:

$$\bar{v} = v_0 + \frac{1}{2}at.$$

Now we substitute this expression for \bar{v} into the equation for displacement, $x = x_0 + \bar{v}t$, yielding

Note:

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2 \quad (\text{constant } a).$$

Example:

Calculating Displacement of an Accelerating Object

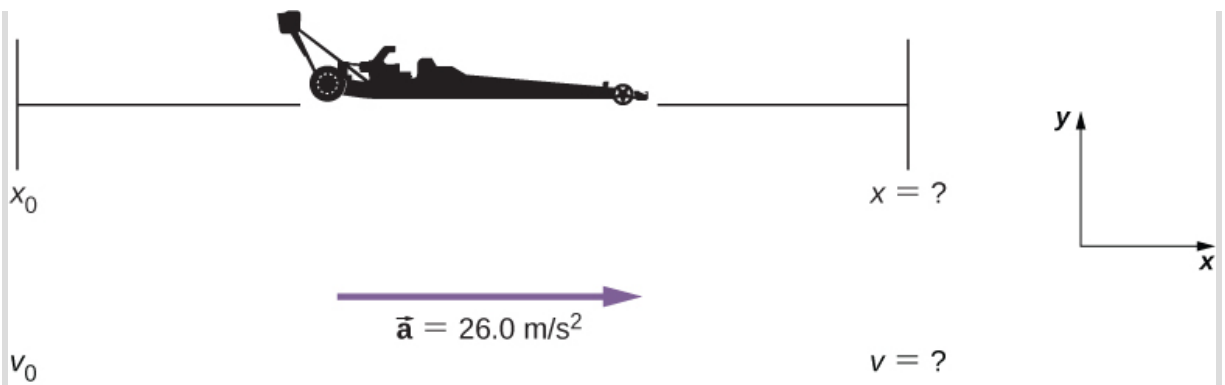
Dragsters can achieve an average acceleration of 26.0 m/s^2 . Suppose a dragster accelerates from rest at this rate for 5.56 s [\[link\]](#). How far does it travel in this time?



U.S. Army Top Fuel pilot Tony “The Sarge” Schumacher begins a race with a controlled burnout.
(credit: Lt. Col. William Thurmond. Photo Courtesy of U.S. Army.)

Strategy

First, let's draw a sketch [\[link\]](#). We are asked to find displacement, which is x if we take x_0 to be zero. (Think about x_0 as the starting line of a race. It can be anywhere, but we call it zero and measure all other positions relative to it.) We can use the equation $x = x_0 + v_0t + \frac{1}{2}at^2$ when we identify v_0 , a , and t from the statement of the problem.



Sketch of an accelerating dragster.

Solution

First, we need to identify the knowns. Starting from rest means that $v_0 = 0$, a is given as 26.0 m/s^2 and t is given as 5.56 s .

Second, we substitute the known values into the equation to solve for the unknown:

Equation:

$$x = x_0 + v_0 t + \frac{1}{2} a t^2.$$

Since the initial position and velocity are both zero, this equation simplifies to

Equation:

$$x = \frac{1}{2} a t^2.$$

Substituting the identified values of a and t gives

Equation:

$$x = \frac{1}{2} (26.0 \text{ m/s}^2) (5.56 \text{ s})^2 = 402 \text{ m}.$$

Significance

If we convert 402 m to miles, we find that the distance covered is very close to one-quarter of a mile, the standard distance for drag racing. So, our

answer is reasonable. This is an impressive displacement to cover in only 5.56 s, but top-notch dragsters can do a quarter mile in even less time than this. If the dragster were given an initial velocity, this would add another term to the distance equation. If the same acceleration and time are used in the equation, the distance covered would be much greater.

What else can we learn by examining the equation $x = x_0 + v_0t + \frac{1}{2}at^2$? We can see the following relationships:

- Displacement depends on the square of the elapsed time when acceleration is not zero. In [\[link\]](#), the dragster covers only one-fourth of the total distance in the first half of the elapsed time.
- If acceleration is zero, then initial velocity equals average velocity ($v_0 = \bar{v}$), and $x = x_0 + v_0t + \frac{1}{2}at^2$ becomes $x = x_0 + v_0t$.

Solving for Final Velocity from Distance and Acceleration

A fourth useful equation can be obtained from another algebraic manipulation of previous equations. If we solve $v = v_0 + at$ for t , we get
Equation:

$$t = \frac{v - v_0}{a}.$$

Substituting this and $\bar{v} = \frac{v_0 + v}{2}$ into $x = x_0 + \bar{v}t$, we get

Note:

Equation:

$$v^2 = v_0^2 + 2a(x - x_0) \quad (\text{constant } a).$$

Example:**Calculating Final Velocity**

Calculate the final velocity of the dragster in [\[link\]](#) without using information about time.

Strategy

The equation $v^2 = v_0^2 + 2a(x - x_0)$ is ideally suited to this task because it relates velocities, acceleration, and displacement, and no time information is required.

Solution

First, we identify the known values. We know that $v_0 = 0$, since the dragster starts from rest. We also know that $x - x_0 = 402 \text{ m}$ (this was the answer in [\[link\]](#)). The average acceleration was given by $a = 26.0 \text{ m/s}^2$. Second, we substitute the knowns into the equation $v^2 = v_0^2 + 2a(x - x_0)$ and solve for v :

Equation:

$$v^2 = 0 + 2 \left(26.0 \text{ m/s}^2 \right) (402 \text{ m}).$$

Thus,

Equation:

$$\begin{aligned} v^2 &= 2.09 \times 10^4 \text{ m}^2/\text{s}^2 \\ v &= \sqrt{2.09 \times 10^4 \text{ m}^2/\text{s}^2} = 145 \text{ m/s}. \end{aligned}$$

Significance

A velocity of 145 m/s is about 522 km/h, or about 324 mi/h, but even this breakneck speed is short of the record for the quarter mile. Also, note that a square root has two values; we took the positive value to indicate a velocity in the same direction as the acceleration.

An examination of the equation $v^2 = v_0^2 + 2a(x - x_0)$ can produce additional insights into the general relationships among physical quantities:

- The final velocity depends on how large the acceleration is and the distance over which it acts.
- For a fixed acceleration, a car that is going twice as fast doesn't simply stop in twice the distance. It takes much farther to stop. (This is why we have reduced speed zones near schools.)

Putting Equations Together

In the following examples, we continue to explore one-dimensional motion, but in situations requiring slightly more algebraic manipulation. The examples also give insight into problem-solving techniques. The note that follows is provided for easy reference to the equations needed. Be aware that these equations are not independent. In many situations we have two unknowns and need two equations from the set to solve for the unknowns. We need as many equations as there are unknowns to solve a given situation.

Note:

Summary of Kinematic Equations (constant a)

Equation:

$$x = x_0 + \bar{v}t$$

Equation:

$$\bar{v} = \frac{v_0 + v}{2}$$

Equation:

$$v = v_0 + at$$

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2$$

Equation:

$$v^2 = v_0^2 + 2a(x - x_0)$$

Before we get into the examples, let's look at some of the equations more closely to see the behavior of acceleration at extreme values. Rearranging [\[link\]](#), we have

Equation:

$$a = \frac{v - v_0}{t}.$$

From this we see that, for a finite time, if the difference between the initial and final velocities is small, the acceleration is small, approaching zero in the limit that the initial and final velocities are equal. On the contrary, in the limit $t \rightarrow 0$ for a finite difference between the initial and final velocities, acceleration becomes infinite.

Similarly, rearranging [\[link\]](#), we can express acceleration in terms of velocities and displacement:

Equation:

$$a = \frac{v^2 - v_0^2}{2(x - x_0)}.$$

Thus, for a finite difference between the initial and final velocities acceleration becomes infinite in the limit the displacement approaches zero. Acceleration approaches zero in the limit the difference in initial and final velocities approaches zero for a finite displacement.

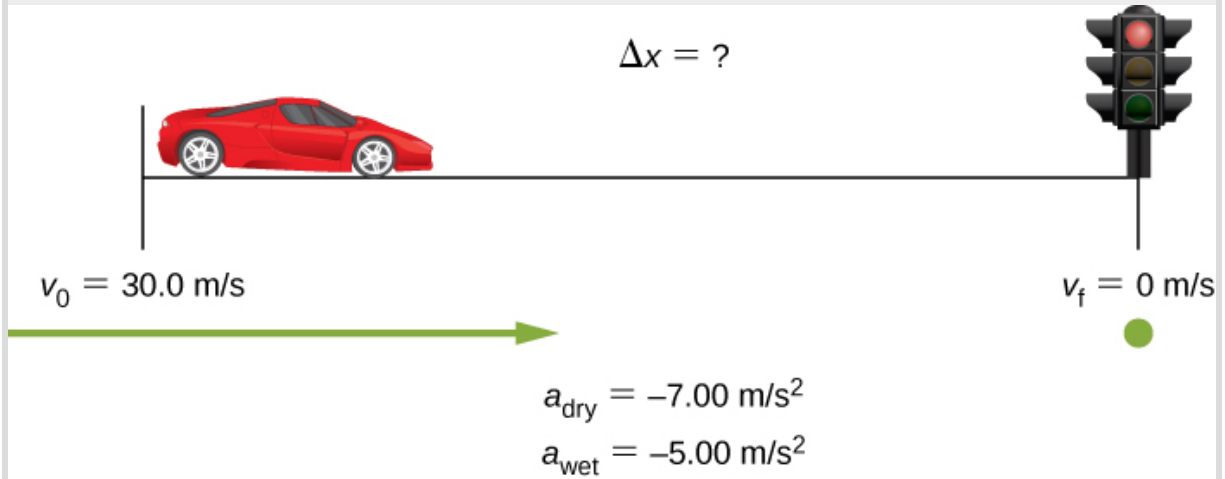
Example:

How Far Does a Car Go?

On dry concrete, a car can accelerate opposite to the motion at a rate of 7.00 m/s^2 , whereas on wet concrete it can accelerate opposite to the motion at only 5.00 m/s^2 . Find the distances necessary to stop a car moving at 30.0 m/s (about 110 km/h) on (a) dry concrete and (b) wet concrete. (c) Repeat both calculations and find the displacement from the point where the driver sees a traffic light turn red, taking into account his reaction time of 0.500 s to get his foot on the brake.

Strategy

First, we need to draw a sketch [\[link\]](#). To determine which equations are best to use, we need to list all the known values and identify exactly what we need to solve for.



Sample sketch to visualize acceleration opposite to the motion and stopping distance of a car.

Solution

- a. First, we need to identify the knowns and what we want to solve for. We know that $v_0 = 30.0 \text{ m/s}$, $v = 0$, and $a = -7.00 \text{ m/s}^2$ (a is negative because it is in a direction opposite to velocity). We take x_0 to be zero. We are looking for displacement Δx , or $x - x_0$. Second, we identify the equation that will help us solve the problem. The best equation to use is

Equation:

$$v^2 = v_0^2 + 2a(x - x_0).$$

This equation is best because it includes only one unknown, x . We know the values of all the other variables in this equation. (Other equations would allow us to solve for x , but they require us to know the stopping time, t , which we do not know. We could use them, but it would entail additional calculations.)

Third, we rearrange the equation to solve for x :

Equation:

$$x - x_0 = \frac{v^2 - v_0^2}{2a}$$

and substitute the known values:

Equation:

$$x - 0 = \frac{0^2 - (30.0 \text{ m/s})^2}{2(-7.00 \text{ m/s}^2)}.$$

Thus,

Equation:

$$x = 64.3 \text{ m on dry concrete.}$$

- b. This part can be solved in exactly the same manner as (a). The only difference is that the acceleration is -5.00 m/s^2 . The result is

Equation:

$$x_{\text{wet}} = 90.0 \text{ m on wet concrete.}$$

- c. When the driver reacts, the stopping distance is the same as it is in (a) and (b) for dry and wet concrete. So, to answer this question, we need to calculate how far the car travels during the reaction time, and then add that to the stopping time. It is reasonable to assume the velocity remains constant during the driver's reaction time.

To do this, we, again, identify the knowns and what we want to solve for. We know that $\bar{v} = 30.0 \text{ m/s}$, $t_{\text{reaction}} = 0.500 \text{ s}$, and $a_{\text{reaction}} = 0$. We take $x_{0\text{-reaction}}$ to be zero. We are looking for

x_{reaction} .

Second, as before, we identify the best equation to use. In this case, $x = x_0 + \bar{v}t$ works well because the only unknown value is x , which is what we want to solve for.

Third, we substitute the knowns to solve the equation:

Equation:

$$x = 0 + (30.0 \text{ m/s}) (0.500 \text{ s}) = 15.0 \text{ m}.$$

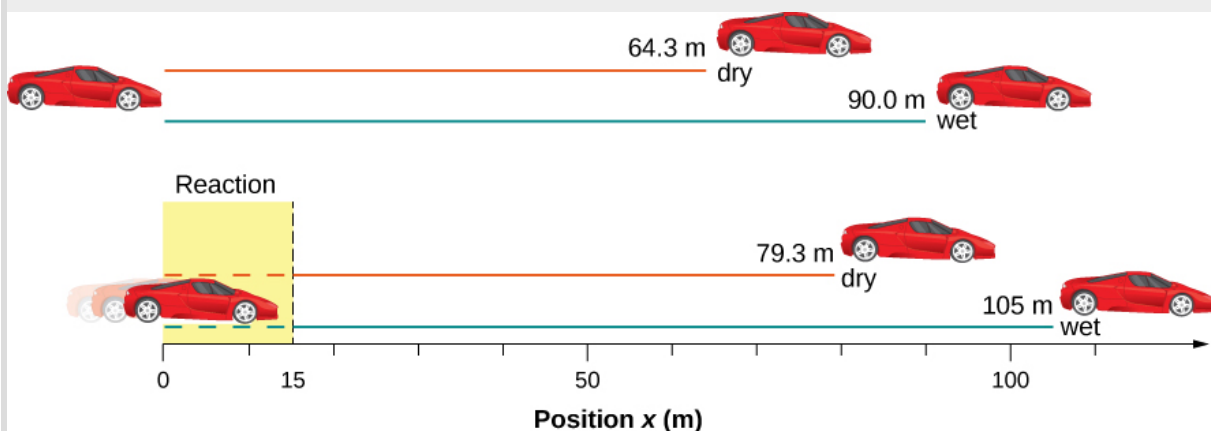
This means the car travels 15.0 m while the driver reacts, making the total displacements in the two cases of dry and wet concrete 15.0 m greater than if he reacted instantly.

Last, we then add the displacement during the reaction time to the displacement when braking ([\[link\]](#)),

Equation:

$$x_{\text{braking}} + x_{\text{reaction}} = x_{\text{total}},$$

and find (a) to be $64.3 \text{ m} + 15.0 \text{ m} = 79.3 \text{ m}$ when dry and (b) to be $90.0 \text{ m} + 15.0 \text{ m} = 105 \text{ m}$ when wet.



The distance necessary to stop a car varies greatly, depending on road conditions and driver reaction time. Shown here are the braking distances for dry and wet pavement, as calculated in this example, for a car traveling initially at 30.0 m/s. Also shown are the total distances traveled from the point when the driver first sees a light turn red, assuming a 0.500-s reaction time.

Significance

The displacements found in this example seem reasonable for stopping a fast-moving car. It should take longer to stop a car on wet pavement than dry. It is interesting that reaction time adds significantly to the displacements, but more important is the general approach to solving problems. We identify the knowns and the quantities to be determined, then find an appropriate equation. If there is more than one unknown, we need as many independent equations as there are unknowns to solve. There is often more than one way to solve a problem. The various parts of this example can, in fact, be solved by other methods, but the solutions presented here are the shortest.

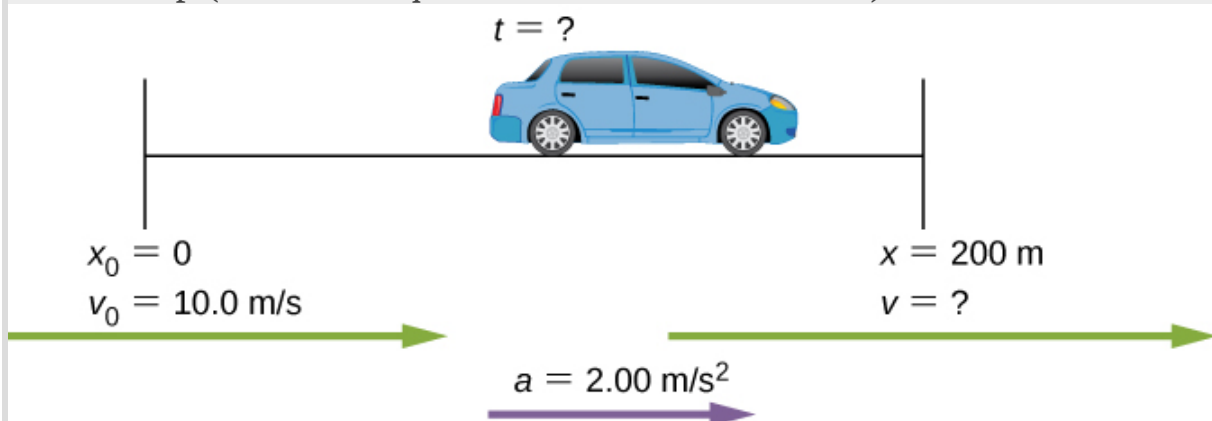
Example:

Calculating Time

Suppose a car merges into freeway traffic on a 200-m-long ramp. If its initial velocity is 10.0 m/s and it accelerates at 2.00 m/s^2 , how long does it take the car to travel the 200 m up the ramp? (Such information might be useful to a traffic engineer.)

Strategy

First, we draw a sketch [\[link\]](#). We are asked to solve for time t . As before, we identify the known quantities to choose a convenient physical relationship (that is, an equation with one unknown, t .)



Sketch of a car accelerating on a freeway ramp.

Solution

Again, we identify the knowns and what we want to solve for. We know that $x_0 = 0$,

$v_0 = 10 \text{ m/s}$, $a = 2.00 \text{ m/s}^2$, and $x = 200 \text{ m}$.

We need to solve for t . The equation $x = x_0 + v_0t + \frac{1}{2}at^2$ works best because the only unknown in the equation is the variable t , for which we need to solve. From this insight we see that when we input the knowns into the equation, we end up with a quadratic equation.

We need to rearrange the equation to solve for t , then substituting the knowns into the equation:

Equation:

$$200 \text{ m} = 0 \text{ m} + (10.0 \text{ m/s})t + \frac{1}{2} (2.00 \text{ m/s}^2)t^2.$$

We then simplify the equation. The units of meters cancel because they are in each term. We can get the units of seconds to cancel by taking $t = t \text{ s}$, where t is the magnitude of time and s is the unit. Doing so leaves

Equation:

$$200 = 10t + t^2.$$

We then use the quadratic formula to solve for t ,

Equation:

$$t^2 + 10t - 200 = 0$$

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

which yields two solutions: $t = 10.0$ and $t = -20.0$. A negative value for time is unreasonable, since it would mean the event happened 20 s before the motion began. We can discard that solution. Thus,

Equation:

$$t = 10.0 \text{ s}.$$

Significance

Whenever an equation contains an unknown squared, there are two solutions. In some problems both solutions are meaningful; in others, only one solution is reasonable. The 10.0-s answer seems reasonable for a typical freeway on-ramp.

Note:

Exercise:

Problem:

Check Your Understanding A rocket accelerates at a rate of 20 m/s^2 during launch. How long does it take the rocket to reach a velocity of 400 m/s ?

Solution:

To answer this, choose an equation that allows us to solve for time t , given only a , v_0 , and v :

$$v = v_0 + at.$$

Rearrange to solve for t :

$$t = \frac{v - v_0}{a} = \frac{400 \text{ m/s} - 0 \text{ m/s}}{20 \text{ m/s}^2} = 20 \text{ s}.$$

Example:

Acceleration of a Spaceship

A spaceship has left Earth's orbit and is on its way to the Moon. It accelerates at 20 m/s^2 for 2 min and covers a distance of 1000 km. What are the initial and final velocities of the spaceship?

Strategy

We are asked to find the initial and final velocities of the spaceship. Looking at the kinematic equations, we see that one equation will not give the answer. We must use one kinematic equation to solve for one of the velocities and substitute it into another kinematic equation to get the

second velocity. Thus, we solve two of the kinematic equations simultaneously.

Solution

First we solve for v_0 using $x = x_0 + v_0t + \frac{1}{2}at^2$:

Equation:

$$x - x_0 = v_0t + \frac{1}{2}at^2$$

Equation:

$$1.0 \times 10^6 \text{ m} = v_0(120.0 \text{ s}) + \frac{1}{2}(20.0 \text{ m/s}^2)(120.0 \text{ s})^2$$

Equation:

$$v_0 = 7133.3 \text{ m/s.}$$

Then we substitute v_0 into $v = v_0 + at$ to solve for the final velocity:

Equation:

$$v = v_0 + at = 7133.3 \text{ m/s} + (20.0 \text{ m/s}^2)(120.0 \text{ s}) = 9533.3 \text{ m/s.}$$

Significance

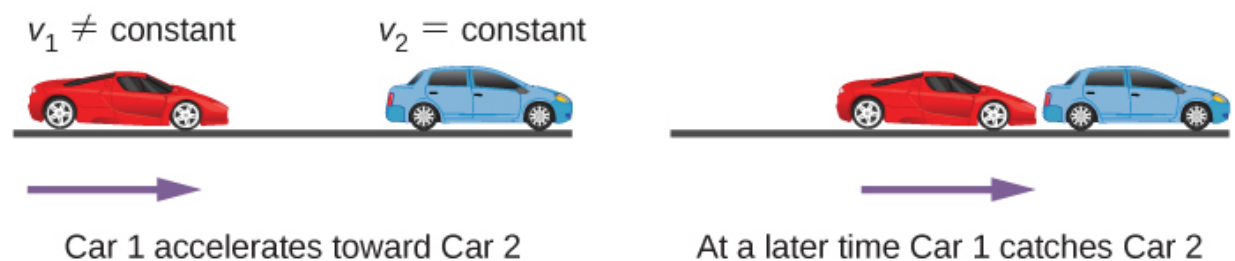
There are six variables in displacement, time, velocity, and acceleration that describe motion in one dimension. The initial conditions of a given problem can be many combinations of these variables. Because of this diversity, solutions may not be as easy as simple substitutions into one of the equations. This example illustrates that solutions to kinematics may require solving two simultaneous kinematic equations.

With the basics of kinematics established, we can go on to many other interesting examples and applications. In the process of developing kinematics, we have also glimpsed a general approach to problem solving that produces both correct answers and insights into physical relationships.

The next level of complexity in our kinematics problems involves the motion of two interrelated bodies, called *two-body pursuit problems*.

Two-Body Pursuit Problems

Up until this point we have looked at examples of motion involving a single body. Even for the problem with two cars and the stopping distances on wet and dry roads, we divided this problem into two separate problems to find the answers. In a **two-body pursuit problem**, the motions of the objects are coupled—meaning, the unknown we seek depends on the motion of both objects. To solve these problems we write the equations of motion for each object and then solve them simultaneously to find the unknown. This is illustrated in [\[link\]](#).



A two-body pursuit scenario where car 2 has a constant velocity and car 1 is behind with a constant acceleration. Car 1 catches up with car 2 at a later time.

The time and distance required for car 1 to catch car 2 depends on the initial distance car 1 is from car 2 as well as the velocities of both cars and the acceleration of car 1. The kinematic equations describing the motion of both cars must be solved to find these unknowns.

Consider the following example.

Example:**Cheetah Catching a Gazelle**

A cheetah waits in hiding behind a bush. The cheetah spots a gazelle running past at 10 m/s. At the instant the gazelle passes the cheetah, the cheetah accelerates from rest at 4 m/s^2 to catch the gazelle. (a) How long does it take the cheetah to catch the gazelle? (b) What is the displacement of the gazelle and cheetah?

Strategy

We use the set of equations for constant acceleration to solve this problem. Since there are two objects in motion, we have separate equations of motion describing each animal. But what links the equations is a common parameter that has the same value for each animal. If we look at the problem closely, it is clear the common parameter to each animal is their position x at a later time t . Since they both start at $x_0 = 0$, their displacements are the same at a later time t , when the cheetah catches up with the gazelle. If we pick the equation of motion that solves for the displacement for each animal, we can then set the equations equal to each other and solve for the unknown, which is time.

Solution

- a. Equation for the gazelle: The gazelle has a constant velocity, which is its average velocity, since it is not accelerating. Therefore, we use [\[link\]](#) with $x_0 = 0$:

Equation:

$$x = x_0 + \bar{v}t = \bar{v}t.$$

Equation for the cheetah: The cheetah is accelerating from rest, so we use [\[link\]](#) with $x_0 = 0$ and $v_0 = 0$:

Equation:

$$x = x_0 + v_0t + \frac{1}{2}at^2 = \frac{1}{2}at^2.$$

Now we have an equation of motion for each animal with a common parameter, which can be eliminated to find the solution. In this case, we solve for t :

Equation:

$$x = \bar{v}t = \frac{1}{2}at^2$$

$$t = \frac{2\bar{v}}{a}.$$

The gazelle has a constant velocity of 10 m/s, which is its average velocity. The acceleration of the cheetah is 4 m/s². Evaluating t , the time for the cheetah to reach the gazelle, we have

Equation:

$$t = \frac{2\bar{v}}{a} = \frac{2(10 \text{ m/s})}{4\text{m/s}^2} = 5 \text{ s}.$$

- b. To get the displacement, we use either the equation of motion for the cheetah or the gazelle, since they should both give the same answer. Displacement of the cheetah:

Equation:

$$x = \frac{1}{2}at^2 = \frac{1}{2}(4\text{m/s}^2)(5)^2 = 50 \text{ m}.$$

Displacement of the gazelle:

Equation:

$$x = \bar{v}t = 10 \text{ m/s}(5) = 50 \text{ m}.$$

We see that both displacements are equal, as expected.

Significance

It is important to analyze the motion of each object and to use the appropriate kinematic equations to describe the individual motion. It is also important to have a good visual perspective of the two-body pursuit problem to see the common parameter that links the motion of both objects.

Note:

Exercise:

Problem:

Check Your Understanding A bicycle has a constant velocity of 10 m/s. A person starts from rest and begins to run to catch up to the bicycle in 30 s when the bicycle is at the same position as the person. What is the acceleration of the person?

Solution:

$$a = \frac{2}{3} \text{ m/s}^2.$$

Summary

- When analyzing one-dimensional motion with constant acceleration, identify the known quantities and choose the appropriate equations to solve for the unknowns. Either one or two of the kinematic equations are needed to solve for the unknowns, depending on the known and unknown quantities.
- Two-body pursuit problems always require two equations to be solved simultaneously for the unknowns.

Conceptual Questions

Exercise:**Problem:**

When analyzing the motion of a single object, what is the required number of known physical variables that are needed to solve for the unknown quantities using the kinematic equations?

Exercise:

Problem:

State two scenarios of the kinematics of single object where three known quantities require two kinematic equations to solve for the unknowns.

Solution:

If the acceleration, time, and displacement are the knowns, and the initial and final velocities are the unknowns, then two kinematic equations must be solved simultaneously. Also if the final velocity, time, and displacement are the knowns then two kinematic equations must be solved for the initial velocity and acceleration.

Problems**Exercise:****Problem:**

A particle moves in a straight line at a constant velocity of 30 m/s. What is its displacement between $t = 0$ and $t = 5.0$ s?

Solution:

150 m

Exercise:**Problem:**

A particle moves in a straight line with an initial velocity of 0 m/s and a constant acceleration of 30 m/s^2 . If $x = 0$ at $t = 0$, what is the particle's position at $t = 5$ s?

Exercise:

Problem:

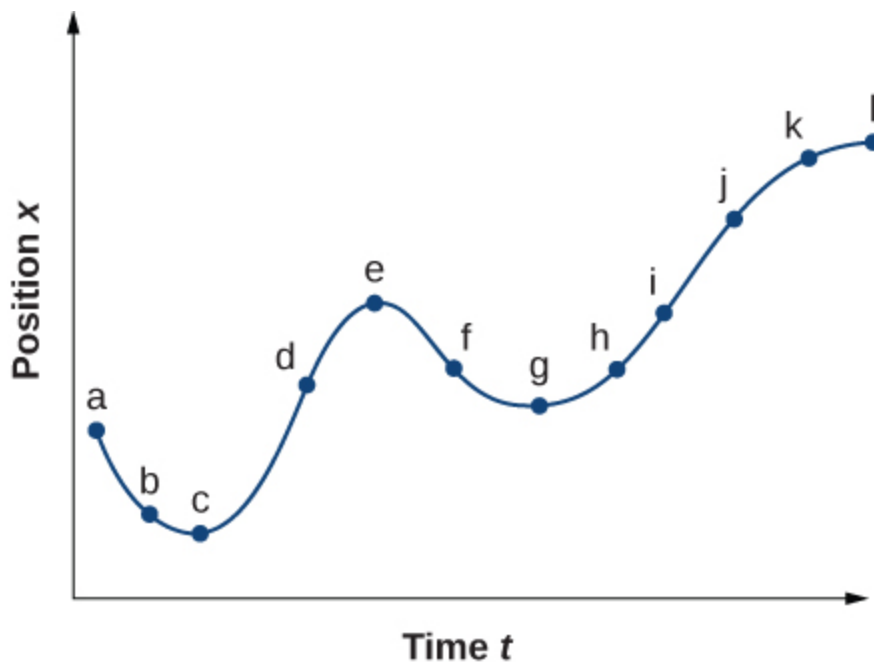
A particle moves in a straight line with an initial velocity of 30 m/s and constant acceleration 30 m/s². (a) What is its displacement at $t = 5$ s? (b) What is its velocity at this same time?

Solution:

- a. 525 m;
- b. $v = 180$ m/s

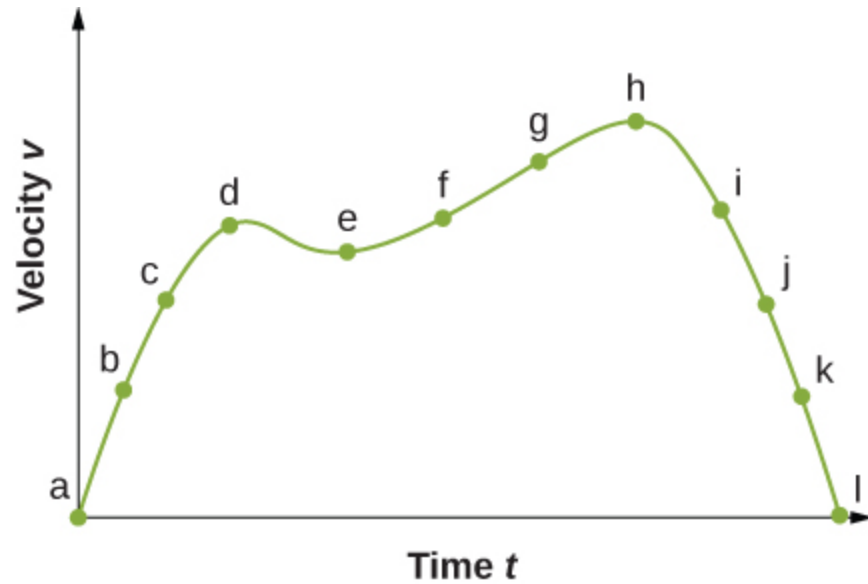
Exercise:**Problem:**

(a) Sketch a graph of velocity versus time corresponding to the graph of displacement versus time given in the following figure. (b) Identify the time or times (t_a , t_b , t_c , etc.) at which the instantaneous velocity has the greatest positive value. (c) At which times is it zero? (d) At which times is it negative?

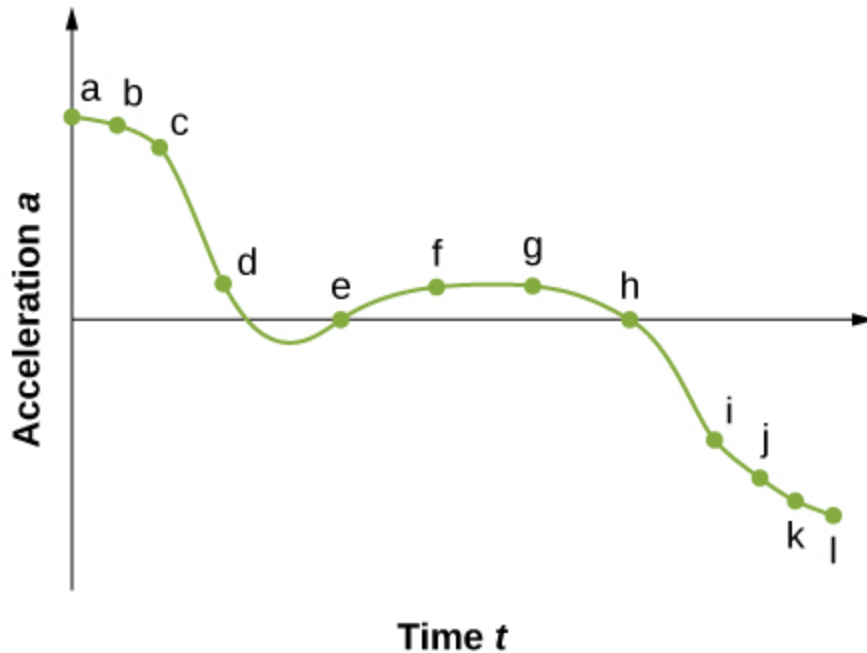
**Exercise:**

Problem:

(a) Sketch a graph of acceleration versus time corresponding to the graph of velocity versus time given in the following figure. (b) Identify the time or times (t_a , t_b , t_c , etc.) at which the acceleration has the greatest positive value. (c) At which times is it zero? (d) At which times is it negative?

**Solution:**

a.



- b. The acceleration has the greatest positive value at t_a
- c. The acceleration is zero at t_e and t_h
- d. The acceleration is negative at t_i, t_j, t_k, t_l

Exercise:

Problem:

A particle has a constant acceleration of 6.0 m/s^2 . (a) If its initial velocity is 2.0 m/s , at what time is its displacement 5.0 m ? (b) What is its velocity at that time?

Exercise:

Problem:

At $t = 10 \text{ s}$, a particle is moving from left to right with a speed of 5.0 m/s . At $t = 20 \text{ s}$, the particle is moving right to left with a speed of 8.0 m/s . Assuming the particle's acceleration is constant, determine (a) its acceleration, (b) its initial velocity, and (c) the instant when its velocity is zero.

Solution:

- a. $a = -1.3 \text{ m/s}^2$;
- b. $v_0 = 18 \text{ m/s}$;
- c. $t = 13.8 \text{ s}$

Exercise:

Problem:

A well-thrown ball is caught in a well-padded mitt. If the acceleration of the ball is $2.10 \times 10^4 \text{ m/s}^2$, and 1.85 ms ($1 \text{ ms} = 10^{-3} \text{ s}$) elapses from the time the ball first touches the mitt until it stops, what is the initial velocity of the ball?

Exercise:

Problem:

A bullet in a gun is accelerated from the firing chamber to the end of the barrel at an average rate of $6.20 \times 10^5 \text{ m/s}^2$ for $8.10 \times 10^{-4} \text{ s}$. What is its muzzle velocity (that is, its final velocity)?

Solution:

$$v = 502.20 \text{ m/s}$$

Exercise:

Problem:

- (a) A light-rail commuter train accelerates at a rate of 1.35 m/s^2 . How long does it take to reach its top speed of 80.0 km/h , starting from rest?
- (b) The same train ordinarily accelerates opposite to the motion at a rate of 1.65 m/s^2 . How long does it take to come to a stop from its top speed?
- (c) In emergencies, the train can accelerate opposite to the motion more rapidly, coming to rest from 80.0 km/h in 8.30 s . What is its emergency acceleration in meters per second squared?

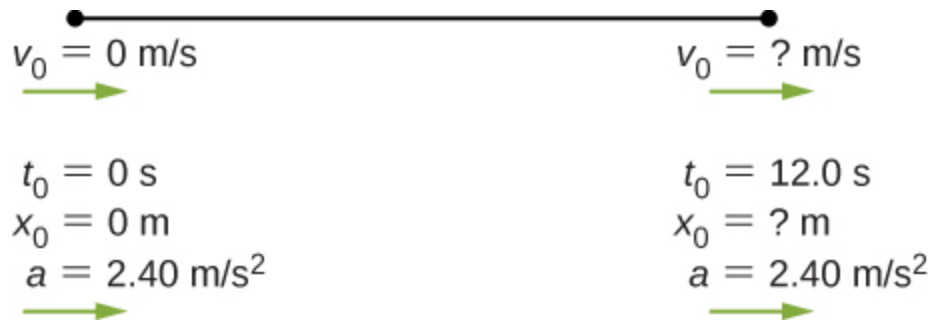
Exercise:

Problem:

While entering a freeway, a car accelerates from rest at a rate of 2.40 m/s^2 for 12.0 s . (a) Draw a sketch of the situation. (b) List the knowns in this problem. (c) How far does the car travel in those 12.0 s ? To solve this part, first identify the unknown, then indicate how you chose the appropriate equation to solve for it. After choosing the equation, show your steps in solving for the unknown, check your units, and discuss whether the answer is reasonable. (d) What is the car's final velocity? Solve for this unknown in the same manner as in (c), showing all steps explicitly.

Solution:

a.



b. Knowns: $a = 2.40 \text{ m/s}^2$, $t = 12.0 \text{ s}$, $v_0 = 0 \text{ m/s}$, and $x_0 = 0 \text{ m}$;

c. $x = x_0 + v_0 t + \frac{1}{2} a t^2 = \frac{1}{2} a t^2 = 2.40 \text{ m/s}^2 (12.0 \text{ s})^2 = 172.80 \text{ m}$
, the answer seems reasonable at about 172.8 m ; d. $v = 28.8 \text{ m/s}$

Exercise:**Problem:**

Unreasonable results At the end of a race, a runner accelerates opposite to the motion from a velocity of 9.00 m/s at a rate of 2.00 m/s^2 . (a) How far does she travel in the next 5.00 s ? (b) What is her final velocity? (c) Evaluate the result. Does it make sense?

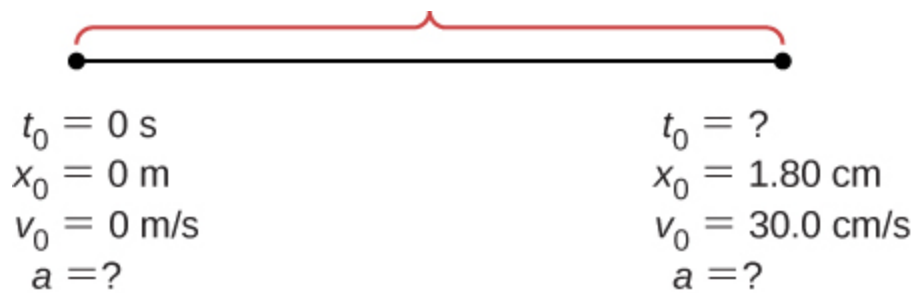
Exercise:

Problem:

Blood is accelerated from rest to 30.0 cm/s in a distance of 1.80 cm by the left ventricle of the heart. (a) Make a sketch of the situation. (b) List the knowns in this problem. (c) How long does the acceleration take? To solve this part, first identify the unknown, then discuss how you chose the appropriate equation to solve for it. After choosing the equation, show your steps in solving for the unknown, checking your units. (d) Is the answer reasonable when compared with the time for a heartbeat?

Solution:

a.



b. Knowns: $v = 30.0 \text{ cm/s}$, $x = 1.80 \text{ cm}$;

c. $a = 250 \text{ cm/s}^2$, $t = 0.12 \text{ s}$;

d. yes

Exercise:**Problem:**

During a slap shot, a hockey player accelerates the puck from a velocity of 8.00 m/s to 40.0 m/s in the same direction. If this shot takes $3.33 \times 10^{-2} \text{ s}$, what is the distance over which the puck accelerates?

Exercise:

Problem:

A powerful motorcycle can accelerate from rest to 26.8 m/s (100 km/h) in only 3.90 s. (a) What is its average acceleration? (b) Assuming constant acceleration, how far does it travel in that time?

Solution:

a. 6.87 m/s^2 ; b. $x = 52.26 \text{ m}$

Exercise:**Problem:**

Freight trains can produce only relatively small accelerations. (a) What is the final velocity of a freight train that accelerates at a rate of 0.0500 m/s^2 for 8.00 min, starting with an initial velocity of 4.00 m/s? (b) If the train can slow down at a rate of 0.550 m/s^2 , how long will it take to come to a stop from this velocity? (c) How far will it travel in each case?

Exercise:**Problem:**

A fireworks shell is accelerated from rest to a velocity of 65.0 m/s over a distance of 0.250 m. (a) Calculate the acceleration. (b) How long did the acceleration last?

Solution:

a. $a = 8450 \text{ m/s}^2$;
b. $t = 0.0077 \text{ s}$

Exercise:

Problem:

A swan on a lake gets airborne by flapping its wings and running on top of the water. (a) If the swan must reach a velocity of 6.00 m/s to take off and it accelerates from rest at an average rate of 0.35 m/s^2 , how far will it travel before becoming airborne? (b) How long does this take?

Exercise:**Problem:**

A woodpecker's brain is specially protected from large accelerations by tendon-like attachments inside the skull. While pecking on a tree, the woodpecker's head comes to a stop from an initial velocity of 0.600 m/s in a distance of only 2.00 mm. (a) Find the acceleration in meters per second squared and in multiples of g , where $g = 9.80 \text{ m/s}^2$. (b) Calculate the stopping time. (c) The tendons cradling the brain stretch, making its stopping distance 4.50 mm (greater than the head and, hence, less acceleration of the brain). What is the brain's acceleration, expressed in multiples of g ?

Solution:

- a. $a = 9.18 g$;
- b. $t = 6.67 \times 10^{-3} \text{ s}$;
- c. $a = -40.0 \text{ m/s}^2$
 $a = 4.08 g$

Exercise:**Problem:**

An unwary football player collides with a padded goalpost while running at a velocity of 7.50 m/s and comes to a full stop after compressing the padding and his body 0.350 m. (a) What is his acceleration? (b) How long does the collision last?

Exercise:

Problem:

A care package is dropped out of a cargo plane and lands in the forest. If we assume the care package speed on impact is 54 m/s (123 mph), then what is its acceleration? Assume the trees and snow stops it over a distance of 3.0 m.

Solution:

Knowns: $x = 3 \text{ m}$, $v = 0 \text{ m/s}$, $v_0 = 54 \text{ m/s}$. We want a , so we can use this equation: $a = -486 \text{ m/s}^2$.

Exercise:**Problem:**

An express train passes through a station. It enters with an initial velocity of 22.0 m/s and accelerates opposite to the motion at a rate of 0.150 m/s^2 as it goes through. The station is 210.0 m long. (a) How fast is it going when the nose leaves the station? (b) How long is the nose of the train in the station? (c) If the train is 130 m long, what is the velocity of the end of the train as it leaves? (d) When does the end of the train leave the station?

Exercise:**Problem:**

Unreasonable results Dragsters can actually reach a top speed of 145.0 m/s in only 4.45 s. (a) Calculate the average acceleration for such a dragster. (b) Find the final velocity of this dragster starting from rest and accelerating at the rate found in (a) for 402.0 m (a quarter mile) without using any information on time. (c) Why is the final velocity greater than that used to find the average acceleration? (*Hint:* Consider whether the assumption of constant acceleration is valid for a dragster. If not, discuss whether the acceleration would be greater at the beginning or end of the run and what effect that would have on the final velocity.)

Solution:

a. $a = 32.58 \text{ m/s}^2$;

b. $v = 161.85 \text{ m/s}$;

c. $v > v_{\text{max}}$, because the assumption of constant acceleration is not valid for a dragster. A dragster changes gears and would have a greater acceleration in first gear than second gear than third gear, and so on. The acceleration would be greatest at the beginning, so it would not be accelerating at 32.6 m/s^2 during the last few meters, but substantially less, and the final velocity would be less than 162 m/s .

Glossary

two-body pursuit problem

a kinematics problem in which the unknowns are calculated by solving the kinematic equations simultaneously for two moving objects

Free Fall

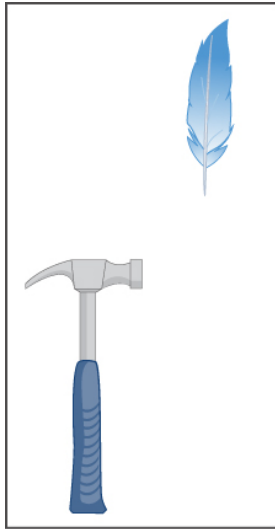
By the end of this section, you will be able to:

- Use the kinematic equations with the variables y and g to analyze free-fall motion.
- Describe how the values of the position, velocity, and acceleration change during a free fall.
- Solve for the position, velocity, and acceleration as functions of time when an object is in a free fall.

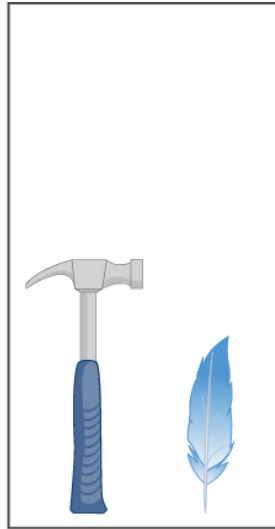
An interesting application of [\[link\]](#) through [\[link\]](#) is called *free fall*, which describes the motion of an object falling in a gravitational field, such as near the surface of Earth or other celestial objects of planetary size. Let's assume the body is falling in a straight line perpendicular to the surface, so its motion is one-dimensional. For example, we can estimate the depth of a vertical mine shaft by dropping a rock into it and listening for the rock to hit the bottom. But "falling," in the context of free fall, does not necessarily imply the body is moving from a greater height to a lesser height. If a ball is thrown upward, the equations of free fall apply equally to its ascent as well as its descent.

Gravity

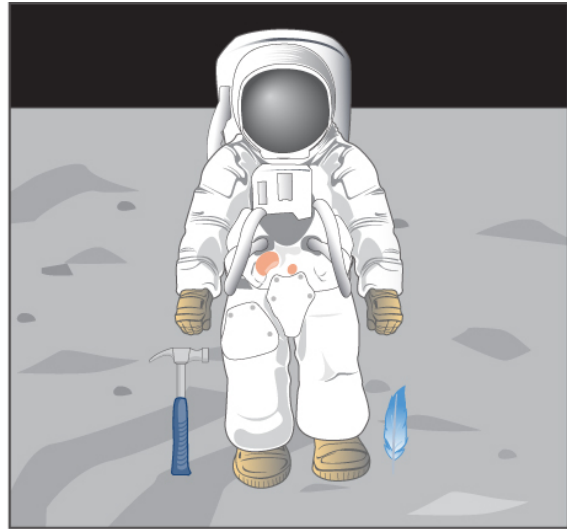
The most remarkable and unexpected fact about falling objects is that if air resistance and friction are negligible, then in a given location all objects fall toward the center of Earth with the *same constant acceleration, independent of their mass*. This experimentally determined fact is unexpected because we are so accustomed to the effects of air resistance and friction that we expect light objects to fall slower than heavy ones. Until Galileo Galilei (1564–1642) proved otherwise, people believed that a heavier object has a greater acceleration in a free fall. We now know this is not the case. In the absence of air resistance, heavy objects arrive at the ground at the same time as lighter objects when dropped from the same height [\[link\]](#).



In air



In a vacuum



In a vacuum (the hard way)

A hammer and a feather fall with the same constant acceleration if air resistance is negligible. This is a general characteristic of gravity not unique to Earth, as astronaut David R. Scott demonstrated in 1971 on the Moon, where the acceleration from gravity is only 1.67 m/s^2 and there is no atmosphere.

In the real world, air resistance can cause a lighter object to fall slower than a heavier object of the same size. A tennis ball reaches the ground after a baseball dropped at the same time. (It might be difficult to observe the difference if the height is not large.) Air resistance opposes the motion of an object through the air, and friction between objects—such as between clothes and a laundry chute or between a stone and a pool into which it is dropped—also opposes motion between them.

For the ideal situations of these first few chapters, an object *falling without air resistance or friction* is defined to be in **free fall**. The force of gravity causes objects to fall toward the center of Earth. The acceleration of free-falling objects is therefore called **acceleration due to gravity**. Acceleration due to gravity is constant, which means we can apply the kinematic equations to any falling object where air resistance and friction are negligible. This opens to us a broad class of interesting situations.

Acceleration due to gravity is so important that its magnitude is given its own symbol, g . It is constant at any given location on Earth and has the average value

Equation:

$$g = 9.81 \text{ m/s}^2 \text{ (or } 32.2 \text{ ft/s}^2\text{)}.$$

Although g varies from 9.78 m/s^2 to 9.83 m/s^2 , depending on latitude, altitude, underlying geological formations, and local topography, let's use an average value of 9.8 m/s^2 rounded to

two significant figures in this text unless specified otherwise. Neglecting these effects on the value of g as a result of position on Earth's surface, as well as effects resulting from Earth's rotation, we take the direction of acceleration due to gravity to be downward (toward the center of Earth). In fact, its direction *defines* what we call vertical. Note that whether acceleration a in the kinematic equations has the value $+g$ or $-g$ depends on how we define our coordinate system. If we define the upward direction as positive, then $a = -g = -9.8 \text{ m/s}^2$, and if we define the downward direction as positive, then $a = g = 9.8 \text{ m/s}^2$.

One-Dimensional Motion Involving Gravity

The best way to see the basic features of motion involving gravity is to start with the simplest situations and then progress toward more complex ones. So, we start by considering straight up-and-down motion with no air resistance or friction. These assumptions mean the velocity (if there is any) is vertical. If an object is dropped, we know the initial velocity is zero when in free fall. When the object has left contact with whatever held or threw it, the object is in free fall. When the object is thrown, it has the same initial speed in free fall as it did before it was released. When the object comes in contact with the ground or any other object, it is no longer in free fall and its acceleration of g is no longer valid. Under these circumstances, the motion is one-dimensional and has constant acceleration of magnitude g . We represent vertical displacement with the symbol y .

Note:

Kinematic Equations for Objects in Free Fall

We assume here that acceleration equals $-g$ (with the positive direction upward).

Equation:

$$v = v_0 - gt$$

Equation:

$$y = y_0 + v_0t - \frac{1}{2}gt^2$$

Equation:

$$v^2 = v_0^2 - 2g(y - y_0)$$

Note:

Free Fall

1. Decide on the sign of the acceleration of gravity. In [\[link\]](#) through [\[link\]](#), acceleration g is negative, which says the positive direction is upward and the negative direction is

downward. In some problems, it may be useful to have acceleration g as positive, indicating the positive direction is downward.

2. Draw a sketch of the problem. This helps visualize the physics involved.
3. Record the knowns and unknowns from the problem description. This helps devise a strategy for selecting the appropriate equations to solve the problem.
4. Decide which of [\[link\]](#) through [\[link\]](#) are to be used to solve for the unknowns.

Example:

Free Fall of a Ball

[\[link\]](#) shows the positions of a ball, at 1-s intervals, with an initial velocity of 4.9 m/s downward, that is thrown from the top of a 98-m-high building. (a) How much time elapses before the ball reaches the ground? (b) What is the velocity when it arrives at the ground?

	t (s)	x (m)	v (m/s)
	0	0	-4.9
	1	-9.8	-14.7
	2	-29.4	-24.5
	3	-58.8	-34.3
	4	-98.0	-44.1

The positions and velocities
at 1-s intervals of a ball
thrown downward from a tall
building at 4.9 m/s.

Strategy

Choose the origin at the top of the building with the positive direction upward and the negative direction downward. To find the time when the position is -98 m, we use [\[link\]](#), with $y_0 = 0$, $v_0 = -4.9$ m/s, and $g = 9.8$ m/s².

Solution

- a. Substitute the given values into the equation:

Equation:

$$y = y_0 + v_0 t - \frac{1}{2} g t^2$$

$$-98.0 \text{ m} = 0 - (4.9 \text{ m/s})t - \frac{1}{2}(9.8 \text{ m/s}^2)t^2.$$

This simplifies to

Equation:

$$t^2 + t - 20 = 0.$$

This is a quadratic equation with roots $t = -5.0 \text{ s}$ and $t = 4.0 \text{ s}$. The positive root is the one we are interested in, since time $t = 0$ is the time when the ball is released at the top of the building. (The time $t = -5.0 \text{ s}$ represents the fact that a ball thrown upward from the ground would have been in the air for 5.0 s when it passed by the top of the building moving downward at 4.9 m/s.)

b. Using [\[link\]](#), we have

Equation:

$$v = v_0 - g t = -4.9 \text{ m/s} - (9.8 \text{ m/s}^2)(4.0 \text{ s}) = -44.1 \text{ m/s}.$$

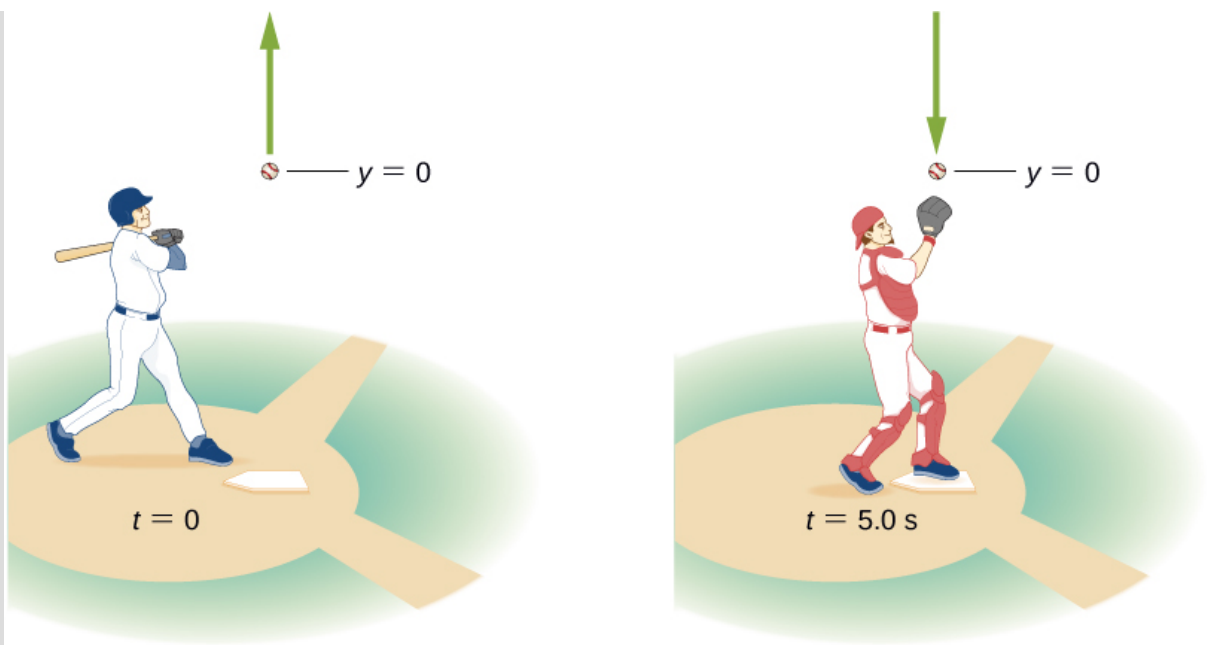
Significance

For situations when two roots are obtained from a quadratic equation in the time variable, we must look at the physical significance of both roots to determine which is correct. Since $t = 0$ corresponds to the time when the ball was released, the negative root would correspond to a time before the ball was released, which is not physically meaningful. When the ball hits the ground, its velocity is not immediately zero, but as soon as the ball interacts with the ground, its acceleration is not g and it accelerates with a different value over a short time to zero velocity. This problem shows how important it is to establish the correct coordinate system and to keep the signs of g in the kinematic equations consistent.

Example:

Vertical Motion of a Baseball

A batter hits a baseball straight upward at home plate and the ball is caught 5.0 s after it is struck [\[link\]](#). (a) What is the initial velocity of the ball? (b) What is the maximum height the ball reaches? (c) How long does it take to reach the maximum height? (d) What is the acceleration at the top of its path? (e) What is the velocity of the ball when it is caught? Assume the ball is hit and caught at the same location.



A baseball hit straight up is caught by the catcher 5.0 s later.

Strategy

Choose a coordinate system with a positive y -axis that is straight up and with an origin that is at the spot where the ball is hit and caught.

Solution

a. [link](#) gives

Equation:

$$y = y_0 + v_0 t - \frac{1}{2} g t^2$$

Equation:

$$0 = 0 + v_0(5.0 \text{ s}) - \frac{1}{2}(9.8 \text{ m/s}^2)(5.0 \text{ s})^2,$$

which gives $v_0 = 24.5 \text{ m/s}$.

b. At the maximum height, $v = 0$. With $v_0 = 24.5 \text{ m/s}$, [link](#) gives

Equation:

$$v^2 = v_0^2 - 2 g(y - y_0)$$

Equation:

$$0 = (24.5 \text{ m/s})^2 - 2(9.8 \text{ m/s}^2)(y - 0)$$

or

Equation:

$$y = 30.6 \text{ m.}$$

- c. To find the time when $v = 0$, we use [\[link\]](#):

Equation:

$$v = v_0 - gt$$

Equation:

$$0 = 24.5 \text{ m/s} - (9.8 \text{ m/s}^2)t.$$

This gives $t = 2.5 \text{ s}$. Since the ball rises for 2.5 s, the time to fall is 2.5 s.

- d. The acceleration is 9.8 m/s^2 everywhere, even when the velocity is zero at the top of the path. Although the velocity is zero at the top, it is changing at the rate of 9.8 m/s^2 downward.

- e. The velocity at $t = 5.0 \text{ s}$ can be determined with [\[link\]](#):

Equation:

$$\begin{aligned} v &= v_0 - gt \\ &= 24.5 \text{ m/s} - 9.8 \text{ m/s}^2(5.0 \text{ s}) \\ &= -24.5 \text{ m/s.} \end{aligned}$$

Significance

The ball returns with the speed it had when it left. This is a general property of free fall for any initial velocity. We used a single equation to go from throw to catch, and did not have to break the motion into two segments, upward and downward. We are used to thinking that the effect of gravity is to create free fall downward toward Earth. It is important to understand, as illustrated in this example, that objects moving upward away from Earth are also in a state of free fall.

Note:

Exercise:

Problem:

Check Your Understanding A chunk of ice breaks off a glacier and falls 30.0 m before it hits the water. Assuming it falls freely (there is no air resistance), how long does it take to hit the water? Which quantity increases faster, the speed of the ice chunk or its distance traveled?

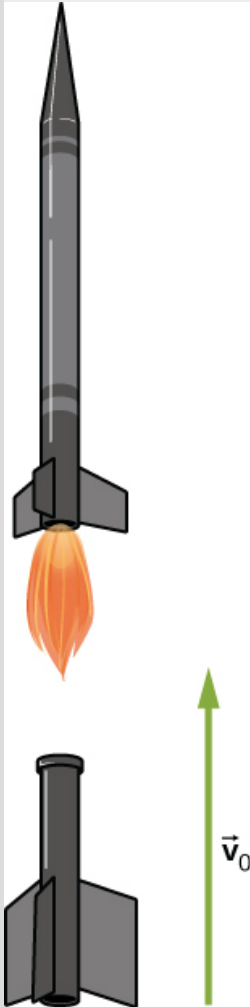
Solution:

It takes 2.47 s to hit the water. The quantity distance traveled increases faster.

Example:

Rocket Booster

A small rocket with a booster blasts off and heads straight upward. When at a height of 5.0 km and velocity of 200.0 m/s, it releases its booster. (a) What is the maximum height the booster attains? (b) What is the velocity of the booster at a height of 6.0 km? Neglect air resistance.



A rocket
releases its
booster at a
given height
and velocity.
How high
and how fast

does the
booster go?

Strategy

We need to select the coordinate system for the acceleration of gravity, which we take as negative downward. We are given the initial velocity of the booster and its height. We consider the point of release as the origin. We know the velocity is zero at the maximum position within the acceleration interval; thus, the velocity of the booster is zero at its maximum height, so we can use this information as well. From these observations, we use [\[link\]](#), which gives us the maximum height of the booster. We also use [\[link\]](#) to give the velocity at 6.0 km. The initial velocity of the booster is 200.0 m/s.

Solution

- a. From [\[link\]](#), $v^2 = v_0^2 - 2g(y - y_0)$. With $v = 0$ and $y_0 = 0$, we can solve for y :

Equation:

$$y = \frac{v_0^2}{2g} = \frac{(2.0 \times 10^2 \text{ m/s})^2}{2(9.8 \text{ m/s}^2)} = 2040.8 \text{ m}.$$

This solution gives the maximum height of the booster in our coordinate system, which has its origin at the point of release, so the maximum height of the booster is roughly 7.0 km.

- b. An altitude of 6.0 km corresponds to $y = 1.0 \times 10^3 \text{ m}$ in the coordinate system we are using. The other initial conditions are $y_0 = 0$, and $v_0 = 200.0 \text{ m/s}$.

We have, from [\[link\]](#),

Equation:

$$v^2 = (200.0 \text{ m/s})^2 - 2(9.8 \text{ m/s}^2)(1.0 \times 10^3 \text{ m}) \Rightarrow v = \pm 142.8 \text{ m/s}.$$

Significance

We have both a positive and negative solution in (b). Since our coordinate system has the positive direction upward, the +142.8 m/s corresponds to a positive upward velocity at 6000 m during the upward leg of the trajectory of the booster. The value $v = -142.8 \text{ m/s}$ corresponds to the velocity at 6000 m on the downward leg. This example is also important in that an object is given an initial velocity at the origin of our coordinate system, but the origin is at an altitude above the surface of Earth, which must be taken into account when forming the solution.

Note:

Visit [this site](#) to learn about graphing polynomials. The shape of the curve changes as the constants are adjusted. View the curves for the individual terms (for example, $y = bx$) to see how they add to generate the polynomial curve.

Summary

- An object in free fall experiences constant acceleration if air resistance is negligible.
- On Earth, all free-falling objects have an acceleration g due to gravity, which averages $g = 9.81 \text{ m/s}^2$.
- For objects in free fall, the upward direction is normally taken as positive for displacement, velocity, and acceleration.

Conceptual Questions

Exercise:

Problem:

What is the acceleration of a rock thrown straight upward on the way up? At the top of its flight? On the way down? Assume there is no air resistance.

Exercise:

Problem:

An object that is thrown straight up falls back to Earth. This is one-dimensional motion. (a) When is its velocity zero? (b) Does its velocity change direction? (c) Does the acceleration have the same sign on the way up as on the way down?

Solution:

a. at the top of its trajectory; b. yes, at the top of its trajectory; c. yes

Exercise:

Problem:

Suppose you throw a rock nearly straight up at a coconut in a palm tree and the rock just misses the coconut on the way up but hits the coconut on the way down. Neglecting air resistance and the slight horizontal variation in motion to account for the hit and miss of the coconut, how does the speed of the rock when it hits the coconut on the way down compare with what it would have been if it had hit the coconut on the way up? Is it more likely to dislodge the coconut on the way up or down? Explain.

Exercise:

Problem:

The severity of a fall depends on your speed when you strike the ground. All factors but the acceleration from gravity being the same, how many times higher could a safe fall occur on the Moon than on Earth (gravitational acceleration on the Moon is about one-sixth that of the Earth)?

Solution:

$$\text{Earth } v = v_0 - gt = -gt; \text{ Moon } v' = \frac{g}{6}t' \quad v = v' - gt = -\frac{g}{6}t' \quad t' = 6t; \text{ Earth } y = -\frac{1}{2}gt^2 \quad \text{Moon } y' = -\frac{1}{2}\frac{g}{6}(6t)^2 = -\frac{1}{2}g6t^2 = -6\left(\frac{1}{2}gt^2\right) = -6y$$

Exercise:

Problem:

How many times higher could an astronaut jump on the Moon than on Earth if her takeoff speed is the same in both locations (gravitational acceleration on the Moon is about one-sixth of that on Earth)?

Problems

Exercise:

Problem:

Calculate the displacement and velocity at times of (a) 0.500 s, (b) 1.00 s, (c) 1.50 s, and (d) 2.00 s for a ball thrown straight up with an initial velocity of 15.0 m/s. Take the point of release to be $y_0 = 0$.

Exercise:

Problem:

Calculate the displacement and velocity at times of (a) 0.500 s, (b) 1.00 s, (c) 1.50 s, (d) 2.00 s, and (e) 2.50 s for a rock thrown straight down with an initial velocity of 14.0 m/s from the Verrazano Narrows Bridge in New York City. The roadway of this bridge is 70.0 m above the water.

Solution:

a. $y = -8.23 \text{ m}$;
 $v_1 = -18.9 \text{ m/s}$

b. $y = -18.9 \text{ m}$;
 $v_2 = -23.8 \text{ m/s}$

c. $y = -32.0 \text{ m}$;
 $v_3 = -28.7 \text{ m/s}$

d. $y = -47.6 \text{ m}$;
 $v_4 = -33.6 \text{ m/s}$

e. $y = -65.6 \text{ m}$
 $v_5 = -38.5 \text{ m/s}$

Exercise:

Problem:

A basketball referee tosses the ball straight up for the starting tip-off. At what velocity must a basketball player leave the ground to rise 1.25 m above the floor in an attempt to get the ball?

Exercise:**Problem:**

A rescue helicopter is hovering over a person whose boat has sunk. One of the rescuers throws a life preserver straight down to the victim with an initial velocity of 1.40 m/s and observes that it takes 1.8 s to reach the water. (a) List the knowns in this problem. (b) How high above the water was the preserver released? Note that the downdraft of the helicopter reduces the effects of air resistance on the falling life preserver, so that an acceleration equal to that of gravity is reasonable.

Solution:

a. Knowns: $a = -9.8 \text{ m/s}^2$ $v_0 = -1.4 \text{ m/s}$ $t = 1.8 \text{ s}$ $y_0 = 0 \text{ m}$;

b.

$y = y_0 + v_0 t - \frac{1}{2} g t^2$ $y = v_0 t - \frac{1}{2} g t = -1.4 \text{ m/s}(1.8 \text{ sec}) - \frac{1}{2} (9.8)(1.8 \text{ s})^2 = -18.4 \text{ m}$
and the origin is at the rescuers, who are 18.4 m above the water.

Exercise:**Problem:**

Unreasonable results A dolphin in an aquatic show jumps straight up out of the water at a velocity of 15.0 m/s. (a) List the knowns in this problem. (b) How high does his body rise above the water? To solve this part, first note that the final velocity is now a known, and identify its value. Then, identify the unknown and discuss how you chose the appropriate equation to solve for it. After choosing the equation, show your steps in solving for the unknown, checking units, and discuss whether the answer is reasonable. (c) How long a time is the dolphin in the air? Neglect any effects resulting from his size or orientation.

Exercise:**Problem:**

A diver bounces straight up from a diving board, avoiding the diving board on the way down, and falls feet first into a pool. She starts with a velocity of 4.00 m/s and her takeoff point is 1.80 m above the pool. (a) What is her highest point above the board? (b) How long a time are her feet in the air? (c) What is her velocity when her feet hit the water?

Solution:

a. $v^2 = v_0^2 - 2g(y - y_0)$ $y_0 = 0$ $v = 0$ $y = \frac{v_0^2}{2g} = \frac{(4.0 \text{ m/s})^2}{2(9.80)} = 0.82 \text{ m}$; b. to the apex
 $v = 0.41 \text{ s}$ times 2 to the board = 0.82 s from the board to the water
 $y = y_0 + v_0t - \frac{1}{2}gt^2$ $y = -1.80 \text{ m}$ $y_0 = 0$ $v_0 = 4.0 \text{ m/s}$
 $-1.8 = 4.0t - 4.9t^2$ $4.9t^2 - 4.0t - 1.80 = 0$, solution to quadratic equation gives
1.13 s; c. $v^2 = v_0^2 - 2g(y - y_0)$ $y_0 = 0$ $v_0 = 4.0 \text{ m/s}$ $y = -1.80 \text{ m}$
 $v = 7.16 \text{ m/s}$

Exercise:

Problem:

(a) Calculate the height of a cliff if it takes 2.35 s for a rock to hit the ground when it is thrown straight up from the cliff with an initial velocity of 8.00 m/s. (b) How long a time would it take to reach the ground if it is thrown straight down with the same speed?

Exercise:

Problem:

A very strong, but inept, shot putter puts the shot straight up vertically with an initial velocity of 11.0 m/s. How long a time does he have to get out of the way if the shot was released at a height of 2.20 m and he is 1.80 m tall?

Solution:

Time to the apex: $t = 1.12 \text{ s}$ times 2 equals 2.24 s to a height of 2.20 m. To 1.80 m in height is an additional 0.40 m.

$$y = y_0 + v_0t - \frac{1}{2}gt^2 \quad y = -0.40 \text{ m} \quad y_0 = 0 \quad v_0 = -11.0 \text{ m/s}$$

$$y = y_0 + v_0t - \frac{1}{2}gt^2 \quad y = -0.40 \text{ m} \quad y_0 = 0 \quad v_0 = -11.0 \text{ m/s}$$

$$-0.40 = -11.0t - 4.9t^2 \quad \text{or} \quad 4.9t^2 + 11.0t - 0.40 = 0$$

Take the positive root, so the time to go the additional 0.4 m is 0.04 s. Total time is
 $2.24 \text{ s} + 0.04 \text{ s} = 2.28 \text{ s}$.

Exercise:

Problem:

You throw a ball straight up with an initial velocity of 15.0 m/s. It passes a tree branch on the way up at a height of 7.0 m. How much additional time elapses before the ball passes the tree branch on the way back down?

Exercise:

Problem:

A kangaroo can jump over an object 2.50 m high. (a) Considering just its vertical motion, calculate its vertical speed when it leaves the ground. (b) How long a time is it in the air?

Solution:

a. $v^2 = v_0^2 - 2g(y - y_0)$ $y_0 = 0$ $v = 0$ $y = 2.50$ m; b. $t = 0.72$ s times 2 gives 1.44 s
 $v_0^2 = 2gy \Rightarrow v_0 = \sqrt{2(9.80)(2.50)} = 7.0$ m/s
in the air

Exercise:**Problem:**

Standing at the base of one of the cliffs of Mt. Arapiles in Victoria, Australia, a hiker hears a rock break loose from a height of 105.0 m. He can't see the rock right away, but then does, 1.50 s later. (a) How far above the hiker is the rock when he can hear it? (b) How much time does he have to move before the rock hits his head?

Exercise:**Problem:**

There is a 250-m-high cliff at Half Dome in Yosemite National Park in California. Suppose a boulder breaks loose from the top of this cliff. (a) How fast will it be going when it strikes the ground? (b) Assuming a reaction time of 0.300 s, how long a time will a tourist at the bottom have to get out of the way after hearing the sound of the rock breaking loose (neglecting the height of the tourist, which would become negligible anyway if hit)? The speed of sound is 335.0 m/s on this day.

Solution:

a. $v = 70.0$ m/s; b. time heard after rock begins to fall: 0.75 s, time to reach the ground: 6.09 s

Glossary

acceleration due to gravity

acceleration of an object as a result of gravity

free fall

the state of movement that results from gravitational force only

Finding Velocity and Displacement from Acceleration

By the end of this section, you will be able to:

- Derive the kinematic equations for constant acceleration using integral calculus.
- Use the integral formulation of the kinematic equations in analyzing motion.
- Find the functional form of velocity versus time given the acceleration function.
- Find the functional form of position versus time given the velocity function.

This section assumes you have enough background in calculus to be familiar with integration. In [Instantaneous Velocity and Speed](#) and [Average and Instantaneous Acceleration](#) we introduced the kinematic functions of velocity and acceleration using the derivative. By taking the derivative of the position function we found the velocity function, and likewise by taking the derivative of the velocity function we found the acceleration function. Using integral calculus, we can work backward and calculate the velocity function from the acceleration function, and the position function from the velocity function.

Kinematic Equations from Integral Calculus

Let's begin with a particle with an acceleration $a(t)$ which is a known function of time. Since the time derivative of the velocity function is acceleration,

Equation:

$$\frac{d}{dt}v(t) = a(t),$$

we can take the indefinite integral of both sides, finding

Equation:

$$\int \frac{d}{dt}v(t)dt = \int a(t)dt + C_1,$$

where C_1 is a constant of integration. Since $\int \frac{d}{dt}v(t)dt = v(t)$, the velocity is given by

Note:

Equation:

$$v(t) = \int a(t)dt + C_1.$$

Similarly, the time derivative of the position function is the velocity function,

Equation:

$$\frac{d}{dt}x(t) = v(t).$$

Thus, we can use the same mathematical manipulations we just used and find

Note:

Equation:

$$x(t) = \int v(t)dt + C_2,$$

where C_2 is a second constant of integration.

We can derive the kinematic equations for a constant acceleration using these integrals. With $a(t) = a$ a constant, and doing the integration in [\[link\]](#), we find

Equation:

$$v(t) = \int a dt + C_1 = at + C_1.$$

If the initial velocity is $v(0) = v_0$, then

Equation:

$$v_0 = 0 + C_1.$$

Then, $C_1 = v_0$ and

Equation:

$$v(t) = v_0 + at,$$

which is [\[link\]](#). Substituting this expression into [\[link\]](#) gives

Equation:

$$x(t) = \int (v_0 + at)dt + C_2.$$

Doing the integration, we find

Equation:

$$x(t) = v_0 t + \frac{1}{2} a t^2 + C_2.$$

If $x(0) = x_0$, we have

Equation:

$$x_0 = 0 + 0 + C_2;$$

so, $C_2 = x_0$. Substituting back into the equation for $x(t)$, we finally have

Equation:

$$x(t) = x_0 + v_0 t + \frac{1}{2} a t^2,$$

which is [\[link\]](#).

Example:

Motion of a Motorboat

A motorboat is traveling at a constant velocity of 5.0 m/s when it starts to accelerate opposite to the motion to arrive at the dock. Its acceleration is $a(t) = -\frac{1}{4}t \text{ m/s}^3$. (a) What is the velocity function of the motorboat? (b) At what time does the velocity reach zero? (c) What is the position function of the motorboat? (d) What is the displacement of the motorboat from the time it begins to accelerate opposite to the motion to when the velocity is zero? (e) Graph the velocity and position functions.

Strategy

(a) To get the velocity function we must integrate and use initial conditions to find the constant of integration. (b) We set the velocity function equal to zero and solve for t . (c) Similarly, we must integrate to find the position function and use initial conditions to find the constant of integration. (d) Since the initial position is taken to be zero, we only have to evaluate the position function at the time when the velocity is zero.

Solution

We take $t = 0$ to be the time when the boat starts to accelerate opposite to the motion.

- a. From the functional form of the acceleration we can solve [\[link\]](#) to get $v(t)$:

Equation:

$$v(t) = \int a(t) dt + C_1 = \int -\frac{1}{4}t dt + C_1 = -\frac{1}{8}t^2 + C_1.$$

At $t = 0$ we have $v(0) = 5.0 \text{ m/s} = 0 + C_1$, so $C_1 = 5.0 \text{ m/s}$ or $v(t) = 5.0 \text{ m/s} - \frac{1}{8}t^2$.

- b. $v(t) = 0 = 5.0 \text{ m/s} - \frac{1}{8}t^2 \text{ m/s}^3 \Rightarrow t = 6.3 \text{ s}$

- c. Solve [\[link\]](#):

Equation:

$$x(t) = \int v(t)dt + C_2 = \int (5.0 - \frac{1}{8}t^2)dt + C_2 = 5.0t \text{ m/s} - \frac{1}{24}t^3 \text{ m/s}^3 + C_2.$$

At $t = 0$, we set $x(0) = 0 = x_0$, since we are only interested in the displacement from when the boat starts to accelerate opposite to the motion. We have

Equation:

$$x(0) = 0 = C_2.$$

Therefore, the equation for the position is

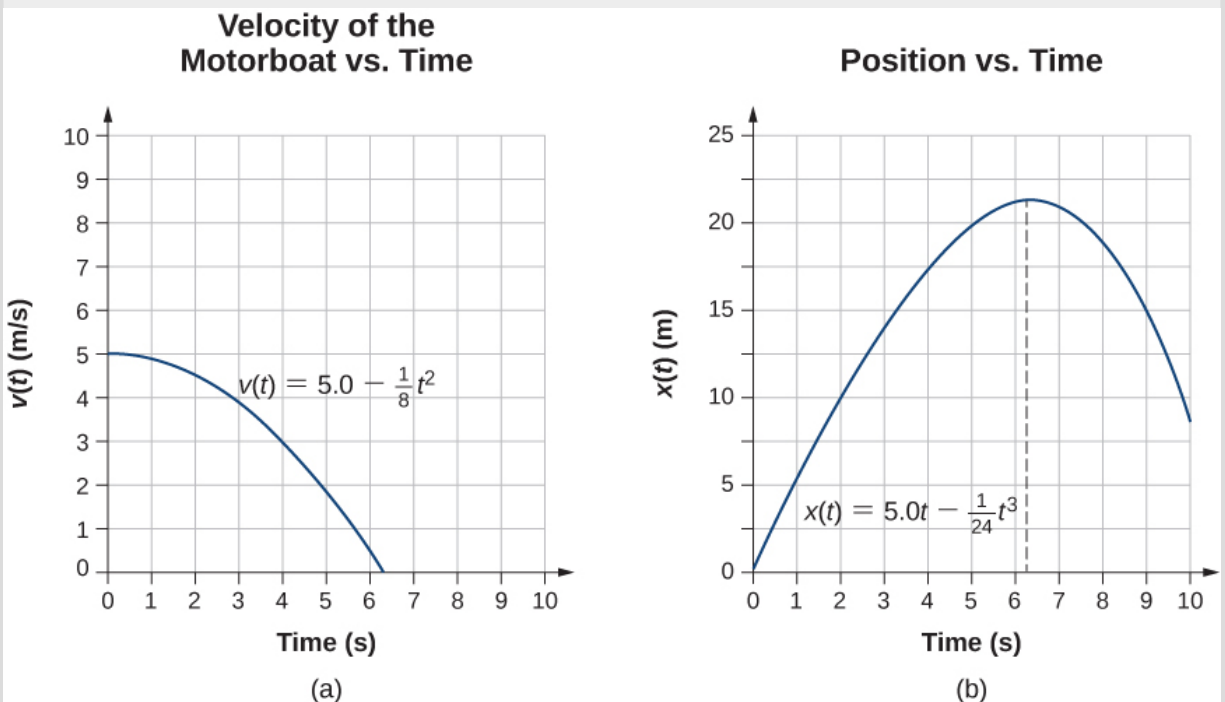
Equation:

$$x(t) = 5.0t - \frac{1}{24}t^3.$$

- d. Since the initial position is taken to be zero, we only have to evaluate the position function at the time when the velocity is zero. This occurs at $t = 6.3$ s. Therefore, the displacement is

Equation:

$$x(6.3) = 5.0(6.3 \text{ s}) - \frac{1}{24}(6.3 \text{ s})^3 = 21.1 \text{ m}.$$



(a) Velocity of the motorboat as a function of time. The motorboat decreases its velocity to zero in 6.3 s. At times greater than this, velocity becomes negative—meaning, the boat is reversing direction. (b) Position of the motorboat as a function of time. At $t = 6.3$ s, the velocity is zero and the boat has stopped. At times greater than this, the velocity becomes

negative—meaning, if the boat continues to move with the same acceleration, it reverses direction and heads back toward where it originated.

Significance

The acceleration function is linear in time so the integration involves simple polynomials. In [\[link\]](#), we see that if we extend the solution beyond the point when the velocity is zero, the velocity becomes negative and the boat reverses direction. This tells us that solutions can give us information outside our immediate interest and we should be careful when interpreting them.

Note:

Exercise:

Problem:

Check Your Understanding A particle starts from rest and has an acceleration function $a(t) = (5 - (10\frac{1}{s})t)\frac{m}{s^2}$. (a) What is the velocity function? (b) What is the position function? (c) When is the velocity zero?

Solution:

- a. The velocity function is the integral of the acceleration function plus a constant of integration. By [\[link\]](#),

$$v(t) = \int a(t)dt + C_1 = \int (5 - 10t)dt + C_1 = 5t - 5t^2 + C_1.$$

Since $v(0) = 0$, we have $C_1 = 0$; so,

$$v(t) = 5t - 5t^2.$$

- b. By [\[link\]](#),

$$x(t) = \int v(t)dt + C_2 = \int (5t - 5t^2)dt + C_2 = \frac{5}{2}t^2 - \frac{5}{3}t^3 + C_2.$$

Since $x(0) = 0$, we have $C_2 = 0$, and

$$x(t) = \frac{5}{2}t^2 - \frac{5}{3}t^3.$$

- c. The velocity can be written as $v(t) = 5t(1 - t)$, which equals zero at $t = 0$, and $t = 1$ s.

Summary

- Integral calculus gives us a more complete formulation of kinematics.
- If acceleration $a(t)$ is known, we can use integral calculus to derive expressions for velocity $v(t)$ and position $x(t)$.
- If acceleration is constant, the integral equations reduce to [\[link\]](#) and [\[link\]](#) for motion with constant acceleration.

Key Equations

Displacement	$\Delta x = x_f - x_i$
Total displacement	$\Delta x_{\text{Total}} = \sum \Delta x_i$
Average velocity (for constant acceleration)	$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1}$
Instantaneous velocity	$v(t) = \frac{dx(t)}{dt}$
Average speed	Average speed = $\bar{s} = \frac{\text{Total distance}}{\text{Elapsed time}}$
Instantaneous speed	Instantaneous speed = $ v(t) $
Average acceleration	$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}$
Instantaneous acceleration	$a(t) = \frac{dv(t)}{dt}$
Position from average velocity	$x = x_0 + \bar{v}t$
Average velocity	$\bar{v} = \frac{v_0 + v}{2}$
Velocity from acceleration	$v = v_0 + at$ (constant a)
Position from velocity and acceleration	$x = x_0 + v_0t + \frac{1}{2}at^2$ (constant a)
Velocity from distance	$v^2 = v_0^2 + 2a(x - x_0)$ (constant a)
Velocity of free fall	$v = v_0 - gt$ (positive upward)
Height of free fall	$y = y_0 + v_0t - \frac{1}{2}gt^2$
Velocity of free fall from height	$v^2 = v_0^2 - 2g(y - y_0)$
Velocity from acceleration	$v(t) = \int a(t)dt + C_1$
Position from velocity	$x(t) = \int v(t)dt + C_2$

Conceptual Questions

Exercise:

Problem:

When given the acceleration function, what additional information is needed to find the velocity function and position function?

Problems

Exercise:

Problem:

The acceleration of a particle varies with time according to the equation $a(t) = pt^2 - qt^3$. Initially, the velocity and position are zero. (a) What is the velocity as a function of time? (b) What is the position as a function of time?

Exercise:

Problem:

Between $t = 0$ and $t = t_0$, a rocket moves straight upward with an acceleration given by $a(t) = A - Bt^{1/2}$, where A and B are constants. (a) If x is in meters and t is in seconds, what are the units of A and B ? (b) If the rocket starts from rest, how does the velocity vary between $t = 0$ and $t = t_0$? (c) If its initial position is zero, what is the rocket's position as a function of time during this same time interval?

Solution:

a. $A = \text{m/s}^2$ $B = \text{m/s}^{5/2}$;

b. $v(t) = \int a(t)dt + C_1 = \int (A - Bt^{1/2})dt + C_1 = At - \frac{2}{3}Bt^{3/2} + C_1$,

$$v(0) = 0 = C_1 \text{ so } v(t_0) = At_0 - \frac{2}{3}Bt_0^{3/2}$$

c. $x(t) = \int v(t)dt + C_2 = \int \left(At - \frac{2}{3}Bt^{3/2} \right)dt + C_2 = \frac{1}{2}At^2 - \frac{4}{15}Bt^{5/2} + C_2$

$$x(0) = 0 = C_2 \text{ so } x(t_0) = \frac{1}{2}At_0^2 - \frac{4}{15}Bt_0^{5/2}$$

Exercise:

Problem:

The velocity of a particle moving along the x -axis varies with time according to $v(t) = A + Bt^{-1}$, where $A = 2 \text{ m/s}$, $B = 0.25 \text{ m}$, and $1.0 \text{ s} \leq t \leq 8.0 \text{ s}$. Determine the acceleration and position of the particle at $t = 2.0 \text{ s}$ and $t = 5.0 \text{ s}$. Assume that $x(t = 1 \text{ s}) = 0$.

Exercise:**Problem:**

A particle at rest leaves the origin with its velocity increasing with time according to $v(t) = 3.2t$ m/s. At 5.0 s, the particle's velocity starts decreasing according to $[16.0 - 1.5(t - 5.0)]$ m/s. This decrease continues until $t = 11.0$ s, after which the particle's velocity remains constant at 7.0 m/s. (a) What is the acceleration of the particle as a function of time? (b) What is the position of the particle at $t = 2.0$ s, $t = 7.0$ s, and $t = 12.0$ s?

Solution:

$$a(t) = 3.2\text{m/s}^2 \quad t \leq 5.0 \text{ s}$$

a. $a(t) = 1.5\text{m/s}^2 \quad 5.0 \text{ s} \leq t \leq 11.0 \text{ s};$

$$a(t) = 0\text{m/s}^2 \quad t > 11.0 \text{ s}$$

b.

$$x(t) = \int v(t)dt + C_2 = \int 3.2tdt + C_2 = 1.6t^2 + C_2$$

$$t \leq 5.0 \text{ s}$$

$$x(0) = 0 \Rightarrow C_2 = 0 \quad \text{therefore, } x(2.0 \text{ s}) = 6.4 \text{ m}$$

$$x(t) = \int v(t)dt + C_2 = \int [16.0 - 1.5(t - 5.0)]dt + C_2 = 16t - 1.5\left(\frac{t^2}{2} - 5.0t\right) + C_2$$

$$5.0 \leq t \leq 11.0 \text{ s}$$

$$x(5 \text{ s}) = 1.6(5.0)^2 = 40 \text{ m} = 16(5.0 \text{ s}) - 1.5\left(\frac{5^2}{2} - 5.0(5.0)\right) + C_2$$

$$40 = 98.75 + C_2 \Rightarrow C_2 = -58.75$$

$$x(7.0 \text{ s}) = 16(7.0) - 1.5\left(\frac{7^2}{2} - 5.0(7)\right) - 58.75 = 69 \text{ m}$$

$$x(t) = \int 7.0dt + C_2 = 7t + C_2$$

$$t \geq 11.0 \text{ s}$$

$$x(11.0 \text{ s}) = 16(11) - 1.5\left(\frac{11^2}{2} - 5.0(11)\right) - 58.75 = 109 = 7(11.0 \text{ s}) + C_2 \Rightarrow C_2 = 32 \text{ m}$$

$$x(t) = 7t + 32 \text{ m}$$

$$x \geq 11.0 \text{ s} \Rightarrow x(12.0 \text{ s}) = 7(12) + 32 = 116 \text{ m}$$

Additional Problems**Exercise:**

Problem:

Professional baseball player Nolan Ryan could pitch a baseball at approximately 160.0 km/h. At that average velocity, how long did it take a ball thrown by Ryan to reach home plate, which is 18.4 m from the pitcher's mound? Compare this with the average reaction time of a human to a visual stimulus, which is 0.25 s.

Exercise:**Problem:**

An airplane leaves Chicago and makes the 3000-km trip to Los Angeles in 5.0 h. A second plane leaves Chicago one-half hour later and arrives in Los Angeles at the same time. Compare the average velocities of the two planes. Ignore the curvature of Earth and the difference in altitude between the two cities.

Solution:

Take west to be the positive direction.

1st plane: $\bar{v} = 600 \text{ km/h}$

2nd plane $\bar{v} = 667.0 \text{ km/h}$

Exercise:**Problem:**

Unreasonable Results A cyclist rides 16.0 km east, then 8.0 km west, then 8.0 km east, then 32.0 km west, and finally 11.2 km east. If his average velocity is 24 km/h, how long did it take him to complete the trip? Is this a reasonable time?

Exercise:**Problem:**

An object has an acceleration of $+1.2 \text{ cm/s}^2$. At $t = 4.0 \text{ s}$, its velocity is -3.4 cm/s . Determine the object's velocities at $t = 1.0 \text{ s}$ and $t = 6.0 \text{ s}$.

Solution:

$$a = \frac{v - v_0}{t - t_0}, t = 0, a = \frac{-3.4 \text{ cm/s} - v_0}{4 \text{ s}} = 1.2 \text{ cm/s}^2 \Rightarrow v_0 = -8.2 \text{ cm/s}$$

$$v = v_0 + at = -8.2 + 1.2 t; v = -7.0 \text{ cm/s} \quad v = -1.0 \text{ cm/s}$$

Exercise:**Problem:**

A particle moves along the x -axis according to the equation $x(t) = 2.0 - 4.0t^2 \text{ m}$. What are the velocity and acceleration at $t = 2.0 \text{ s}$ and $t = 5.0 \text{ s}$?

Exercise:

Problem:

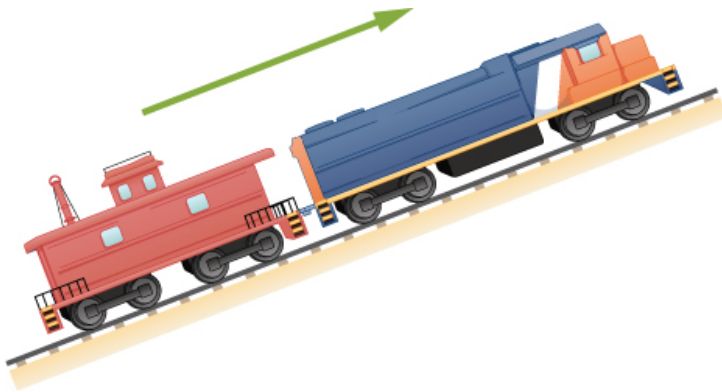
A particle moving at constant acceleration has velocities of 2.0 m/s at $t = 2.0 \text{ s}$ and -7.6 m/s at $t = 5.2 \text{ s}$. What is the acceleration of the particle?

Solution:

$$a = -3 \text{ m/s}^2$$

Exercise:**Problem:**

A train is moving up a steep grade at constant velocity (see following figure) when its caboose breaks loose and starts rolling freely along the track. After 5.0 s , the caboose is 30 m behind the train. What is the acceleration of the caboose?

**Exercise:****Problem:**

An electron is moving in a straight line with a velocity of $4.0 \times 10^5 \text{ m/s}$. It enters a region 5.0 cm long where it undergoes an acceleration of $6.0 \times 10^{12} \text{ m/s}^2$ along the same straight line. (a) What is the electron's velocity when it emerges from this region? b) How long does the electron take to cross the region?

Solution:

a.

$$v = 8.7 \times 10^5 \text{ m/s};$$

b. $t = 7.8 \times 10^{-8} \text{ s}$

Exercise:

Problem:

An ambulance driver is rushing a patient to the hospital. While traveling at 72 km/h, she notices the traffic light at the upcoming intersections has turned amber. To reach the intersection before the light turns red, she must travel 50 m in 2.0 s. (a) What minimum acceleration must the ambulance have to reach the intersection before the light turns red? (b) What is the speed of the ambulance when it reaches the intersection?

Exercise:**Problem:**

A motorcycle that is slowing down uniformly covers 2.0 successive km in 80 s and 120 s, respectively. Calculate (a) the acceleration of the motorcycle and (b) its velocity at the beginning and end of the 2-km trip.

Solution:

$1 \text{ km} = v_0(80.0 \text{ s}) + \frac{1}{2}a(80.0)^2$; $2 \text{ km} = v_0(200.0) + \frac{1}{2}a(200.0)^2$ solve simultaneously to get $a = -\frac{0.1}{2400.0} \text{ km/s}^2$ and $v_0 = 0.014167 \text{ km/s}$, which is 51.0 km/h. Velocity at the end of the trip is $v = 21.0 \text{ km/h}$.

Exercise:**Problem:**

A cyclist travels from point A to point B in 10 min. During the first 2.0 min of her trip, she maintains a uniform acceleration of 0.090 m/s^2 . She then travels at constant velocity for the next 5.0 min. Next, she accelerates opposite to the motion at a constant rate so that she comes to a rest at point B 3.0 min later. (a) Sketch the velocity-versus-time graph for the trip. (b) What is the acceleration during the last 3 min? (c) How far does the cyclist travel?

Exercise:**Problem:**

Two trains are moving at 30 m/s in opposite directions on the same track. The engineers see simultaneously that they are on a collision course and apply the brakes when they are 1000 m apart. Assuming both trains have the same acceleration, what must this acceleration be if the trains are to stop just short of colliding?

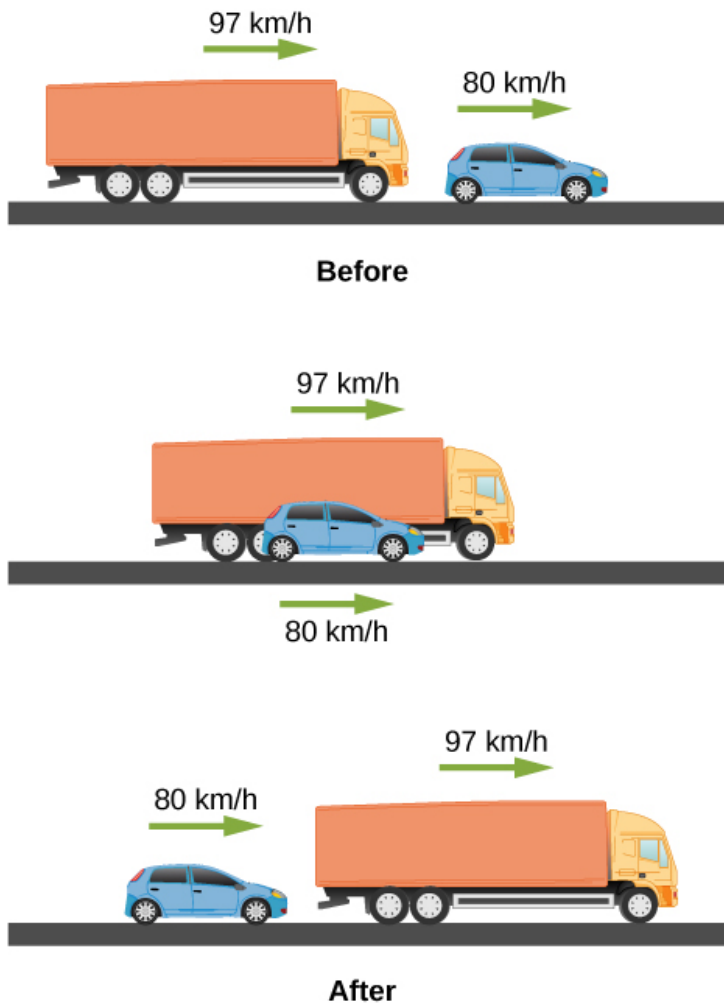
Solution:

$$a = -0.9 \text{ m/s}^2$$

Exercise:

Problem:

A 10.0-m-long truck moving with a constant velocity of 97.0 km/h passes a 3.0-m-long car moving with a constant velocity of 80.0 km/h. How much time elapses between the moment the front of the truck is even with the back of the car and the moment the back of the truck is even with the front of the car?

**Exercise:****Problem:**

A police car waits in hiding slightly off the highway. A speeding car is spotted by the police car doing 40 m/s. At the instant the speeding car passes the police car, the police car accelerates from rest at 4 m/s^2 to catch the speeding car. How long does it take the police car to catch the speeding car?

Solution:

Equation for the speeding car: This car has a constant velocity, which is the average velocity, and is not accelerating, so use the equation for displacement with $x_0 = 0$: $x = x_0 + \bar{v}t = \bar{v}t$;
Equation for the police car: This car is accelerating, so use the equation for displacement with $x_0 = 0$ and $v_0 = 0$, since the police car starts from rest:

$x = x_0 + v_0t + \frac{1}{2}at^2 = \frac{1}{2}at^2$; Now we have an equation of motion for each car with a common parameter, which can be eliminated to find the solution. In this case, we solve for t .
Step 1, eliminating x : $x = \bar{v}t = \frac{1}{2}at^2$; Step 2, solving for t : $t = \frac{2\bar{v}}{a}$. The speeding car has a constant velocity of 40 m/s, which is its average velocity. The acceleration of the police car is 4 m/s². Evaluating t , the time for the police car to reach the speeding car, we have
 $t = \frac{2\bar{v}}{a} = \frac{2(40)}{4} = 20$ s.

Exercise:

Problem:

Pablo is running in a half marathon at a velocity of 3 m/s. Another runner, Jacob, is 50 meters behind Pablo with the same velocity. Jacob begins to accelerate at 0.05 m/s². (a) How long does it take Jacob to catch Pablo? (b) What is the distance covered by Jacob? (c) What is the final velocity of Jacob?

Exercise:

Problem:

Unreasonable results A runner approaches the finish line and is 75 m away; her speed at this position is 8 m/s. She accelerates opposite to the motion at this point at 0.5 m/s². How long does it take her to cross the finish line from 75 m away? Is this reasonable?

Solution:

At this acceleration she comes to a full stop in $t = \frac{-v_0}{a} = \frac{8}{0.5} = 16$ s, but the distance covered is $x = 8 \text{ m/s}(16 \text{ s}) - \frac{1}{2}(0.5)(16 \text{ s})^2 = 64 \text{ m}$, which is less than the distance she is away from the finish line, so she never finishes the race.

Exercise:

Problem:

An airplane accelerates at 5.0 m/s² for 30.0 s. During this time, it covers a distance of 10.0 km. What are the initial and final velocities of the airplane?

Exercise:

Problem:

Compare the distance traveled of an object that undergoes a change in velocity that is twice its initial velocity with an object that changes its velocity by four times its initial velocity over the same time period. The accelerations of both objects are constant.

Solution:

$$x_1 = \frac{3}{2}v_0t$$

$$x_2 = \frac{5}{3}x_1$$

Exercise:

Problem:

An object is moving east with a constant velocity and is at position x_0 at time $t_0 = 0$. (a) With what acceleration must the object have for its total displacement to be zero at a later time t ? (b) What is the physical interpretation of the solution in the case for $t \rightarrow \infty$?

Exercise:

Problem:

A ball is thrown straight up. It passes a 2.00-m-high window 7.50 m off the ground on its path up and takes 1.30 s to go past the window. What was the ball's initial velocity?

Solution:

$$v_0 = 7.9 \text{ m/s velocity at the bottom of the window.}$$

$$v = 7.9 \text{ m/s}$$

$$v_0 = 14.1 \text{ m/s}$$

Exercise:

Problem:

A coin is dropped from a hot-air balloon that is 300 m above the ground and rising at 10.0 m/s upward. For the coin, find (a) the maximum height reached, (b) its position and velocity 4.00 s after being released, and (c) the time before it hits the ground.

Exercise:

Problem:

A soft tennis ball is dropped onto a hard floor from a height of 1.50 m and rebounds to a height of 1.10 m. (a) Calculate its velocity just before it strikes the floor. (b) Calculate its velocity just after it leaves the floor on its way back up. (c) Calculate its acceleration during contact with the floor if that contact lasts 3.50 ms (3.50×10^{-3} s) (d) How much did the ball compress during its collision with the floor, assuming the floor is absolutely rigid?

Solution:

- a. $v = 5.42 \text{ m/s}$;
- b. $v = 4.64 \text{ m/s}$;
- c. $a = 2874.28 \text{ m/s}^2$;
- d. $(x - x_0) = 5.11 \times 10^{-3} \text{ m}$

Exercise:

Problem:

Unreasonable results. A raindrop falls from a cloud 100 m above the ground. Neglect air resistance. What is the speed of the raindrop when it hits the ground? Is this a reasonable number?

Exercise:**Problem:**

Compare the time in the air of a basketball player who jumps 1.0 m vertically off the floor with that of a player who jumps 0.3 m vertically.

Solution:

Consider the players fall from rest at the height 1.0 m and 0.3 m.

0.9 s

0.5 s

Exercise:**Problem:**

Suppose that a person takes 0.5 s to react and move his hand to catch an object he has dropped. (a) How far does the object fall on Earth, where $g = 9.8 \text{ m/s}^2$? (b) How far does the object fall on the Moon, where the acceleration due to gravity is $1/6$ of that on Earth?

Exercise:**Problem:**

A hot-air balloon rises from ground level at a constant velocity of 3.0 m/s. One minute after liftoff, a sandbag is dropped accidentally from the balloon. Calculate (a) the time it takes for the sandbag to reach the ground and (b) the velocity of the sandbag when it hits the ground.

Solution:

a. $t = 6.37 \text{ s}$ taking the positive root;

b. $v = 59.5 \text{ m/s}$

Exercise:**Problem:**

(a) A world record was set for the men's 100-m dash in the 2008 Olympic Games in Beijing by Usain Bolt of Jamaica. Bolt "coasted" across the finish line with a time of 9.69 s. If we assume that Bolt accelerated for 3.00 s to reach his maximum speed, and maintained that speed for the rest of the race, calculate his maximum speed and his acceleration. (b) During the same Olympics, Bolt also set the world record in the 200-m dash with a time of 19.30 s. Using the same assumptions as for the 100-m dash, what was his maximum speed for this race?

Exercise:

Problem:

An object is dropped from a height of 75.0 m above ground level. (a) Determine the distance traveled during the first second. (b) Determine the final velocity at which the object hits the ground. (c) Determine the distance traveled during the last second of motion before hitting the ground.

Solution:

- a. $y = 4.9$ m;
- b. $v = 38.3$ m/s;
- c. -33.3 m

Exercise:**Problem:**

A steel ball is dropped onto a hard floor from a height of 1.50 m and rebounds to a height of 1.45 m. (a) Calculate its velocity just before it strikes the floor. (b) Calculate its velocity just after it leaves the floor on its way back up. (c) Calculate its acceleration during contact with the floor if that contact lasts 0.0800 ms (8.00×10^{-5} s) (d) How much did the ball compress during its collision with the floor, assuming the floor is absolutely rigid?

Exercise:**Problem:**

An object is dropped from a roof of a building of height h . During the last second of its descent, it drops a distance $h/3$. Calculate the height of the building.

Solution:

$h = \frac{1}{2}gt^2$, h = total height and time to drop to ground

$\frac{2}{3}h = \frac{1}{2}g(t-1)^2$ in $t-1$ seconds it drops $2/3h$

$\frac{2}{3}(\frac{1}{2}gt^2) = \frac{1}{2}g(t-1)^2$ or $\frac{t^2}{3} = \frac{1}{2}(t-1)^2$

$0 = t^2 - 6t + 3$ $t = \frac{6 \pm \sqrt{6^2 - 4 \cdot 3}}{2} = 3 \pm \frac{\sqrt{24}}{2}$

$t = 5.45$ s and $h = 145.5$ m. Other root is less than 1 s. Check for $t = 4.45$ s $h = \frac{1}{2}gt^2 = 97.0$

m = $\frac{2}{3}(145.5)$

Challenge Problems**Exercise:****Problem:**

In a 100-m race, the winner is timed at 11.2 s. The second-place finisher's time is 11.6 s. How far is the second-place finisher behind the winner when she crosses the finish line? Assume the velocity of each runner is constant throughout the race.

Exercise:**Problem:**

The position of a particle moving along the x-axis varies with time according to $x(t) = 5.0t^2 - 4.0t^3$ m. Find (a) the velocity and acceleration of the particle as functions of time, (b) the velocity and acceleration at $t = 2.0$ s, (c) the time at which the position is a maximum, (d) the time at which the velocity is zero, and (e) the maximum position.

Solution:

- a. $v(t) = 10t - 12t^2$ m/s, $a(t) = 10 - 24t$ m/s²;
 b. $v(2 \text{ s}) = -28$ m/s, $a(2 \text{ s}) = -38$ m/s²; c. The slope of the position function is zero or the velocity is zero. There are two possible solutions: $t = 0$, which gives $x = 0$, or $t = 10.0/12.0 = 0.83$ s, which gives $x = 1.16$ m. The second answer is the correct choice; d. 0.83 s
 (e) 1.16 m

Exercise:**Problem:**

A cyclist sprints at the end of a race to clinch a victory. She has an initial velocity of 11.5 m/s and accelerates at a rate of 0.500 m/s² for 7.00 s. (a) What is her final velocity? (b) The cyclist continues at this velocity to the finish line. If she is 300 m from the finish line when she starts to accelerate, how much time did she save? (c) The second-place winner was 5.00 m ahead when the winner started to accelerate, but he was unable to accelerate, and traveled at 11.8 m/s until the finish line. What was the difference in finish time in seconds between the winner and runner-up? How far back was the runner-up when the winner crossed the finish line?

Exercise:**Problem:**

In 1967, New Zealander Burt Munro set the world record for an Indian motorcycle, on the Bonneville Salt Flats in Utah, of 295.38 km/h. The one-way course was 8.00 km long. Acceleration rates are often described by the time it takes to reach 96.0 km/h from rest. If this time was 4.00 s and Burt accelerated at this rate until he reached his maximum speed, how long did it take Burt to complete the course?

Solution:

$96 \text{ km/h} = 26.67 \text{ m/s}$, $a = \frac{26.67 \text{ m/s}}{4.0 \text{ s}} = 6.67 \text{ m/s}^2$, $295.38 \text{ km/h} = 82.05 \text{ m/s}$, $t = 12.3 \text{ s}$
 time to accelerate to maximum speed
 $x = 504.55 \text{ m}$ distance covered during acceleration
 7495.44 m at a constant speed
 $\frac{7495.44 \text{ m}}{82.05 \text{ m/s}} = 91.35 \text{ s}$ so total time is $91.35 \text{ s} + 12.3 \text{ s} = 103.65 \text{ s}$.

Introduction

class="introduction"

The Red
Arrows is
the
aerobatics
display team
of Britain's
Royal Air
Force. Based
in
Lincolnshire
, England,
they perform
precision
flying shows
at high
speeds,
which
requires
accurate
measuremen
t of position,
velocity, and
acceleration
in three
dimensions.

(credit:
modification
of work by
Phil Long)



To give a complete description of kinematics, we must explore motion in two and three dimensions. After all, most objects in our universe do not move in straight lines; rather, they follow curved paths. From kicked footballs to the flight paths of birds to the orbital motions of celestial bodies and down to the flow of blood plasma in your veins, most motion follows curved trajectories.

Fortunately, the treatment of motion in one dimension in the previous chapter has given us a foundation on which to build, as the concepts of position, displacement, velocity, and acceleration defined in one dimension can be expanded to two and three dimensions. Consider the Red Arrows, also known as the Royal Air Force Aerobatic team of the United Kingdom. Each jet follows a unique curved trajectory in three-dimensional airspace, as well as has a unique velocity and acceleration. Thus, to describe the motion of any of the jets accurately, we must assign to each jet a unique position vector in three dimensions as well as a unique velocity and acceleration vector. We can apply the same basic equations for displacement, velocity, and acceleration we derived in [Motion Along a Straight Line](#) to describe the motion of the jets in two and three dimensions, but with some modifications—in particular, the inclusion of vectors.

In this chapter we also explore two special types of motion in two dimensions: projectile motion and circular motion. Last, we conclude with a discussion of relative motion. In the chapter-opening picture, each jet has a

relative motion with respect to any other jet in the group or to the people observing the air show on the ground.

Displacement and Velocity Vectors

By the end of this section, you will be able to:

- Calculate position vectors in a multidimensional displacement problem.
- Solve for the displacement in two or three dimensions.
- Calculate the velocity vector given the position vector as a function of time.
- Calculate the average velocity in multiple dimensions.

Displacement and velocity in two or three dimensions are straightforward extensions of the one-dimensional definitions. However, now they are vector quantities, so calculations with them have to follow the rules of vector algebra, not scalar algebra.

Displacement Vector

To describe motion in two and three dimensions, we must first establish a coordinate system and a convention for the axes. We generally use the coordinates x , y , and z to locate a particle at point $P(x, y, z)$ in three dimensions. If the particle is moving, the variables x , y , and z are functions of time (t):

Equation:

$$x = x(t) \quad y = y(t) \quad z = z(t).$$

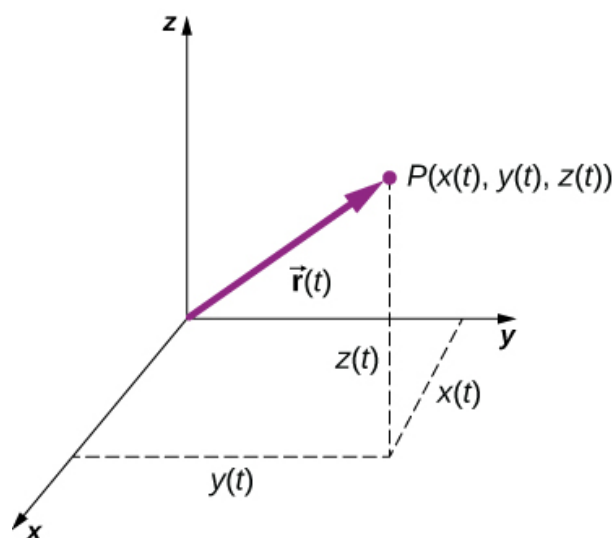
The **position vector** from the origin of the coordinate system to point P is $\vec{\mathbf{r}}(t)$. In unit vector notation, introduced in [Coordinate Systems and Components of a Vector](#), $\vec{\mathbf{r}}(t)$ is

Note:

Equation:

$$\vec{\mathbf{r}}(t) = x(t)\hat{\mathbf{i}} + y(t)\hat{\mathbf{j}} + z(t)\hat{\mathbf{k}}.$$

[\[link\]](#) shows the coordinate system and the vector to point P , where a particle could be located at a particular time t . Note the orientation of the x , y , and z axes. This orientation is called a right-handed coordinate system ([Coordinate Systems and Components of a Vector](#)) and it is used throughout the chapter.



A three-dimensional coordinate system with a particle at position $P(x(t), y(t), z(t))$.

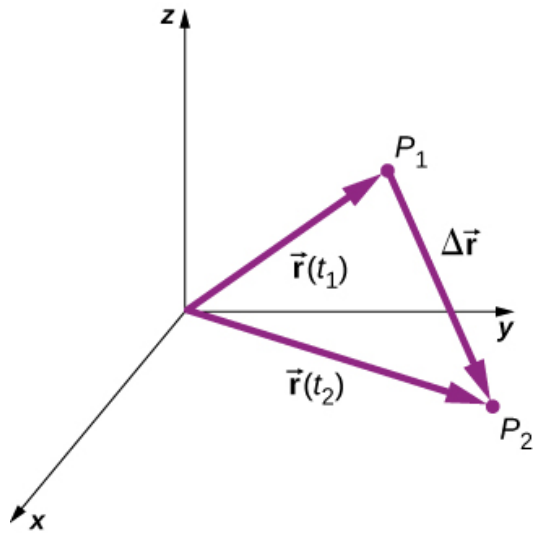
With our definition of the position of a particle in three-dimensional space, we can formulate the three-dimensional displacement. [\[link\]](#) shows a particle at time t_1 located at P_1 with position vector $\vec{r}(t_1)$. At a later time t_2 , the particle is located at P_2 with position vector $\vec{r}(t_2)$. The **displacement vector** $\Delta\vec{r}$ is found by subtracting $\vec{r}(t_1)$ from $\vec{r}(t_2)$:

Note:

Equation:

$$\Delta\vec{r} = \vec{r}(t_2) - \vec{r}(t_1).$$

Vector addition is discussed in [Vectors](#). Note that this is the same operation we did in one dimension, but now the vectors are in three-dimensional space.



The displacement
 $\Delta \vec{r} = \vec{r}(t_2) - \vec{r}(t_1)$ is the vector
 from P_1 to P_2 .

The following examples illustrate the concept of displacement in multiple dimensions.

Example:

Polar Orbiting Satellite

A satellite is in a circular polar orbit around Earth at an altitude of 400 km—meaning, it passes directly overhead at the North and South Poles. What is the magnitude and direction of the displacement vector from when it is directly over the North Pole to when it is at -45° latitude?

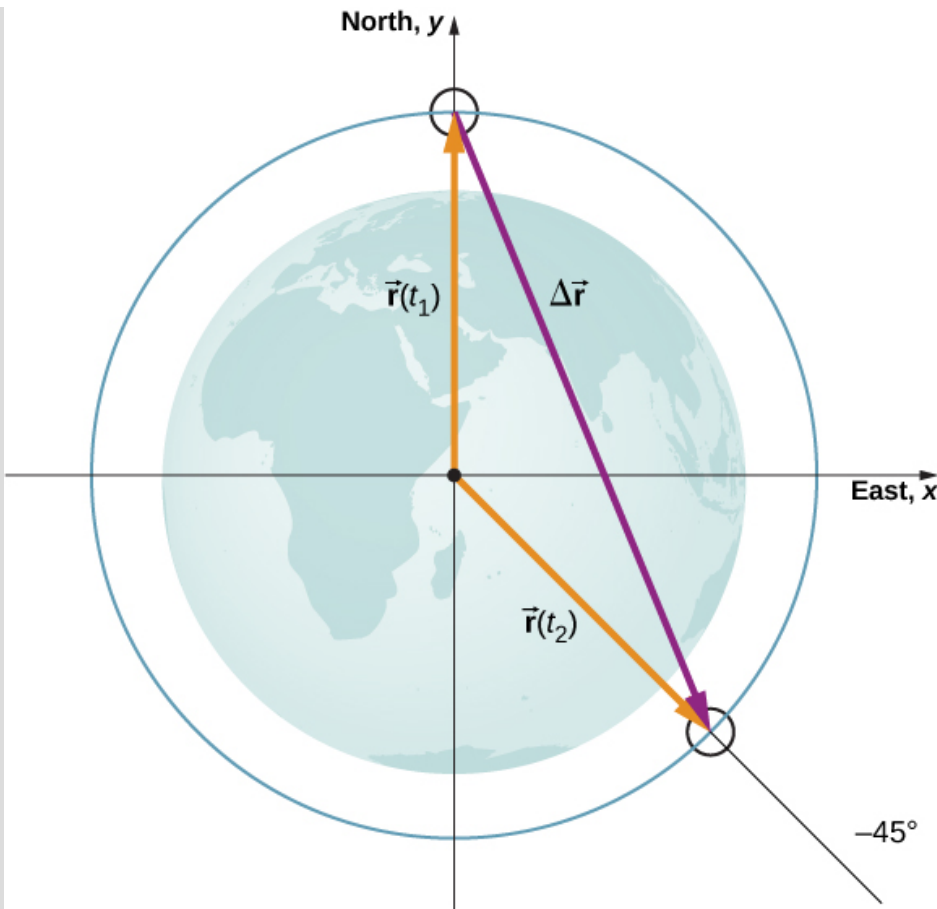
Strategy

We make a picture of the problem to visualize the solution graphically. This will aid in our understanding of the displacement. We then use unit vectors to solve for the displacement.

Solution

[\[link\]](#) shows the surface of Earth and a circle that represents the orbit of the satellite.

Although satellites are moving in three-dimensional space, they follow trajectories of ellipses, which can be graphed in two dimensions. The position vectors are drawn from the center of Earth, which we take to be the origin of the coordinate system, with the y-axis as north and the x-axis as east. The vector between them is the displacement of the satellite. We take the radius of Earth as 6370 km, so the length of each position vector is 6770 km.



Two position vectors are drawn from the center of Earth, which is the origin of the coordinate system, with the y -axis as north and the x -axis as east. The vector between them is the displacement of the satellite.

In unit vector notation, the position vectors are

Equation:

$$\vec{r}(t_1) = 6770. \text{ km} \hat{j}$$

$$\vec{r}(t_2) = 6770. \text{ km} (\cos(-45^\circ)) \hat{i} + 6770. \text{ km} (\sin(-45^\circ)) \hat{j}.$$

Evaluating the sine and cosine, we have

Equation:

$$\vec{r}(t_1) = 6770. \hat{j}$$

$$\vec{r}(t_2) = 4787 \hat{i} - 4787 \hat{j}.$$

Now we can find $\Delta\vec{r}$, the displacement of the satellite:

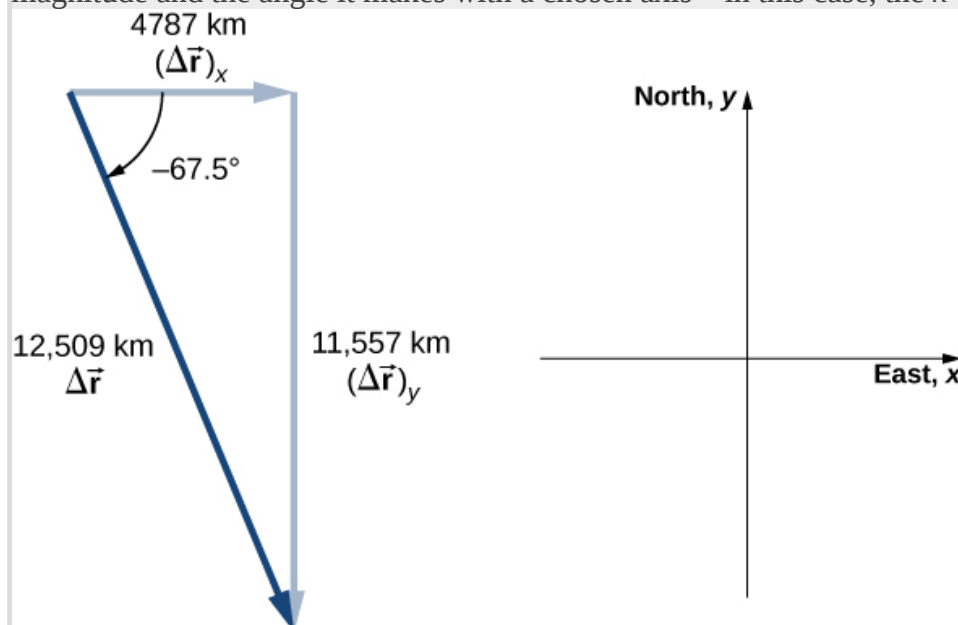
Equation:

$$\Delta\vec{r} = \vec{r}(t_2) - \vec{r}(t_1) = 4787\hat{i} - 11,557\hat{j}.$$

The magnitude of the displacement is $|\Delta\vec{r}| = \sqrt{(4787)^2 + (-11,557)^2} = 12,509 \text{ km}$. The angle the displacement makes with the x -axis is $\theta = \tan^{-1}\left(\frac{-11,557}{4787}\right) = -67.5^\circ$.

Significance

Plotting the displacement gives information and meaning to the unit vector solution to the problem. When plotting the displacement, we need to include its components as well as its magnitude and the angle it makes with a chosen axis—in this case, the x -axis ([link](#)).



Displacement vector with components, angle, and magnitude.

Note that the satellite took a curved path along its circular orbit to get from its initial position to its final position in this example. It also could have traveled 4787 km east, then 11,557 km south to arrive at the same location. Both of these paths are longer than the length of the displacement vector. In fact, the displacement vector gives the shortest path between two points in one, two, or three dimensions.

Many applications in physics can have a series of displacements, as discussed in the previous chapter. The total displacement is the sum of the individual displacements, only this time, we need to be careful, because we are adding vectors. We illustrate this concept with an example of Brownian motion.

Example:

Brownian Motion

Brownian motion is a chaotic random motion of particles suspended in a fluid, resulting from collisions with the molecules of the fluid. This motion is three-dimensional. The displacements in numerical order of a particle undergoing Brownian motion could look like the following, in micrometers ([link](#)):

Equation:

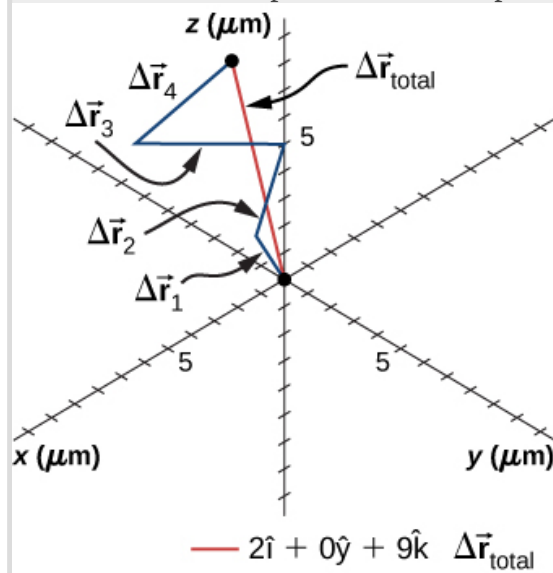
$$\Delta \vec{r}_1 = 2.0\hat{i} + \hat{j} + 3.0\hat{k}$$

$$\Delta \vec{r}_2 = -\hat{i} + 3.0\hat{k}$$

$$\Delta \vec{r}_3 = 4.0\hat{i} - 2.0\hat{j} + \hat{k}$$

$$\Delta \vec{r}_4 = -3.0\hat{i} + \hat{j} + 2.0\hat{k}.$$

What is the total displacement of the particle from the origin?



Trajectory of a particle undergoing random displacements of Brownian motion. The total displacement is shown in red.

Solution

We form the sum of the displacements and add them as vectors:

Equation:

$$\begin{aligned}
\Delta \vec{r}_{\text{Total}} &= \sum \Delta \vec{r}_i = \Delta \vec{r}_1 + \Delta \vec{r}_2 + \Delta \vec{r}_3 + \Delta \vec{r}_4 \\
&= (2.0 - 1.0 + 4.0 - 3.0)\hat{i} + (1.0 + 0 - 2.0 + 1.0)\hat{j} + (3.0 + 3.0 + 1.0 + 2.0)\hat{k} \\
&= 2.0\hat{i} + 0\hat{j} + 9.0\hat{k} \mu\text{m}.
\end{aligned}$$

To complete the solution, we express the displacement as a magnitude and direction,

Equation:

$$|\Delta \vec{r}_{\text{Total}}| = \sqrt{2.0^2 + 0^2 + 9.0^2} = 9.2 \mu\text{m}, \quad \theta = \tan^{-1} \left(\frac{9}{2} \right) = 77^\circ,$$

with respect to the x-axis in the xz-plane.

Significance

From the figure we can see the magnitude of the total displacement is less than the sum of the magnitudes of the individual displacements.

Velocity Vector

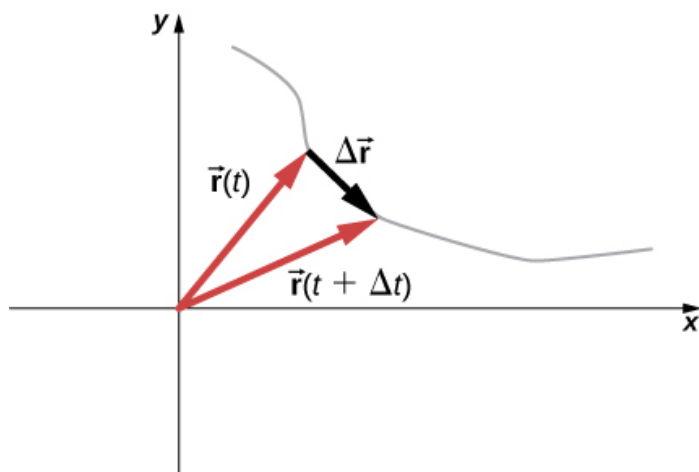
In the previous chapter we found the instantaneous velocity by calculating the derivative of the position function with respect to time. We can do the same operation in two and three dimensions, but we use vectors. The instantaneous **velocity vector** is now

Note:

Equation:

$$\vec{v}(t) = \lim_{\Delta t \rightarrow 0} \frac{\vec{r}(t + \Delta t) - \vec{r}(t)}{\Delta t} = \frac{d\vec{r}}{dt}.$$

Let's look at the relative orientation of the position vector and velocity vector graphically. In [\[link\]](#) we show the vectors $\vec{r}(t)$ and $\vec{r}(t + \Delta t)$, which give the position of a particle moving along a path represented by the gray line. As Δt goes to zero, the velocity vector, given by [\[link\]](#), becomes tangent to the path of the particle at time t .



A particle moves along a path given by the gray line. In the limit as Δt approaches zero, the velocity vector becomes tangent to the path of the particle.

[\[link\]](#) can also be written in terms of the components of $\vec{v}(t)$. Since

Equation:

$$\vec{r}(t) = x(t)\hat{\mathbf{i}} + y(t)\hat{\mathbf{j}} + z(t)\hat{\mathbf{k}},$$

we can write

Note:

Equation:

$$\vec{v}(t) = v_x(t)\hat{\mathbf{i}} + v_y(t)\hat{\mathbf{j}} + v_z(t)\hat{\mathbf{k}}$$

where

Note:

Equation:

$$v_x(t) = \frac{dx(t)}{dt}, \quad v_y(t) = \frac{dy(t)}{dt}, \quad v_z(t) = \frac{dz(t)}{dt}.$$

If only the average velocity is of concern, we have the vector equivalent of the one-dimensional average velocity for two and three dimensions:

Note:

Equation:

$$\vec{v}_{\text{avg}} = \frac{\vec{r}(t_2) - \vec{r}(t_1)}{t_2 - t_1}.$$

Example:

Calculating the Velocity Vector

The position function of a particle is $\vec{r}(t) = 2.0t^2\hat{i} + (2.0 + 3.0t)\hat{j} + 5.0t\hat{k}$ m. (a) What is the instantaneous velocity and speed at $t = 2.0$ s? (b) What is the average velocity between 1.0 s and 3.0 s?

Solution

Using [\[link\]](#) and [\[link\]](#), and taking the derivative of the position function with respect to time, we find

$$(a) \ v(t) = \frac{d\vec{r}(t)}{dt} = 4.0t\hat{i} + 3.0\hat{j} + 5.0\hat{k} \text{ m/s}$$

$$\vec{v}(2.0 \text{ s}) = 8.0\hat{i} + 3.0\hat{j} + 5.0\hat{k} \text{ m/s}$$

$$\text{Speed } |\vec{v}(2.0 \text{ s})| = \sqrt{8^2 + 3^2 + 5^2} = 9.9 \text{ m/s}.$$

(b) From [\[link\]](#),

$$\begin{aligned} \vec{v}_{\text{avg}} &= \frac{\vec{r}(t_2) - \vec{r}(t_1)}{t_2 - t_1} = \frac{\vec{r}(3.0 \text{ s}) - \vec{r}(1.0 \text{ s})}{3.0 \text{ s} - 1.0 \text{ s}} = \frac{(18\hat{i} + 11\hat{j} + 15\hat{k}) \text{ m} - (2\hat{i} + 5\hat{j} + 5\hat{k}) \text{ m}}{2.0 \text{ s}} \\ &= \frac{(16\hat{i} + 6\hat{j} + 10\hat{k}) \text{ m}}{2.0 \text{ s}} = 8.0\hat{i} + 3.0\hat{j} + 5.0\hat{k} \text{ m/s}. \end{aligned}$$

Significance

We see the average velocity is the same as the instantaneous velocity at $t = 2.0$ s, as a result of the velocity function being linear. This need not be the case in general. In fact, most of the time, instantaneous and average velocities are not the same.

Note:

Exercise:

Problem:

Check Your Understanding The position function of a particle is

$\vec{r}(t) = 3.0t^3\hat{i} + 4.0\hat{j}$. (a) What is the instantaneous velocity at $t = 3$ s? (b) Is the average velocity between 2 s and 4 s equal to the instantaneous velocity at $t = 3$ s?

Solution:

(a) Taking the derivative with respect to time of the position function, we have

$\vec{v}(t) = 9.0t^2\hat{i}$ and $\vec{v}(3.0\text{ s}) = 81.0\hat{i}\text{ m/s}$. (b) Since the velocity function is nonlinear, we suspect the average velocity is not equal to the instantaneous velocity. We check this and find

$$\vec{v}_{\text{avg}} = \frac{\vec{r}(t_2) - \vec{r}(t_1)}{t_2 - t_1} = \frac{\vec{r}(4.0\text{ s}) - \vec{r}(2.0\text{ s})}{4.0\text{ s} - 2.0\text{ s}} = \frac{(188\hat{i} - 20\hat{i})\text{ m}}{2.0\text{ s}} = 84\hat{i}\text{ m/s},$$

which is different from $\vec{v}(3.0\text{ s}) = 81.0\hat{i}\text{ m/s}$.

The Independence of Perpendicular Motions

When we look at the three-dimensional equations for position and velocity written in unit vector notation, [\[link\]](#) and [\[link\]](#), we see the components of these equations are separate and unique functions of time that do not depend on one another. Motion along the x direction has no part of its motion along the y and z directions, and similarly for the other two coordinate axes. Thus, the motion of an object in two or three dimensions can be divided into separate, independent motions along the perpendicular axes of the coordinate system in which the motion takes place.

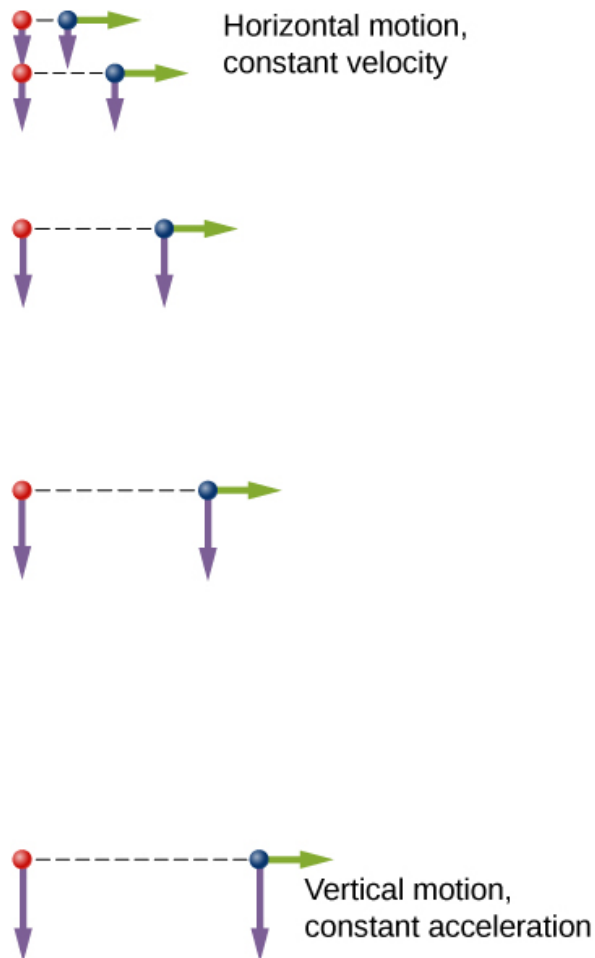
To illustrate this concept with respect to displacement, consider a woman walking from point A to point B in a city with square blocks. The woman taking the path from A to B may walk east for so many blocks and then north (two perpendicular directions) for another set of blocks to arrive at B . How far she walks east is affected only by her motion eastward. Similarly, how far she walks north is affected only by her motion northward.

Note:**Independence of Motion**

In the kinematic description of motion, we are able to treat the horizontal and vertical components of motion separately. In many cases, motion in the horizontal direction does not affect motion in the vertical direction, and vice versa.

An example illustrating the independence of vertical and horizontal motions is given by two baseballs. One baseball is dropped from rest. At the same instant, another is thrown

horizontally from the same height and it follows a curved path. A stroboscope captures the positions of the balls at fixed time intervals as they fall ([link](#)).



A diagram of the motions of two identical balls: one falls from rest and the other has an initial horizontal velocity. Each subsequent position is an equal time interval. Arrows represent the horizontal and vertical velocities at each position. The ball on the right has an initial horizontal velocity whereas the ball on the left has no horizontal velocity. Despite the difference in horizontal velocities, the vertical velocities and positions are identical for both balls, which shows the vertical and horizontal motions are independent.

It is remarkable that for each flash of the strobe, the vertical positions of the two balls are the same. This similarity implies vertical motion is independent of whether the ball is moving horizontally. (Assuming no air resistance, the vertical motion of a falling object is influenced by gravity only, not by any horizontal forces.) Careful examination of the ball thrown horizontally shows it travels the same horizontal distance between flashes. This is because there are no additional forces on the ball in the horizontal direction after it is thrown. This result means horizontal velocity is constant and is affected neither by vertical motion nor by gravity (which is vertical). Note this case is true for ideal conditions only. In the real world, air resistance affects the speed of the balls in both directions.

The two-dimensional curved path of the horizontally thrown ball is composed of two independent one-dimensional motions (horizontal and vertical). The key to analyzing such motion, called *projectile motion*, is to resolve it into motions along perpendicular directions. Resolving two-dimensional motion into perpendicular components is possible because the components are independent.

Summary

- The position function $\vec{\mathbf{r}}(t)$ gives the position as a function of time of a particle moving in two or three dimensions. Graphically, it is a vector from the origin of a chosen coordinate system to the point where the particle is located at a specific time.
- The displacement vector $\Delta\vec{\mathbf{r}}$ gives the shortest distance between any two points on the trajectory of a particle in two or three dimensions.
- Instantaneous velocity gives the speed and direction of a particle at a specific time on its trajectory in two or three dimensions, and is a vector in two and three dimensions.
- The velocity vector is tangent to the trajectory of the particle.
- Displacement $\vec{\mathbf{r}}(t)$ can be written as a vector sum of the one-dimensional displacements $\vec{x}(t)$, $\vec{y}(t)$, $\vec{z}(t)$ along the x , y , and z directions.
- Velocity $\vec{\mathbf{v}}(t)$ can be written as a vector sum of the one-dimensional velocities $v_x(t)$, $v_y(t)$, $v_z(t)$ along the x , y , and z directions.
- Motion in any given direction is independent of motion in a perpendicular direction.

Conceptual Questions

Exercise:

Problem:

What form does the trajectory of a particle have if the distance from any point A to point B is equal to the magnitude of the displacement from A to B ?

Solution:

straight line

Exercise:

Problem:

Give an example of a trajectory in two or three dimensions caused by independent perpendicular motions.

Exercise:

Problem:

If the instantaneous velocity is zero, what can be said about the slope of the position function?

Solution:

The slope must be zero because the velocity vector is tangent to the graph of the position function.

Problems

Exercise:

Problem:

The coordinates of a particle in a rectangular coordinate system are $(1.0, -4.0, 6.0)$. What is the position vector of the particle?

Solution:

$$\vec{r} = 1.0\hat{i} - 4.0\hat{j} + 6.0\hat{k}$$

Exercise:

Problem:

The position of a particle changes from $\vec{r}_1 = (2.0\hat{i} + 3.0\hat{j})\text{cm}$ to $\vec{r}_2 = (-4.0\hat{i} + 3.0\hat{j})\text{cm}$. What is the particle's displacement?

Exercise:

Problem:

The 18th hole at Pebble Beach Golf Course is a dogleg to the left of length 496.0 m. The fairway off the tee is taken to be the x direction. A golfer hits his tee shot a distance of 300.0 m, corresponding to a displacement $\Delta\vec{r}_1 = 300.0\text{ m}\hat{i}$, and hits his second shot 189.0 m with a displacement $\Delta\vec{r}_2 = 172.0\text{ m}\hat{i} + 80.3\text{ m}\hat{j}$. What is the final displacement of the golf ball from the tee?

Solution:

$$\Delta \vec{r}_{\text{Total}} = 472.0 \text{ m} \hat{i} + 80.3 \text{ m} \hat{j}$$

Exercise:**Problem:**

A bird flies straight northeast a distance of 95.0 km for 3.0 h. With the x -axis due east and the y -axis due north, what is the displacement in unit vector notation for the bird? What is the average velocity for the trip?

Exercise:**Problem:**

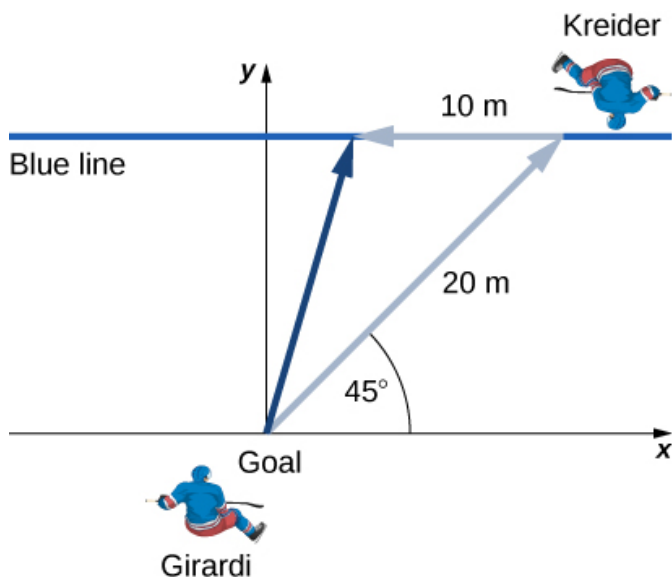
A cyclist rides 5.0 km due east, then 10.0 km 20° west of north. From this point she rides 8.0 km due west. What is the final displacement from where the cyclist started?

Solution:

$$\text{Sum of displacements} = -6.4 \text{ km} \hat{i} + 9.4 \text{ km} \hat{j}$$

Exercise:**Problem:**

New York Rangers defenseman Daniel Girardi stands at the goal and passes a hockey puck 20 m and 45° from straight down the ice to left wing Chris Kreider waiting at the blue line. Kreider waits for Girardi to reach the blue line and passes the puck directly across the ice to him 10 m away. What is the final displacement of the puck? See the following figure.



Exercise:**Problem:**

The position of a particle is $\vec{r}(t) = 4.0t^2\hat{i} - 3.0\hat{j} + 2.0t^3\hat{k}$ m. (a) What is the velocity of the particle at 0 s and at 1.0 s? (b) What is the average velocity between 0 s and 1.0 s?

Solution:

a. $\vec{v}(t) = 8.0t\hat{i} + 6.0t^2\hat{k}$, $\vec{v}(0) = 0$, $\vec{v}(1.0) = 8.0\hat{i} + 6.0\hat{k}$ m/s,

b. $\vec{v}_{\text{avg}} = 4.0\hat{i} + 2.0\hat{k}$ m/s

Exercise:**Problem:**

Clay Matthews, a linebacker for the Green Bay Packers, can reach a speed of 10.0 m/s. At the start of a play, Matthews runs downfield at 45° with respect to the 50-yard line and covers 8.0 m in 1 s. He then runs straight down the field at 90° with respect to the 50-yard line for 12 m, with an elapsed time of 1.2 s. (a) What is Matthews' final displacement from the start of the play? (b) What is his average velocity?

Exercise:**Problem:**

The F-35B Lighting II is a short-takeoff and vertical landing fighter jet. If it does a vertical takeoff to 20.00-m height above the ground and then follows a flight path angled at 30° with respect to the ground for 20.00 km, what is the final displacement?

Solution:

$$\Delta\vec{r}_1 = 20.00\text{ m}\hat{j}, \Delta\vec{r}_2 = (2.000 \times 10^4\text{ m})(\cos 30^\circ\hat{i} + \sin 30^\circ\hat{j})$$

$$\Delta\vec{r} = 1.700 \times 10^4\text{ m}\hat{i} + 1.002 \times 10^4\text{ m}\hat{j}$$

Glossary

displacement vector

vector from the initial position to a final position on a trajectory of a particle

position vector

vector from the origin of a chosen coordinate system to the position of a particle in two- or three-dimensional space

velocity vector

vector that gives the instantaneous speed and direction of a particle; tangent to the trajectory

Acceleration Vector

By the end of this section, you will be able to:

- Calculate the acceleration vector given the velocity function in unit vector notation.
- Describe the motion of a particle with a constant acceleration in three dimensions.
- Use the one-dimensional motion equations along perpendicular axes to solve a problem in two or three dimensions with a constant acceleration.
- Express the acceleration in unit vector notation.

Instantaneous Acceleration

In addition to obtaining the displacement and velocity vectors of an object in motion, we often want to know its **acceleration vector** at any point in time along its trajectory. This acceleration vector is the instantaneous acceleration and it can be obtained from the derivative with respect to time of the velocity function, as we have seen in a previous chapter. The only difference in two or three dimensions is that these are now vector quantities. Taking the derivative with respect to time $\vec{v}(t)$, we find

Note:

Equation:

$$\vec{a}(t) = \lim_{t \rightarrow 0} \frac{\vec{v}(t + \Delta t) - \vec{v}(t)}{\Delta t} = \frac{d\vec{v}(t)}{dt}.$$

The acceleration in terms of components is

Note:

Equation:

$$\vec{a}(t) = \frac{dv_x(t)}{dt} \hat{\mathbf{i}} + \frac{dv_y(t)}{dt} \hat{\mathbf{j}} + \frac{dv_z(t)}{dt} \hat{\mathbf{k}}.$$

Also, since the velocity is the derivative of the position function, we can write the acceleration in terms of the second derivative of the position function:

Note:

Equation:

$$\vec{\mathbf{a}}(t) = \frac{d^2x(t)}{dt^2}\hat{\mathbf{i}} + \frac{d^2y(t)}{dt^2}\hat{\mathbf{j}} + \frac{d^2z(t)}{dt^2}\hat{\mathbf{k}}.$$

Example:

Finding an Acceleration Vector

A particle has a velocity of $\vec{\mathbf{v}}(t) = 5.0t\hat{\mathbf{i}} + t^2\hat{\mathbf{j}} - 2.0t^3\hat{\mathbf{k}}$ m/s. (a) What is the acceleration function? (b) What is the acceleration vector at $t = 2.0$ s? Find its magnitude and direction.

Solution

(a) We take the first derivative with respect to time of the velocity function to find the acceleration. The derivative is taken component by component:

Equation:

$$\vec{\mathbf{a}}(t) = 5.0\hat{\mathbf{i}} + 2.0t\hat{\mathbf{j}} - 6.0t^2\hat{\mathbf{k}} \text{ m/s}^2.$$

(b) Evaluating $\vec{\mathbf{a}}(2.0 \text{ s}) = 5.0\hat{\mathbf{i}} + 4.0\hat{\mathbf{j}} - 24.0\hat{\mathbf{k}}$ m/s² gives us the direction in unit vector notation. The magnitude of the acceleration is

$$|\vec{\mathbf{a}}(2.0 \text{ s})| = \sqrt{5.0^2 + 4.0^2 + (-24.0)^2} = 24.8 \text{ m/s}^2.$$

Significance

In this example we find that acceleration has a time dependence and is changing throughout the motion. Let's consider a different velocity function for the particle.

Example:

Finding a Particle Acceleration

A particle has a position function $\vec{\mathbf{r}}(t) = (10t - t^2)\hat{\mathbf{i}} + 5t\hat{\mathbf{j}} + 5t\hat{\mathbf{k}}$ m. (a) What is the velocity? (b) What is the acceleration? (c) Describe the motion from $t = 0$ s.

Strategy

We can gain some insight into the problem by looking at the position function. It is linear in y and z , so we know the acceleration in these directions is zero when we take

the second derivative. Also, note that the position in the x direction is zero for $t = 0$ s and $t = 10$ s.

Solution

(a) Taking the derivative with respect to time of the position function, we find

Equation:

$$\vec{v}(t) = (10 - 2t)\hat{i} + 5\hat{j} + 5\hat{k} \text{ m/s}.$$

The velocity function is linear in time in the x direction and is constant in the y and z directions.

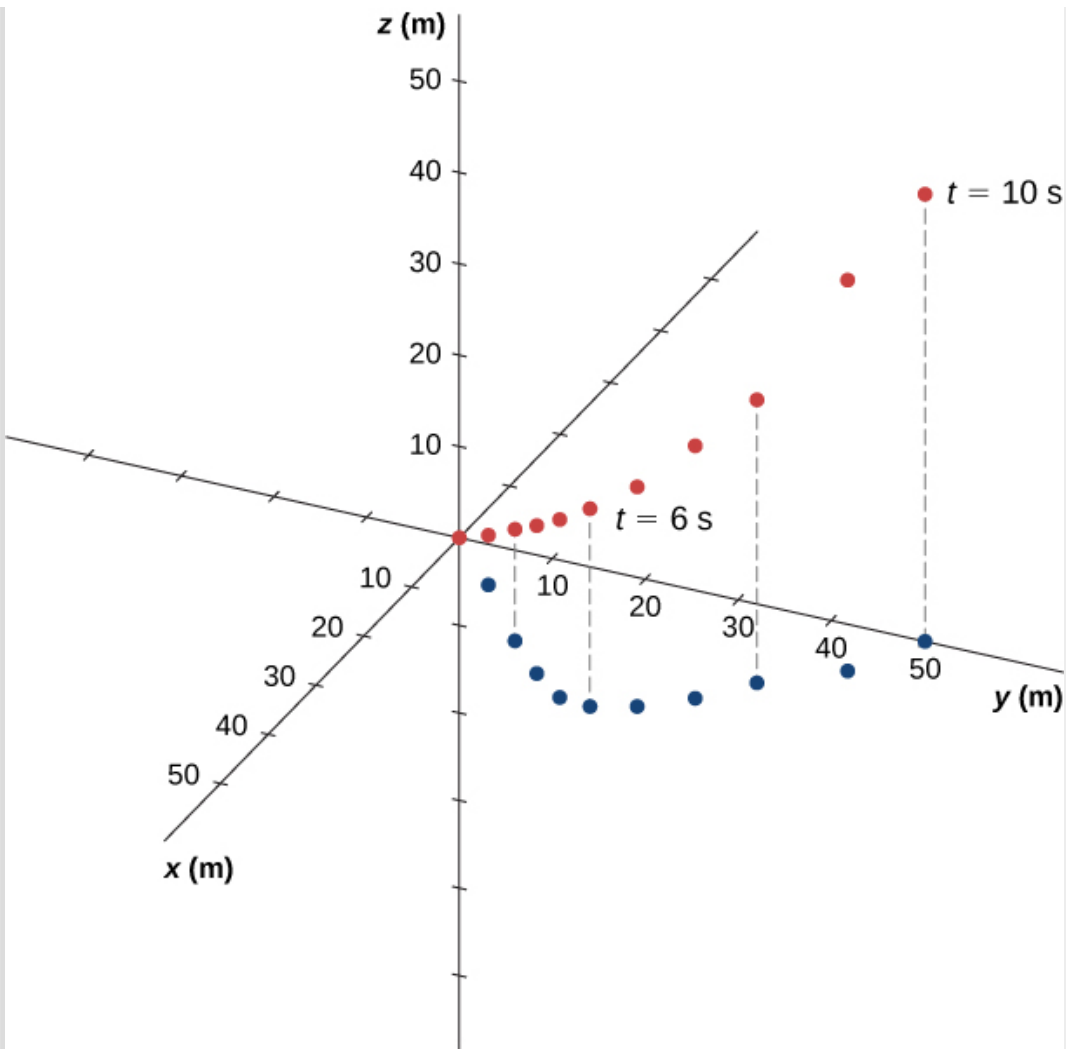
(b) Taking the derivative of the velocity function, we find

Equation:

$$\vec{a}(t) = -2\hat{i} \text{ m/s}^2.$$

The acceleration vector is a constant in the negative x -direction.

(c) The trajectory of the particle can be seen in [\[link\]](#). Let's look in the y and z directions first. The particle's position increases steadily as a function of time with a constant velocity in these directions. In the x direction, however, the particle follows a path in positive x until $t = 5$ s, when it reverses direction. We know this from looking at the velocity function, which becomes zero at this time and negative thereafter. We also know this because the acceleration is negative and constant—meaning, the particle is accelerating in the opposite direction. The particle's position reaches 25 m, where it then reverses direction and begins to accelerate in the negative x direction. The position reaches zero at $t = 10$ s.



The particle starts at point $(x, y, z) = (0, 0, 0)$ with position vector $\vec{r} = 0$. The projection of the trajectory onto the xy -plane is shown. The values of y and z increase linearly as a function of time, whereas x has a turning point at $t = 5$ s and 25 m, when it reverses direction. At this point, the x component of the velocity becomes negative. At $t = 10$ s, the particle is back to 0 m in the x direction.

Significance

By graphing the trajectory of the particle, we can better understand its motion, given by the numerical results of the kinematic equations.

Note:

Exercise:**Problem:**

Check Your Understanding Suppose the acceleration function has the form $\vec{a}(t) = a\hat{i} + b\hat{j} + c\hat{k} \text{ m/s}^2$, where a , b , and c are constants. What can be said about the functional form of the velocity function?

Solution:

The acceleration vector is constant and doesn't change with time. If a , b , and c are not zero, then the velocity function must be linear in time. We have

$\vec{v}(t) = \int \vec{a} dt = \int (a\hat{i} + b\hat{j} + c\hat{k}) dt = (a\hat{i} + b\hat{j} + c\hat{k})t \text{ m/s}$, since taking the derivative of the velocity function produces $\vec{a}(t)$. If any of the components of the acceleration are zero, then that component of the velocity would be a constant.

Constant Acceleration

Multidimensional motion with constant acceleration can be treated the same way as shown in the previous chapter for one-dimensional motion. Earlier we showed that three-dimensional motion is equivalent to three one-dimensional motions, each along an axis perpendicular to the others. To develop the relevant equations in each direction, let's consider the two-dimensional problem of a particle moving in the xy plane with constant acceleration, ignoring the z -component for the moment. The acceleration vector is

Equation:

$$\vec{a} = a_{0x}\hat{i} + a_{0y}\hat{j}.$$

Each component of the motion has a separate set of equations similar to [\[link\]](#)–[\[link\]](#) of the previous chapter on one-dimensional motion. We show only the equations for position and velocity in the x - and y -directions. A similar set of kinematic equations could be written for motion in the z -direction:

Equation:

$$x(t) = x_0 + (v_x)_{\text{avg}}t$$

Equation:

$$v_x(t) = v_{0x} + a_x t$$

Equation:

$$x(t) = x_0 + v_{0x}t + \frac{1}{2}a_x t^2$$

Equation:

$$v_x^2(t) = v_{0x}^2 + 2a_x(x - x_0)$$

Equation:

$$y(t) = y_0 + (v_y)_{\text{avg}}t$$

Equation:

$$v_y(t) = v_{0y} + a_y t$$

Equation:

$$y(t) = y_0 + v_{0y}t + \frac{1}{2}a_y t^2$$

Equation:

$$v_y^2(t) = v_{0y}^2 + 2a_y(y - y_0).$$

Here the subscript 0 denotes the initial position or velocity. [\[link\]](#) to [\[link\]](#) can be substituted into [\[link\]](#) and [\[link\]](#) without the z-component to obtain the position vector and velocity vector as a function of time in two dimensions:

Equation:

$$\vec{\mathbf{r}}(t) = x(t)\hat{\mathbf{i}} + y(t)\hat{\mathbf{j}} \text{ and } \vec{\mathbf{v}}(t) = v_x(t)\hat{\mathbf{i}} + v_y(t)\hat{\mathbf{j}}.$$

The following example illustrates a practical use of the kinematic equations in two dimensions.

Example:
A Skier

[\[link\]](#) shows a skier moving with an acceleration of 2.1 m/s^2 down a slope of 15° at $t = 0$. With the origin of the coordinate system at the front of the lodge, her initial position and velocity are

Equation:

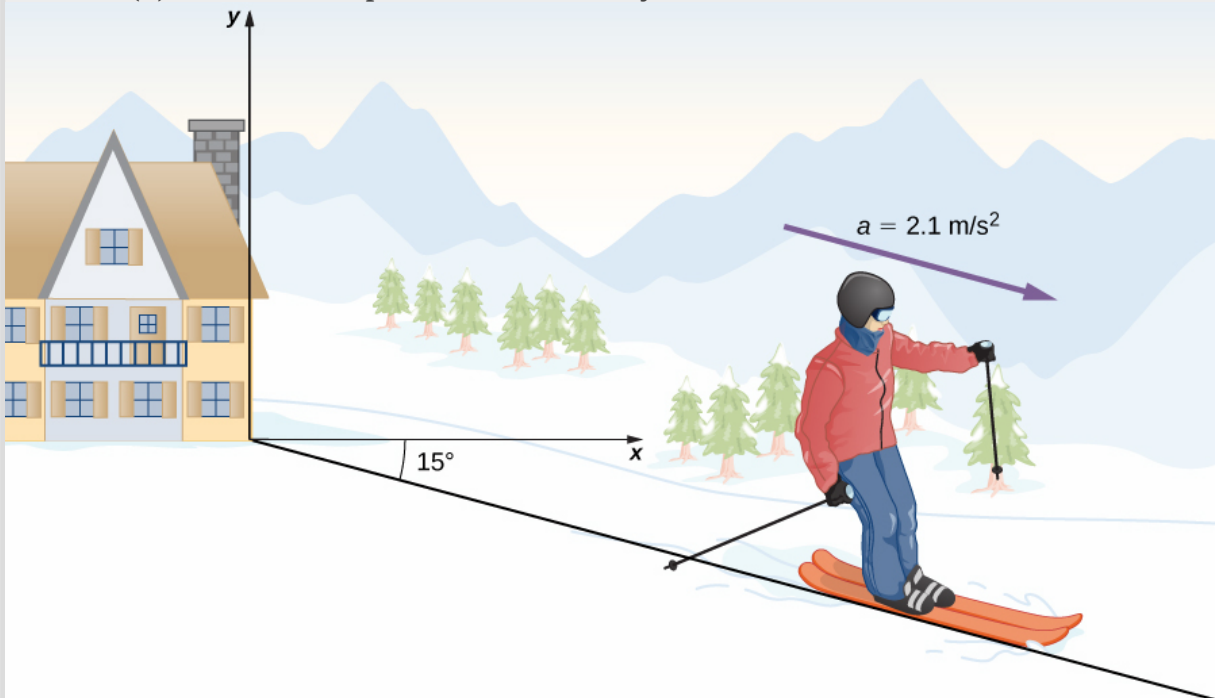
$$\vec{r}(0) = (75.0\hat{i} - 50.0\hat{j}) \text{ m}$$

and

Equation:

$$\vec{v}(0) = (4.1\hat{i} - 1.1\hat{j}) \text{ m/s}.$$

(a) What are the x - and y -components of the skier's position and velocity as functions of time? (b) What are her position and velocity at $t = 10.0 \text{ s}$?



A skier has an acceleration of 2.1 m/s^2 down a slope of 15° . The origin of the coordinate system is at the ski lodge.

Strategy

Since we are evaluating the components of the motion equations in the x and y directions, we need to find the components of the acceleration and put them into the kinematic equations. The components of the acceleration are found by referring to the coordinate system in [\[link\]](#). Then, by inserting the components of the initial position

and velocity into the motion equations, we can solve for her position and velocity at a later time t .

Solution

(a) The origin of the coordinate system is at the top of the hill with y -axis vertically upward and the x -axis horizontal. By looking at the trajectory of the skier, the x -component of the acceleration is positive and the y -component is negative. Since the angle is 15° down the slope, we find

Equation:

$$a_x = (2.1 \text{ m/s}^2) \cos(15^\circ) = 2.0 \text{ m/s}^2$$

Equation:

$$a_y = (-2.1 \text{ m/s}^2) \sin 15^\circ = -0.54 \text{ m/s}^2.$$

Inserting the initial position and velocity into [\[link\]](#) and [\[link\]](#) for x , we have

Equation:

$$x(t) = 75.0 \text{ m} + (4.1 \text{ m/s})t + \frac{1}{2}(2.0 \text{ m/s}^2)t^2$$

Equation:

$$v_x(t) = 4.1 \text{ m/s} + (2.0 \text{ m/s}^2)t.$$

For y , we have

Equation:

$$y(t) = -50.0 \text{ m} + (-1.1 \text{ m/s})t + \frac{1}{2}(-0.54 \text{ m/s}^2)t^2$$

Equation:

$$v_y(t) = -1.1 \text{ m/s} + (-0.54 \text{ m/s}^2)t.$$

(b) Now that we have the equations of motion for x and y as functions of time, we can evaluate them at $t = 10.0 \text{ s}$:

Equation:

$$x(10.0 \text{ s}) = 75.0 \text{ m} + (4.1 \text{ m/s}^2)(10.0 \text{ s}) + \frac{1}{2}(2.0 \text{ m/s}^2)(10.0 \text{ s})^2 = 216.0 \text{ m}$$

Equation:

$$v_x(10.0 \text{ s}) = 4.1 \text{ m/s} + (2.0 \text{ m/s}^2)(10.0 \text{ s}) = 24.1 \text{ m/s}$$

Equation:

$$y(10.0 \text{ s}) = -50.0 \text{ m} + (-1.1 \text{ m/s})(10.0 \text{ s}) + \frac{1}{2}(-0.54 \text{ m/s}^2)(10.0 \text{ s})^2 = -88.0 \text{ m}$$

Equation:

$$v_y(10.0 \text{ s}) = -1.1 \text{ m/s} + (-0.54 \text{ m/s}^2)(10.0 \text{ s}) = -6.5 \text{ m/s}.$$

The position and velocity at $t = 10.0 \text{ s}$ are, finally,

Equation:

$$\vec{r}(10.0 \text{ s}) = (216.0\hat{i} - 88.0\hat{j}) \text{ m}$$

Equation:

$$\vec{v}(10.0 \text{ s}) = (24.1\hat{i} - 6.5\hat{j}) \text{ m/s}.$$

The magnitude of the velocity of the skier at 10.0 s is 25 m/s , which is 60 mi/h .

Significance

It is useful to know that, given the initial conditions of position, velocity, and acceleration of an object, we can find the position, velocity, and acceleration at any later time.

With [\[link\]](#) through [\[link\]](#) we have completed the set of expressions for the position, velocity, and acceleration of an object moving in two or three dimensions. If the trajectories of the objects look something like the “Red Arrows” in the opening picture for the chapter, then the expressions for the position, velocity, and acceleration can be quite complicated. In the sections to follow we examine two special cases of motion in two and three dimensions by looking at projectile motion and circular motion.

Note:

At this [University of Colorado Boulder website](#), you can explore the position velocity and acceleration of a ladybug with an interactive simulation that allows you to change these parameters.

Summary

- In two and three dimensions, the acceleration vector can have an arbitrary direction and does not necessarily point along a given component of the velocity.
- The instantaneous acceleration is produced by a change in velocity taken over a very short (infinitesimal) time period. Instantaneous acceleration is a vector in two or three dimensions. It is found by taking the derivative of the velocity function with respect to time.
- In three dimensions, acceleration $\vec{a}(t)$ can be written as a vector sum of the one-dimensional accelerations $a_x(t)$, $a_y(t)$, and $a_z(t)$ along the x -, y -, and z -axes.
- The kinematic equations for constant acceleration can be written as the vector sum of the constant acceleration equations in the x , y , and z directions.

Conceptual Questions

Exercise:

Problem:

If the position function of a particle is a linear function of time, what can be said about its acceleration?

Exercise:

Problem:

If an object has a constant x -component of the velocity and suddenly experiences an acceleration in the y direction, does the x -component of its velocity change?

Solution:

No, motions in perpendicular directions are independent.

Exercise:

Problem:

If an object has a constant x -component of velocity and suddenly experiences an acceleration at an angle of 70° in the x direction, does the x -component of velocity change?

Problems

Exercise:

Problem:

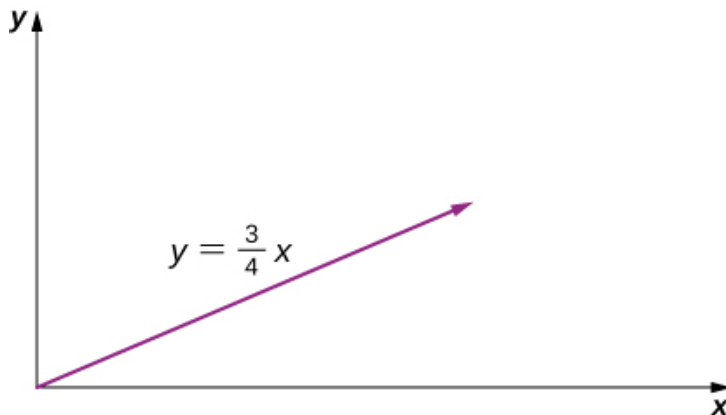
The position of a particle is $\vec{r}(t) = (3.0t^2\hat{i} + 5.0\hat{j} - 6.0t\hat{k})$ m. (a) Determine its velocity and acceleration as functions of time. (b) What are its velocity and acceleration at time $t = 0$?

Exercise:**Problem:**

A particle's acceleration is $(4.0\hat{i} + 3.0\hat{j})\text{m/s}^2$. At $t = 0$, its position and velocity are zero. (a) What are the particle's position and velocity as functions of time? (b) Find the equation of the path of the particle. Draw the x - and y -axes and sketch the trajectory of the particle.

Solution:

a. $\vec{v}(t) = (4.0t\hat{i} + 3.0t\hat{j})\text{m/s}$, $\vec{r}(t) = (2.0t^2\hat{i} + \frac{3}{2}t^2\hat{j})$ m,
b. $x(t) = 2.0t^2\text{m}$, $y(t) = \frac{3}{2}t^2\text{m}$, $t^2 = \frac{x}{2} \Rightarrow y = \frac{3}{4}x$

**Exercise:****Problem:**

A boat leaves the dock at $t = 0$ and heads out into a lake with an acceleration of $2.0\text{ m/s}^2\hat{i}$. A strong wind is pushing the boat, giving it an additional velocity of $2.0\text{ m/s}\hat{i} + 1.0\text{ m/s}\hat{j}$. (a) What is the velocity of the boat at $t = 10\text{ s}$? (b) What is the position of the boat at $t = 10\text{ s}$? Draw a sketch of the boat's trajectory and position at $t = 10\text{ s}$, showing the x - and y -axes.

Exercise:

Problem:

The position of a particle for $t > 0$ is given by

$\vec{r}(t) = (3.0t^2\hat{i} - 7.0t^3\hat{j} - 5.0t^{-2}\hat{k})$ m. (a) What is the velocity as a function of time? (b) What is the acceleration as a function of time? (c) What is the particle's velocity at $t = 2.0$ s? (d) What is its speed at $t = 1.0$ s and $t = 3.0$ s? (e) What is the average velocity between $t = 1.0$ s and $t = 2.0$ s?

Solution:

a. $\vec{v}(t) = (6.0t\hat{i} - 21.0t^2\hat{j} + 10.0t^{-3}\hat{k})$ m/s,

b. $\vec{a}(t) = (6.0\hat{i} - 42.0t\hat{j} - 30t^{-4}\hat{k})$ m/s²,

c. $\vec{v}(2.0\text{ s}) = (12.0\hat{i} - 84.0\hat{j} + 1.25\hat{k})$ m/s,

d. $\vec{v}(1.0\text{ s}) = 6.0\hat{i} - 21.0\hat{j} + 10.0\hat{k}$ m/s, $|\vec{v}(1.0\text{ s})| = 24.0$ m/s

$\vec{v}(3.0\text{ s}) = 18.0\hat{i} - 189.0\hat{j} + 0.37\hat{k}$ m/s, $|\vec{v}(3.0\text{ s})| = 190$ m/s,

e. $\vec{r}(t) = (3.0t^2\hat{i} - 7.0t^3\hat{j} - 5.0t^{-2}\hat{k})$ m

$$\vec{v}_{\text{avg}} = 9.0\hat{i} - 49.0\hat{j} + 3.75\hat{k} \text{ m/s}$$

Exercise:**Problem:**

The acceleration of a particle is a constant. At $t = 0$ the velocity of the particle is $(10\hat{i} + 20\hat{j})$ m/s. At $t = 4$ s the velocity is $10\hat{j}$ m/s. (a) What is the particle's acceleration? (b) How do the position and velocity vary with time? Assume the particle is initially at the origin.

Exercise:**Problem:**

A particle has a position function $\vec{r}(t) = \cos(1.0t)\hat{i} + \sin(1.0t)\hat{j} + t\hat{k}$, where the arguments of the cosine and sine functions are in radians. (a) What is the velocity vector? (b) What is the acceleration vector?

Solution:

a. $\vec{v}(t) = -\sin(1.0t)\hat{i} + \cos(1.0t)\hat{j} + \hat{k}$, b. $\vec{a}(t) = -\cos(1.0t)\hat{i} - \sin(1.0t)\hat{j}$

Exercise:

Problem:

A Lockheed Martin F-35 II Lighting jet takes off from an aircraft carrier with a runway length of 90 m and a takeoff speed 70 m/s at the end of the runway. Jets are catapulted into airspace from the deck of an aircraft carrier with two sources of propulsion: the jet propulsion and the catapult. At the point of leaving the deck of the aircraft carrier, the F-35's acceleration decreases to a constant acceleration of 5.0 m/s^2 at 30° with respect to the horizontal. (a) What is the initial acceleration of the F-35 on the deck of the aircraft carrier to make it airborne? (b) Write the position and velocity of the F-35 in unit vector notation from the point it leaves the deck of the aircraft carrier. (c) At what altitude is the fighter 5.0 s after it leaves the deck of the aircraft carrier? (d) What is its velocity and speed at this time? (e) How far has it traveled horizontally?

Glossary

acceleration vector

instantaneous acceleration found by taking the derivative of the velocity function with respect to time in unit vector notation

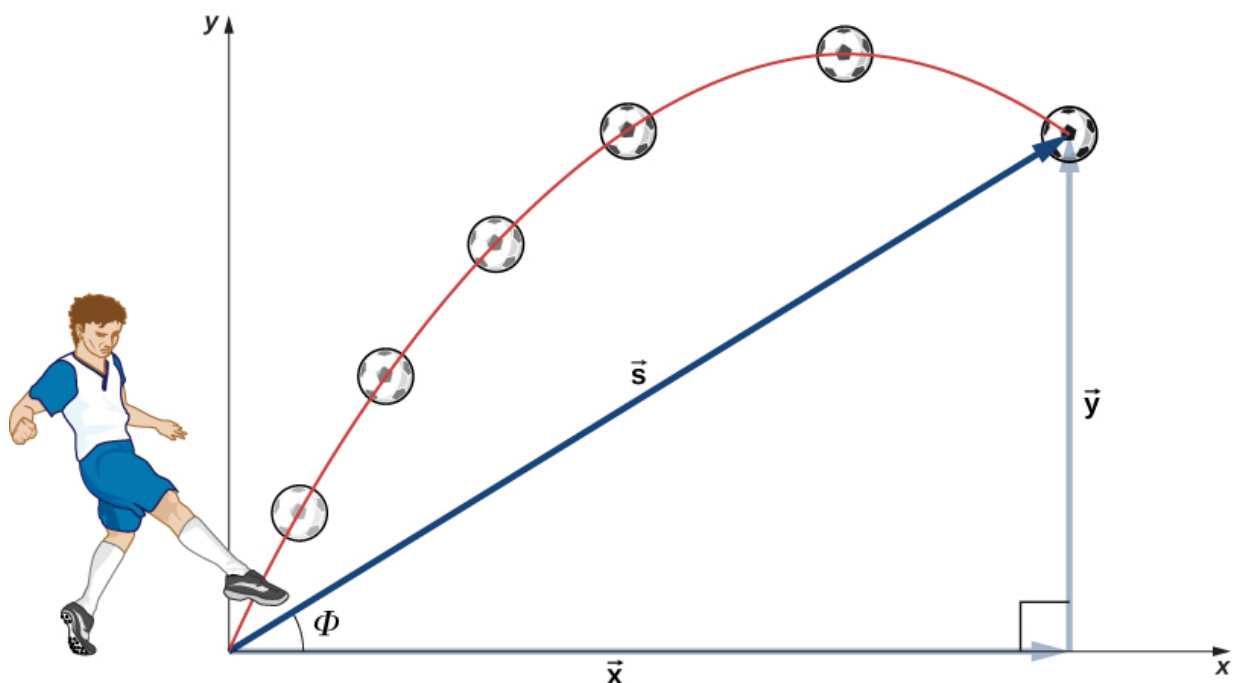
Projectile Motion

By the end of this section, you will be able to:

- Use one-dimensional motion in perpendicular directions to analyze projectile motion.
- Calculate the range, time of flight, and maximum height of a projectile that is launched and impacts a flat, horizontal surface.
- Find the time of flight and impact velocity of a projectile that lands at a different height from that of launch.
- Calculate the trajectory of a projectile.

Projectile motion is the motion of an object thrown or projected into the air, subject only to acceleration as a result of gravity. The applications of projectile motion in physics and engineering are numerous. Some examples include meteors as they enter Earth's atmosphere, fireworks, and the motion of any ball in sports. Such objects are called *projectiles* and their path is called a **trajectory**. The motion of falling objects as discussed in [Motion Along a Straight Line](#) is a simple one-dimensional type of projectile motion in which there is no horizontal movement. In this section, we consider two-dimensional projectile motion, and our treatment neglects the effects of air resistance.

The most important fact to remember here is that *motions along perpendicular axes are independent* and thus can be analyzed separately. We discussed this fact in [Displacement and Velocity Vectors](#), where we saw that vertical and horizontal motions are independent. The key to analyzing two-dimensional projectile motion is to break it into two motions: one along the horizontal axis and the other along the vertical. (This choice of axes is the most sensible because acceleration resulting from gravity is vertical; thus, there is no acceleration along the horizontal axis when air resistance is negligible.) As is customary, we call the horizontal axis the x -axis and the vertical axis the y -axis. It is not required that we use this choice of axes; it is simply convenient in the case of gravitational acceleration. In other cases we may choose a different set of axes. [\[link\]](#) illustrates the notation for displacement, where we define \vec{s} to be the total displacement, and \vec{x} and \vec{y} are its component vectors along the horizontal and vertical axes, respectively. The magnitudes of these vectors are s , x , and y .



The total displacement s of a soccer ball at a point along its path. The vector \vec{s} has components \vec{x} and \vec{y} along the horizontal and vertical axes. Its magnitude is s and it makes an angle Φ with the horizontal.

To describe projectile motion completely, we must include velocity and acceleration, as well as displacement. We must find their components along the x - and y -axes. Let's assume all forces except gravity (such as air resistance and friction, for example) are negligible. Defining the positive direction to be upward, the components of acceleration are then very simple:

Equation:

$$a_y = -g = -9.8 \text{ m/s}^2 \quad (-32 \text{ ft/s}^2).$$

Because gravity is vertical, $a_x = 0$. If $a_x = 0$, this means the initial velocity in the x direction is equal to the final velocity in the x direction, or $v_x = v_{0x}$. With these conditions on acceleration and velocity, we can write the kinematic [\[link\]](#) through [\[link\]](#) for motion in a uniform gravitational field, including the rest of the kinematic equations for a constant acceleration from [Motion with Constant Acceleration](#). The kinematic equations for motion in a

uniform gravitational field become kinematic equations with
 $a_y = -g, a_x = 0$:

Horizontal Motion

Equation:

$$v_{0x} = v_x, x = x_0 + v_x t$$

Vertical Motion

Equation:

$$y = y_0 + \frac{1}{2}(v_{0y} + v_y)t$$

Equation:

$$v_y = v_{0y} - gt$$

Equation:

$$y = y_0 + v_{0y}t - \frac{1}{2}gt^2$$

Equation:

$$v_y^2 = v_{0y}^2 - 2g(y - y_0)$$

Using this set of equations, we can analyze projectile motion, keeping in mind some important points.

Note:

Projectile Motion

1. Resolve the motion into horizontal and vertical components along the x - and y -axes. The magnitudes of the components of displacement \vec{s} along these axes are x and y . The magnitudes of the components of

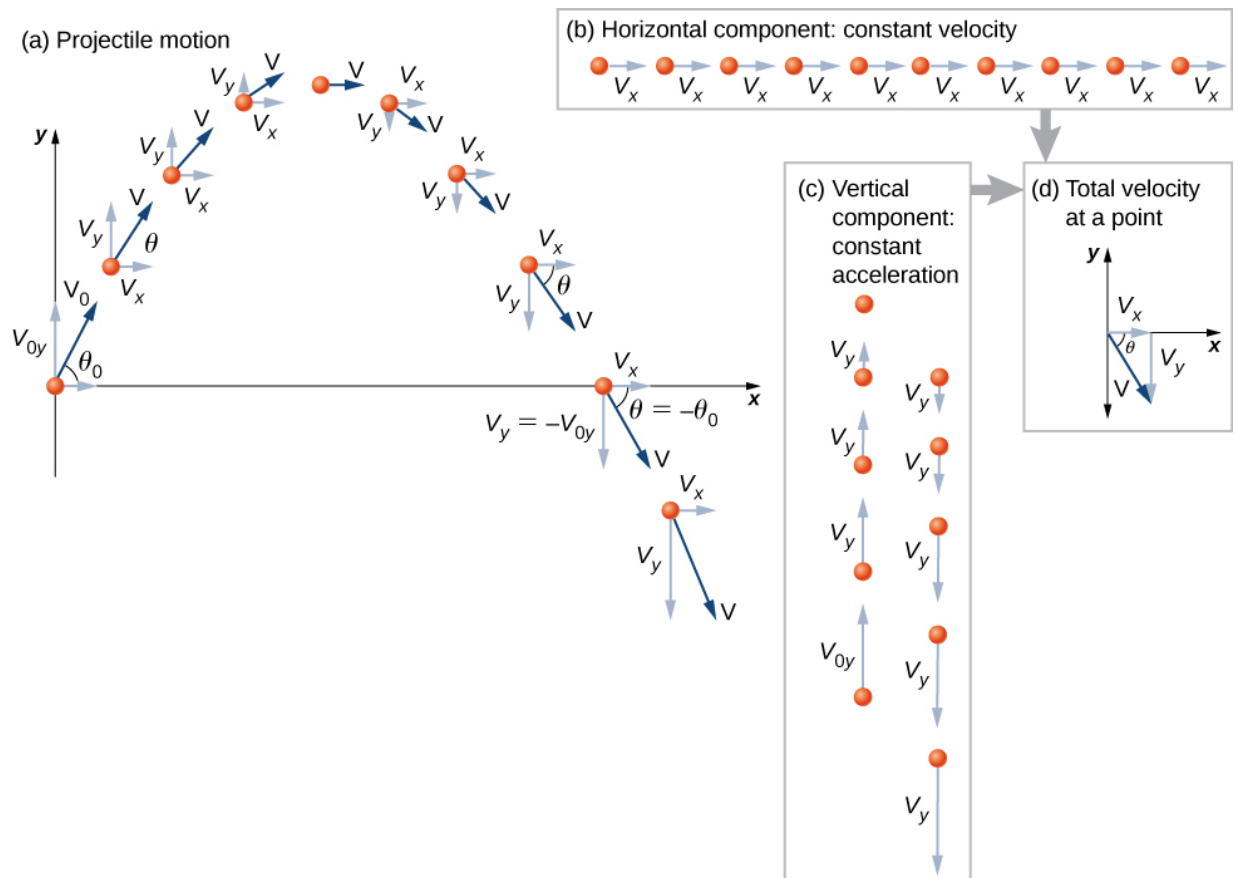
velocity \vec{v} are $v_x = v\cos\theta$ and $v_y = v\sin\theta$, where v is the magnitude of the velocity and θ is its direction relative to the horizontal, as shown in [\[link\]](#).

2. Treat the motion as two independent one-dimensional motions: one horizontal and the other vertical. Use the kinematic equations for horizontal and vertical motion presented earlier.
3. Solve for the unknowns in the two separate motions: one horizontal and one vertical. Note that the only common variable between the motions is time t . The problem-solving procedures here are the same as those for one-dimensional kinematics and are illustrated in the following solved examples.
4. Recombine quantities in the horizontal and vertical directions to find the total displacement \vec{s} and velocity \vec{v} . Solve for the magnitude and direction of the displacement and velocity using

Equation:

$$s = \sqrt{x^2 + y^2}, \quad \Phi = \tan^{-1}(y/x), \quad v = \sqrt{v_x^2 + v_y^2},$$

where Φ is the direction of the displacement \vec{s} .

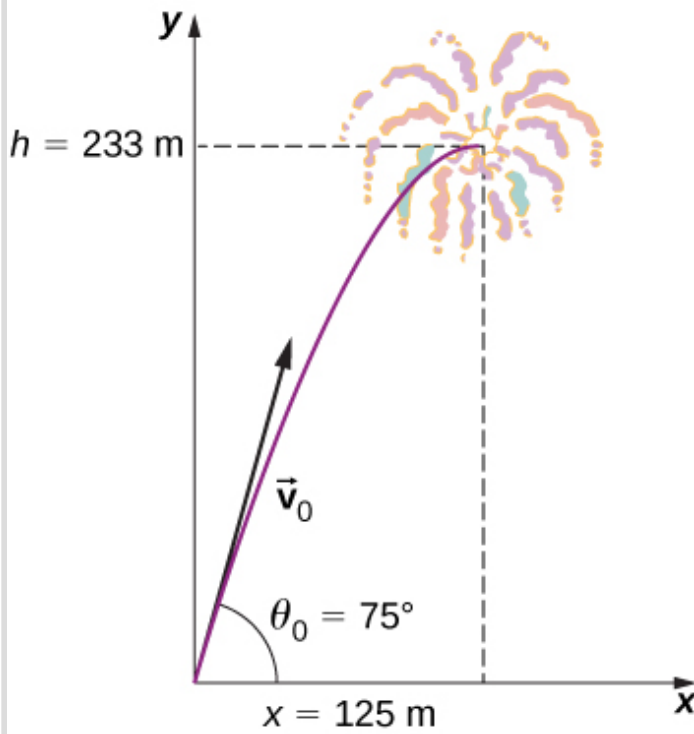


(a) We analyze two-dimensional projectile motion by breaking it into two independent one-dimensional motions along the vertical and horizontal axes. (b) The horizontal motion is simple, because $a_x = 0$ and v_x is a constant. (c) The velocity in the vertical direction begins to decrease as the object rises. At its highest point, the vertical velocity is zero. As the object falls toward Earth again, the vertical velocity increases again in magnitude but points in the opposite direction to the initial vertical velocity. (d) The x and y motions are recombined to give the total velocity at any given point on the trajectory.

Example:

A Fireworks Projectile Explodes High and Away

During a fireworks display, a shell is shot into the air with an initial speed of 70.0 m/s at an angle of 75.0° above the horizontal, as illustrated in [\[link\]](#). The fuse is timed to ignite the shell just as it reaches its highest point above the ground. (a) Calculate the height at which the shell explodes. (b) How much time passes between the launch of the shell and the explosion? (c) What is the horizontal displacement of the shell when it explodes? (d) What is the total displacement from the point of launch to the highest point?



The trajectory of a fireworks shell. The fuse is set to explode the shell at the highest point in its trajectory, which is found to be at a height of 233 m and 125 m away horizontally.

Strategy

The motion can be broken into horizontal and vertical motions in which $a_x = 0$ and $a_y = -g$. We can then define x_0 and y_0 to be zero and solve for the desired quantities.

Solution

(a) By “height” we mean the altitude or vertical position y above the starting point. The highest point in any trajectory, called the *apex*, is reached when $v_y = 0$. Since we know the initial and final velocities, as well as the initial position, we use the following equation to find y :

Equation:

$$v_y^2 = v_{0y}^2 - 2g(y - y_0).$$

Because y_0 and v_y are both zero, the equation simplifies to

Equation:

$$0 = v_{0y}^2 - 2gy.$$

Solving for y gives

Equation:

$$y = \frac{v_{0y}^2}{2g}.$$

Now we must find v_{0y} , the component of the initial velocity in the y direction. It is given by $v_{0y} = v_0 \sin \theta_0$, where v_0 is the initial velocity of 70.0 m/s and $\theta_0 = 75^\circ$ is the initial angle. Thus,

Equation:

$$v_{0y} = v_0 \sin \theta = (70.0 \text{ m/s}) \sin 75^\circ = 67.6 \text{ m/s}$$

and y is

Equation:

$$y = \frac{(67.6 \text{ m/s})^2}{2(9.80 \text{ m/s}^2)}.$$

Thus, we have

Equation:

$$y = 233 \text{ m}.$$

Note that because up is positive, the initial vertical velocity is positive, as is the maximum height, but the acceleration resulting from gravity is negative. Note also that the maximum height depends only on the vertical component of the initial velocity, so that any projectile with a 67.6-m/s initial vertical component of velocity reaches a maximum height of 233 m (neglecting air resistance). The numbers in this example are reasonable for large fireworks displays, the shells of which do reach such heights before exploding. In practice, air resistance is not completely negligible, so the initial velocity would have to be somewhat larger than that given to reach the same height. (b) As in many physics problems, there is more than one way to solve for the time the projectile reaches its highest point. In this case, the easiest method is to use $v_y = v_{0y} - gt$. Because $v_y = 0$ at the apex, this equation reduces to simply

Equation:

$$0 = v_{0y} - gt$$

or

Equation:

$$t = \frac{v_{0y}}{g} = \frac{67.6 \text{ m/s}}{9.80 \text{ m/s}^2} = 6.90\text{s}.$$

This time is also reasonable for large fireworks. If you are able to see the launch of fireworks, notice that several seconds pass before the shell explodes. Another way of finding the time is by using

$y = y_0 + \frac{1}{2}(v_{0y} + v_y)t$. This is left for you as an exercise to complete.

(c) Because air resistance is negligible, $a_x = 0$ and the horizontal velocity is constant, as discussed earlier. The horizontal displacement is the horizontal velocity multiplied by time as given by $x = x_0 + v_x t$, where x_0 is equal to zero. Thus,

Equation:

$$x = v_x t,$$

where v_x is the x-component of the velocity, which is given by

Equation:

$$v_x = v_0 \cos \theta = (70.0 \text{ m/s}) \cos 75^\circ = 18.1 \text{ m/s}.$$

Time t for both motions is the same, so x is

Equation:

$$x = (18.1 \text{ m/s}) 6.90 \text{ s} = 125 \text{ m}.$$

Horizontal motion is a constant velocity in the absence of air resistance. The horizontal displacement found here could be useful in keeping the fireworks fragments from falling on spectators. When the shell explodes, air resistance has a major effect, and many fragments land directly below.

(d) The horizontal and vertical components of the displacement were just calculated, so all that is needed here is to find the magnitude and direction of the displacement at the highest point:

Equation:

$$\vec{s} = 125\hat{i} + 233\hat{j}$$

Equation:

$$|\vec{s}| = \sqrt{125^2 + 233^2} = 264 \text{ m}$$

Equation:

$$\Phi = \tan^{-1} \left(\frac{233}{125} \right) = 61.8^\circ.$$

Note that the angle for the displacement vector is less than the initial angle of launch. To see why this is, review [\[link\]](#), which shows the curvature of the trajectory toward the ground level.

When solving [\[link\]](#)(a), the expression we found for y is valid for any projectile motion when air resistance is negligible. Call the maximum height $y = h$. Then,

Equation:

$$h = \frac{v_{0y}^2}{2g}.$$

This equation defines the *maximum height of a projectile above its launch position* and it depends only on the vertical component of the initial velocity.

Note:

Exercise:

Problem:

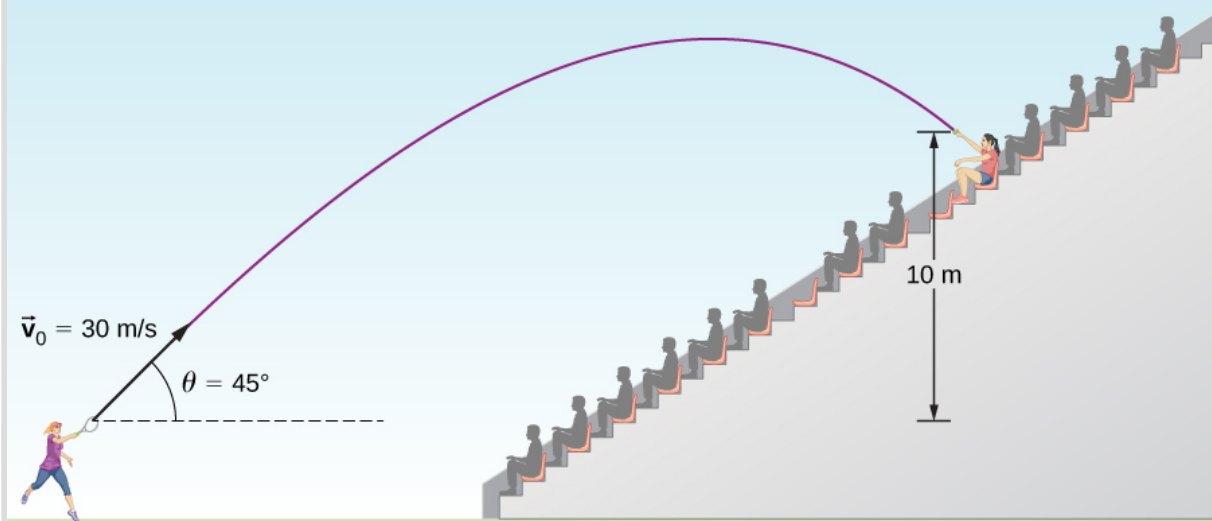
Check Your Understanding A rock is thrown horizontally off a cliff 100.0 m high with a velocity of 15.0 m/s. (a) Define the origin of the coordinate system. (b) Which equation describes the horizontal motion? (c) Which equations describe the vertical motion? (d) What is the rock's velocity at the point of impact?

Solution:

(a) Choose the top of the cliff where the rock is thrown from the origin of the coordinate system. Although it is arbitrary, we typically choose time $t = 0$ to correspond to the origin. (b) The equation that describes the horizontal motion is $x = x_0 + v_x t$. With $x_0 = 0$, this equation becomes $x = v_x t$. (c) [\[link\]](#) through [\[link\]](#) and [\[link\]](#) describe the vertical motion, but since $y_0 = 0$ and $v_{0y} = 0$, these equations simplify greatly to become $y = \frac{1}{2}(v_{0y} + v_y)t = \frac{1}{2}v_y t$, $v_y = -gt$, $y = -\frac{1}{2}gt^2$, and $v_y^2 = -2gy$. (d) We use the kinematic equations to find the x and y components of the velocity at the point of impact. Using $v_y^2 = -2gy$ and noting the point of impact is -100.0 m, we find the y component of the velocity at impact is $v_y = 44.3$ m/s. We are given the x component, $v_x = 15.0$ m/s, so we can calculate the total velocity at impact: $v = 46.8$ m/s and $\theta = 71.3^\circ$ below the horizontal.

Example:**Calculating Projectile Motion: Tennis Player**

A tennis player wins a match at Arthur Ashe stadium and hits a ball into the stands at 30 m/s and at an angle 45° above the horizontal ([\[link\]](#)). On its way down, the ball is caught by a spectator 10 m above the point where the ball was hit. (a) Calculate the time it takes the tennis ball to reach the spectator. (b) What are the magnitude and direction of the ball's velocity at impact?



The trajectory of a tennis ball hit into the stands.

Strategy

Again, resolving this two-dimensional motion into two independent one-dimensional motions allows us to solve for the desired quantities. The time a projectile is in the air is governed by its vertical motion alone. Thus, we solve for t first. While the ball is rising and falling vertically, the horizontal motion continues at a constant velocity. This example asks for the final velocity. Thus, we recombine the vertical and horizontal results to obtain \vec{v} at final time t , determined in the first part of the example.

Solution

(a) While the ball is in the air, it rises and then falls to a final position 10.0 m higher than its starting altitude. We can find the time for this by using [\[link\]](#):

Equation:

$$y = y_0 + v_{0y}t - \frac{1}{2}gt^2.$$

If we take the initial position y_0 to be zero, then the final position is $y = 10$ m. The initial vertical velocity is the vertical component of the initial velocity:

Equation:

$$v_{0y} = v_0 \sin \theta_0 = (30.0 \text{ m/s}) \sin 45^\circ = 21.2 \text{ m/s}.$$

Substituting into [\[link\]](#) for y gives us

Equation:

$$10.0 \text{ m} = (21.2 \text{ m/s})t - (4.90 \text{ m/s}^2)t^2.$$

Rearranging terms gives a quadratic equation in t :

Equation:

$$(4.90 \text{ m/s}^2)t^2 - (21.2 \text{ m/s})t + 10.0 \text{ m} = 0.$$

Use of the quadratic formula yields $t = 3.79$ s and $t = 0.54$ s. Since the ball is at a height of 10 m at two times during its trajectory—once on the way up and once on the way down—we take the longer solution for the time it takes the ball to reach the spectator:

Equation:

$$t = 3.79 \text{ s}.$$

The time for projectile motion is determined completely by the vertical motion. Thus, any projectile that has an initial vertical velocity of 21.2 m/s and lands 10.0 m below its starting altitude spends 3.79 s in the air.

(b) We can find the final horizontal and vertical velocities v_x and v_y with the use of the result from (a). Then, we can combine them to find the magnitude of the total velocity vector \vec{v} and the angle θ it makes with the horizontal. Since v_x is constant, we can solve for it at any horizontal location. We choose the starting point because we know both the initial velocity and the initial angle. Therefore,

Equation:

$$v_x = v_0 \cos \theta_0 = (30 \text{ m/s}) \cos 45^\circ = 21.2 \text{ m/s}.$$

The final vertical velocity is given by [\[link\]](#):

Equation:

$$v_y = v_{0y} - gt.$$

Since v_{0y} was found in part (a) to be 21.2 m/s, we have

Equation:

$$v_y = 21.2 \text{ m/s} - 9.8 \text{ m/s}^2 (3.79 \text{ s}) = -15.9 \text{ m/s}.$$

The magnitude of the final velocity \vec{v} is

Equation:

$$v = \sqrt{v_x^2 + v_y^2} = \sqrt{(21.2 \text{ m/s})^2 + (-15.9 \text{ m/s})^2} = 26.5 \text{ m/s}.$$

The direction θ_v is found using the inverse tangent:

Equation:

$$\theta_v = \tan^{-1} \left(\frac{v_y}{v_x} \right) = \tan^{-1} \left(\frac{-15.9}{21.2} \right) = 36.9^\circ.$$

Significance

(a) As mentioned earlier, the time for projectile motion is determined completely by the vertical motion. Thus, any projectile that has an initial vertical velocity of 21.2 m/s and lands 10.0 m above its starting altitude spends 3.79 s in the air. (b) The negative angle means the velocity is 36.9° below the horizontal at the point of impact. This result is consistent with the fact that the ball is impacting at a point on the other side of the apex of the trajectory and therefore has a negative y component of the velocity. The magnitude of the velocity is less than the magnitude of the initial velocity we expect since it is impacting 10.0 m above the launch elevation.

Time of Flight, Trajectory, and Range

Of interest are the time of flight, trajectory, and range for a projectile launched on a flat horizontal surface and impacting on the same surface. In this case, kinematic equations give useful expressions for these quantities, which are derived in the following sections.

Time of flight

We can solve for the time of flight of a projectile that is both launched and impacts on a flat horizontal surface by performing some manipulations of the kinematic equations. We note the position and displacement in y must be zero at launch and at impact on an even surface. Thus, we set the displacement in y equal to zero and find

Equation:

$$y - y_0 = v_{0y}t - \frac{1}{2}gt^2 = (v_0\sin\theta_0)t - \frac{1}{2}gt^2 = 0.$$

Factoring, we have

Equation:

$$t \left(v_0\sin\theta_0 - \frac{gt}{2} \right) = 0.$$

Solving for t gives us

Note:

Equation:

$$T_{\text{tof}} = \frac{2(v_0\sin\theta_0)}{g}.$$

This is the **time of flight** for a projectile both launched and impacting on a flat horizontal surface. [\[link\]](#) does not apply when the projectile lands at a different elevation than it was launched, as we saw in [\[link\]](#) of the tennis player hitting the ball into the stands. The other solution, $t = 0$, corresponds to the time at launch. The time of flight is linearly proportional to the initial velocity in the y direction and inversely proportional to g . Thus, on the Moon, where gravity is one-sixth that of Earth, a projectile launched with the same velocity as on Earth would be airborne six times as long.

Trajectory

The trajectory of a projectile can be found by eliminating the time variable t from the kinematic equations for arbitrary t and solving for $y(x)$. We take $x_0 = y_0 = 0$ so the projectile is launched from the origin. The kinematic equation for x gives

Equation:

$$x = v_{0x}t \Rightarrow t = \frac{x}{v_{0x}} = \frac{x}{v_0 \cos \theta_0}.$$

Substituting the expression for t into the equation for the position $y = (v_0 \sin \theta_0)t - \frac{1}{2}gt^2$ gives

Equation:

$$y = (v_0 \sin \theta_0) \left(\frac{x}{v_0 \cos \theta_0} \right) - \frac{1}{2}g \left(\frac{x}{v_0 \cos \theta_0} \right)^2.$$

Rearranging terms, we have

Note:

Equation:

$$y = (\tan\theta_0)x - \left[\frac{g}{2(v_0\cos\theta_0)^2} \right] x^2.$$

This trajectory equation is of the form $y = ax + bx^2$, which is an equation of a parabola with coefficients

Equation:

$$a = \tan\theta_0, \quad b = -\frac{g}{2(v_0\cos\theta_0)^2}.$$

Range

From the trajectory equation we can also find the **range**, or the horizontal distance traveled by the projectile. Factoring [\[link\]](#), we have

Equation:

$$y = x \left[\tan\theta_0 - \frac{g}{2(v_0\cos\theta_0)^2} x \right].$$

The position y is zero for both the launch point and the impact point, since we are again considering only a flat horizontal surface. Setting $y = 0$ in this equation gives solutions $x = 0$, corresponding to the launch point, and

Equation:

$$x = \frac{2v_0^2\sin\theta_0\cos\theta_0}{g},$$

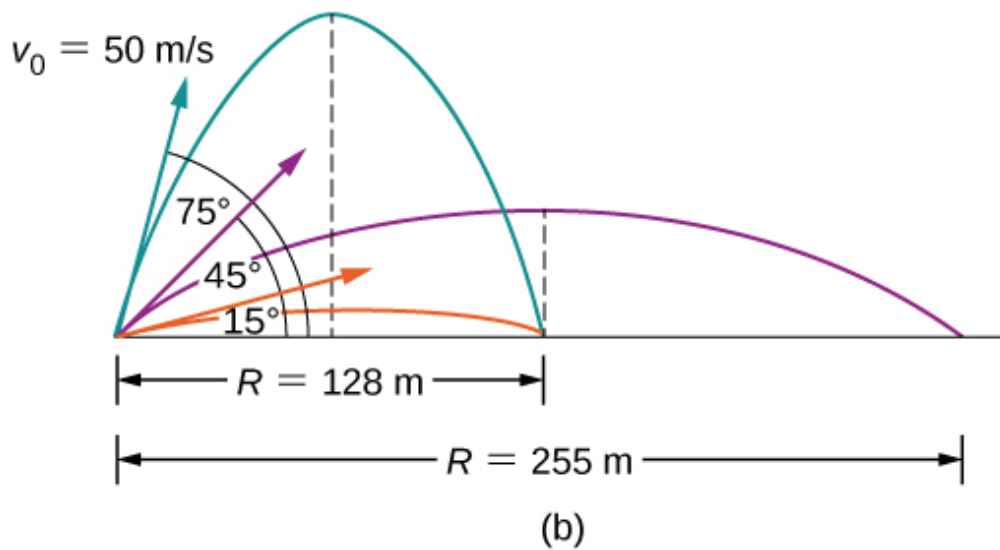
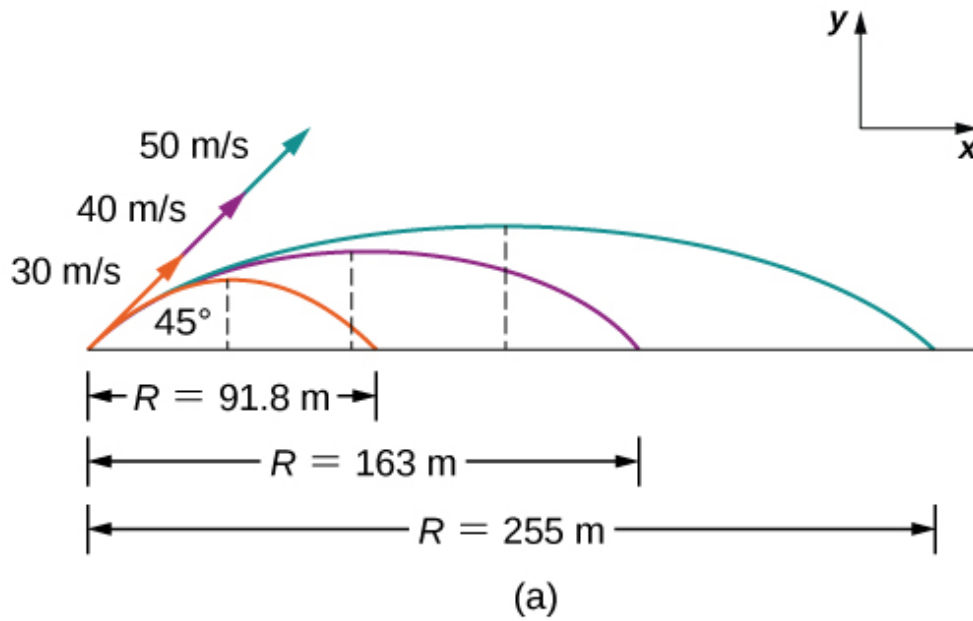
corresponding to the impact point. Using the trigonometric identity $2\sin\theta\cos\theta = \sin 2\theta$ and setting $x = R$ for range, we find

Note:

Equation:

$$R = \frac{v_0^2 \sin 2\theta_0}{g}.$$

Note particularly that [\[link\]](#) is valid only for launch and impact on a horizontal surface. We see the range is directly proportional to the square of the initial speed v_0 and $\sin 2\theta_0$, and it is inversely proportional to the acceleration of gravity. Thus, on the Moon, the range would be six times greater than on Earth for the same initial velocity. Furthermore, we see from the factor $\sin 2\theta_0$ that the range is maximum at 45° . These results are shown in [\[link\]](#). In (a) we see that the greater the initial velocity, the greater the range. In (b), we see that the range is maximum at 45° . This is true only for conditions neglecting air resistance. If air resistance is considered, the maximum angle is somewhat smaller. It is interesting that the same range is found for two initial launch angles that sum to 90° . The projectile launched with the smaller angle has a lower apex than the higher angle, but they both have the same range.



Trajectories of projectiles on level ground. (a) The greater the initial speed v_0 , the greater the range for a given initial angle. (b) The effect of initial angle θ_0 on the range of a projectile with a given initial speed. Note that the range is the same for initial angles of 15° and 75° , although the maximum heights of those paths are different.

Example:**Comparing Golf Shots**

A golfer finds himself in two different situations on different holes. On the second hole he is 120 m from the green and wants to hit the ball 90 m and let it run onto the green. He angles the shot low to the ground at 30° to the horizontal to let the ball roll after impact. On the fourth hole he is 90 m from the green and wants to let the ball drop with a minimum amount of rolling after impact. Here, he angles the shot at 70° to the horizontal to minimize rolling after impact. Both shots are hit and impacted on a level surface.

(a) What is the initial speed of the ball at the second hole?

(b) What is the initial speed of the ball at the fourth hole?

(c) Write the trajectory equation for both cases.

(d) Graph the trajectories.

Strategy

We see that the range equation has the initial speed and angle, so we can solve for the initial speed for both (a) and (b). When we have the initial speed, we can use this value to write the trajectory equation.

Solution

$$(a) R = \frac{v_0^2 \sin 2\theta_0}{g} \Rightarrow v_0 = \sqrt{\frac{Rg}{\sin 2\theta_0}} = \sqrt{\frac{90.0 \text{ m}(9.8 \text{ m/s}^2)}{\sin(2(30^\circ))}} = 31.9 \text{ m/s}$$

$$(b) R = \frac{v_0^2 \sin 2\theta_0}{g} \Rightarrow v_0 = \sqrt{\frac{Rg}{\sin 2\theta_0}} = \sqrt{\frac{90.0 \text{ m}(9.8 \text{ m/s}^2)}{\sin(2(70^\circ))}} = 37.0 \text{ m/s}$$

(c)

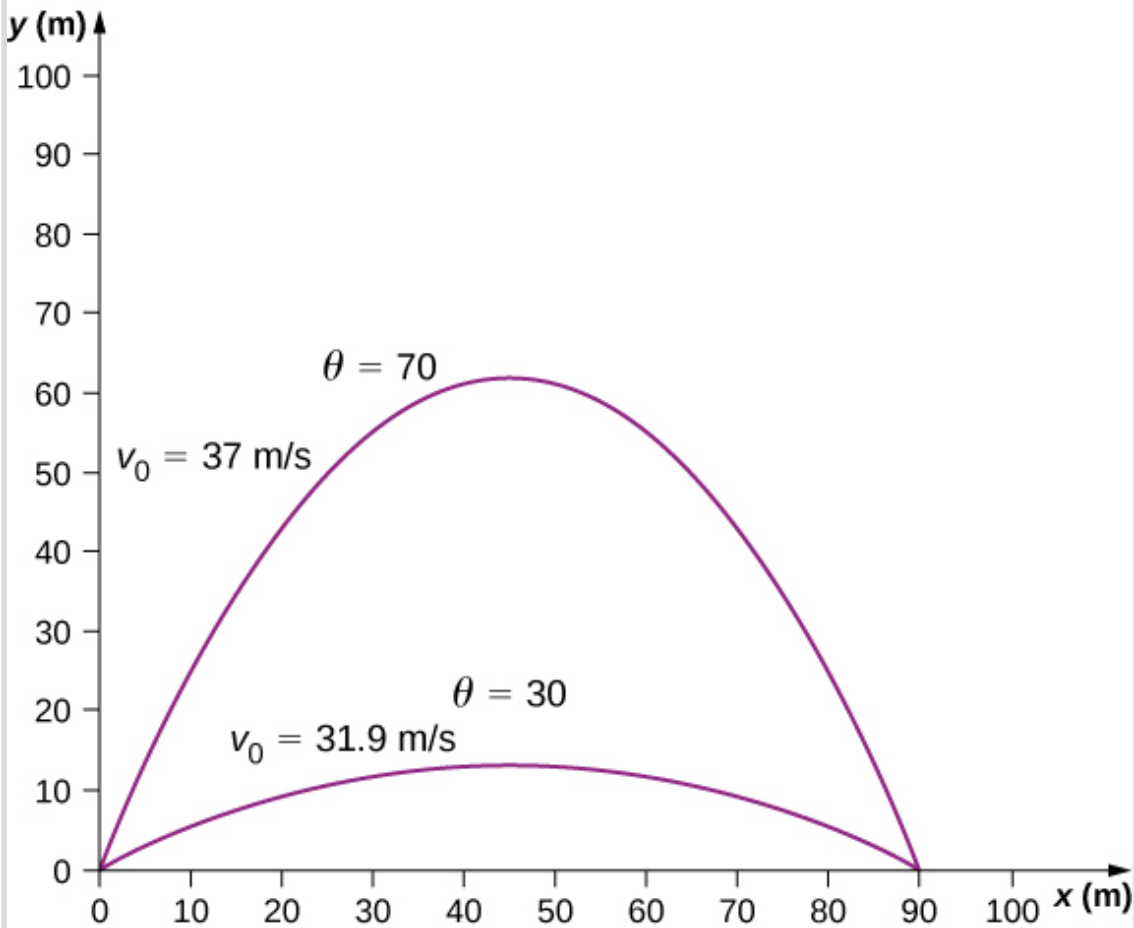
$$y = x \left[\tan \theta_0 - \frac{g}{2(v_0 \cos \theta_0)^2} x \right]$$

$$\text{Second hole: } y = x \left[\tan 30^\circ - \frac{9.8 \text{ m/s}^2}{2[(31.9 \text{ m/s})(\cos 30^\circ)]^2} x \right] = 0.58x - 0.0064x^2$$

$$\text{Fourth hole: } y = x \left[\tan 70^\circ - \frac{9.8 \text{ m/s}^2}{2[(37.0 \text{ m/s})(\cos 70^\circ)]^2} x \right] = 2.75x - 0.0306x^2$$

(d) Using a graphing utility, we can compare the two trajectories, which are shown in [\[link\]](#).

Golf Shot



Two trajectories of a golf ball with a range of 90 m. The impact points of both are at the same level as the launch point.

Significance

The initial speed for the shot at 70° is greater than the initial speed of the shot at 30° . Note from [\[link\]](#) that two projectiles launched at the same speed but at different angles have the same range if the launch angles add to 90° . The launch angles in this example add to give a number greater than 90° . Thus, the shot at 70° has to have a greater launch speed to reach 90 m, otherwise it would land at a shorter distance.

Note:

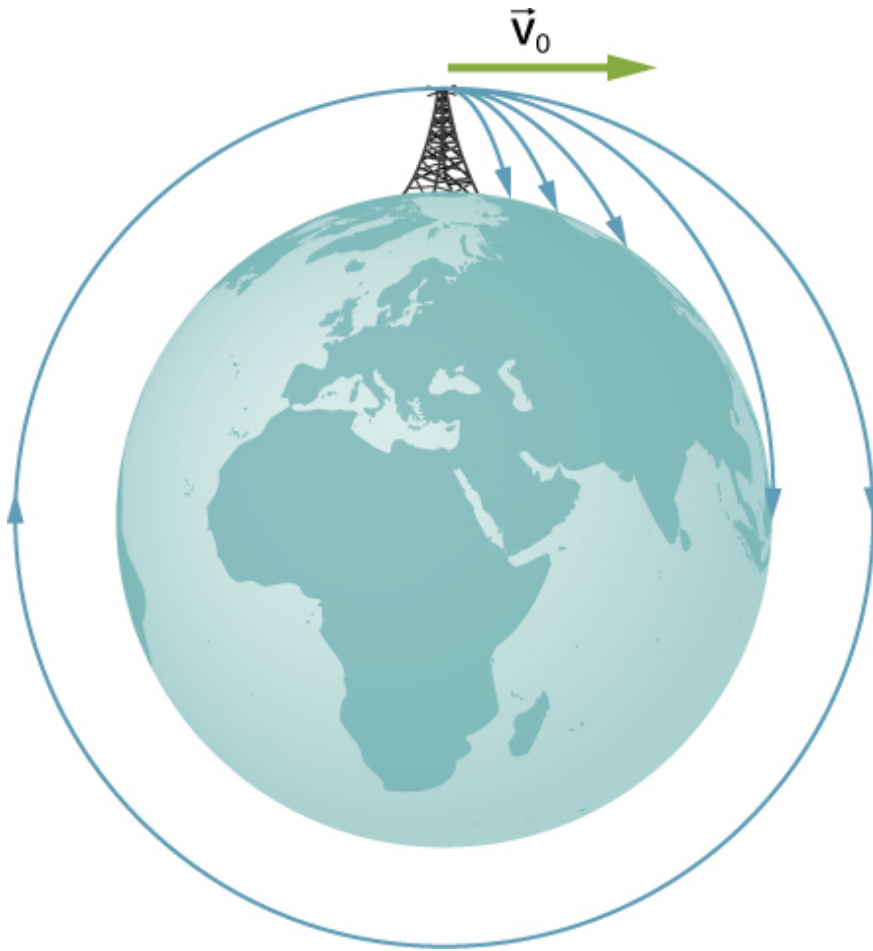
Exercise:**Problem:**

Check Your Understanding If the two golf shots in [\[link\]](#) were launched at the same speed, which shot would have the greatest range?

Solution:

The golf shot at 30° .

When we speak of the range of a projectile on level ground, we assume R is very small compared with the circumference of Earth. If, however, the range is large, Earth curves away below the projectile and the acceleration resulting from gravity changes direction along the path. The range is larger than predicted by the range equation given earlier because the projectile has farther to fall than it would on level ground, as shown in [\[link\]](#), which is based on a drawing in Newton's *Principia*. If the initial speed is great enough, the projectile goes into orbit. Earth's surface drops 5 m every 8000 m. In 1 s an object falls 5 m without air resistance. Thus, if an object is given a horizontal velocity of 8000 m/s (or 18,000 mi/hr) near Earth's surface, it will go into orbit around the planet because the surface continuously falls away from the object. This is roughly the speed of the Space Shuttle in a low Earth orbit when it was operational, or any satellite in a low Earth orbit. These and other aspects of orbital motion, such as Earth's rotation, are covered in greater depth in [Gravitation](#).



Projectile to satellite. In each case shown here, a projectile is launched from a very high tower to avoid air resistance. With increasing initial speed, the range increases and becomes longer than it would be on level ground because Earth curves away beneath its path. With a speed of 8000 m/s, orbit is achieved.

Note:

At [PhET Explorations: Projectile Motion](#), learn about projectile motion in terms of the launch angle and initial velocity.

Summary

- Projectile motion is the motion of an object subject only to the acceleration of gravity, where the acceleration is constant, as near the surface of Earth.
- To solve projectile motion problems, we analyze the motion of the projectile in the horizontal and vertical directions using the one-dimensional kinematic equations for x and y .
- The time of flight of a projectile launched with initial vertical velocity v_{0y} on an even surface is given by

Equation:

$$T_{tof} = \frac{2(v_0 \sin \theta)}{g}.$$

This equation is valid only when the projectile lands at the same elevation from which it was launched.

- The maximum horizontal distance traveled by a projectile is called the range. Again, the equation for range is valid only when the projectile lands at the same elevation from which it was launched.

Conceptual Questions

Exercise:

Problem:

Answer the following questions for projectile motion on level ground assuming negligible air resistance, with the initial angle being neither 0° nor 90° : (a) Is the velocity ever zero? (b) When is the velocity a minimum? A maximum? (c) Can the velocity ever be the same as the initial velocity at a time other than at $t = 0$? (d) Can the speed ever be the same as the initial speed at a time other than at $t = 0$?

Solution:

a. no; b. minimum at apex of trajectory and maximum at launch and impact; c. no, velocity is a vector; d. yes, where it lands

Exercise:**Problem:**

Answer the following questions for projectile motion on level ground assuming negligible air resistance, with the initial angle being neither 0° nor 90° : (a) Is the acceleration ever zero? (b) Is the vector \vec{v} ever parallel or antiparallel to the vector \vec{a} ? (c) Is the vector v ever perpendicular to the vector a ? If so, where is this located?

Exercise:**Problem:**

A dime is placed at the edge of a table so it hangs over slightly. A quarter is slid horizontally on the table surface perpendicular to the edge and hits the dime head on. Which coin hits the ground first?

Solution:

They both hit the ground at the same time.

Problems**Exercise:****Problem:**

A bullet is shot horizontally from shoulder height (1.5 m) with an initial speed 200 m/s. (a) How much time elapses before the bullet hits the ground? (b) How far does the bullet travel horizontally?

Solution:

a. $t = 0.55$ s, b. $x = 110$ m

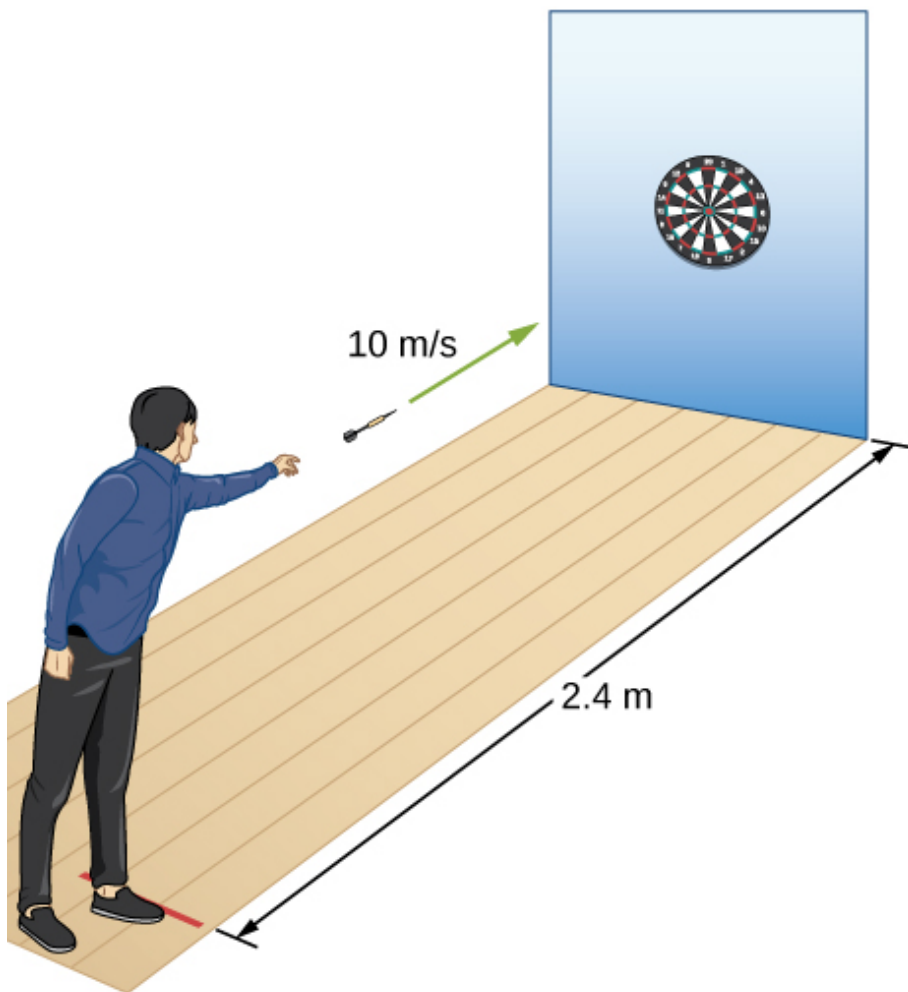
Exercise:

Problem:

A marble rolls off a tabletop 1.0 m high and hits the floor at a point 3.0 m away from the table's edge in the horizontal direction. (a) How long is the marble in the air? (b) What is the speed of the marble when it leaves the table's edge? (c) What is its speed when it hits the floor?

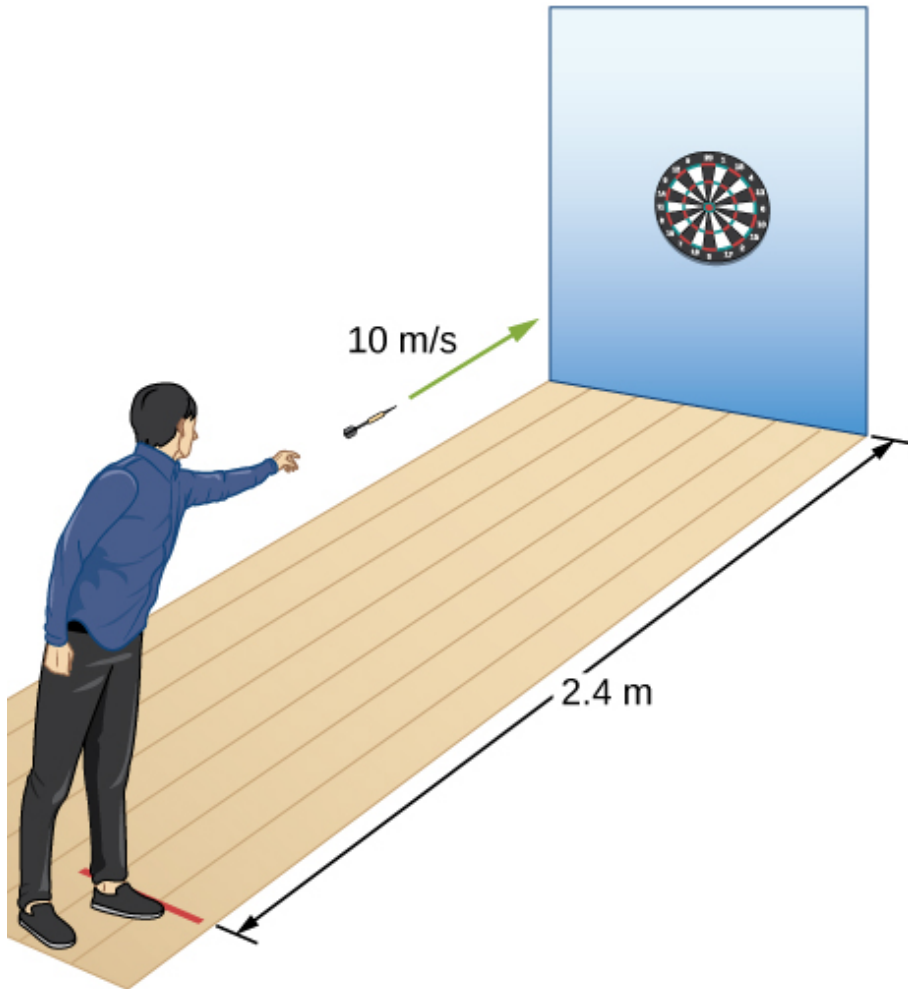
Exercise:**Problem:**

A dart is thrown horizontally at a speed of 10 m/s at the bull's-eye of a dartboard 2.4 m away, as in the following figure. (a) How far below the intended target does the dart hit? (b) What does your answer tell you about how proficient dart players throw their darts?

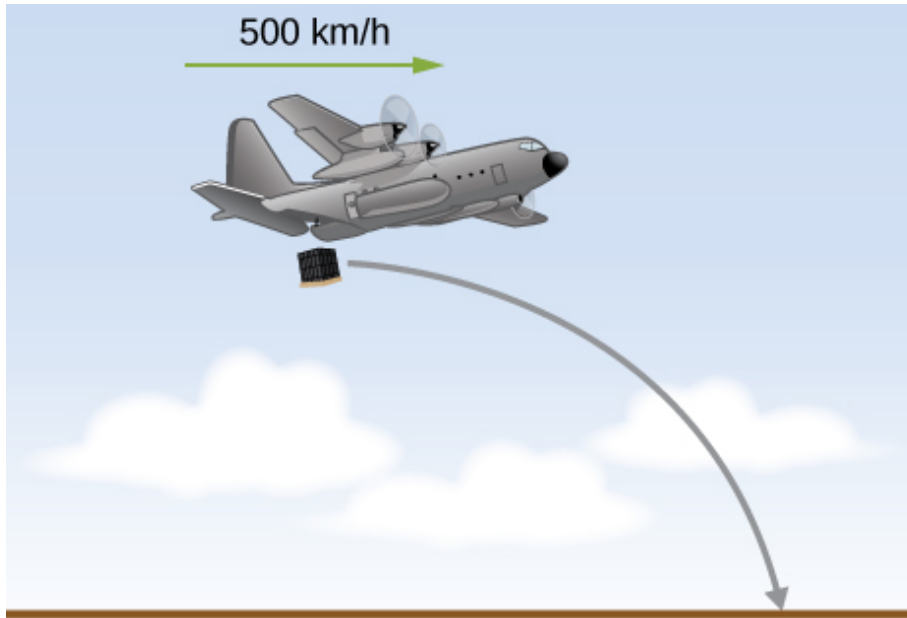


Solution:

a. $t = 0.24\text{s}$, $d = 0.28\text{ m}$, b. They aim high.

**Exercise:****Problem:**

An airplane flying horizontally with a speed of 500 km/h at a height of 800 m drops a crate of supplies (see the following figure). If the parachute fails to open, how far in front of the release point does the crate hit the ground?



Exercise:

Problem:

Suppose the airplane in the preceding problem fires a projectile horizontally in its direction of motion at a speed of 300 m/s relative to the plane. (a) How far in front of the release point does the projectile hit the ground? (b) What is its speed when it hits the ground?

Solution:

a., $t = 12.8 \text{ s}$, $x = 5619 \text{ m}$ b.

$v_y = 125.0 \text{ m/s}$, $v_x = 439.0 \text{ m/s}$, $|\vec{v}| = 456.0 \text{ m/s}$

Exercise:

Problem:

A fastball pitcher can throw a baseball at a speed of 40 m/s (90 mi/h). (a) Assuming the pitcher can release the ball 16.7 m from home plate so the ball is moving horizontally, how long does it take the ball to reach home plate? (b) How far does the ball drop between the pitcher's hand and home plate?

Exercise:

Problem:

A projectile is launched at an angle of 30° and lands 20 s later at the same height as it was launched. (a) What is the initial speed of the projectile? (b) What is the maximum altitude? (c) What is the range? (d) Calculate the displacement from the point of launch to the position on its trajectory at 15 s.

Solution:

a.

$$v_y = v_{0y} - gt, \quad t = 10\text{s}, \quad v_y = 0, \quad v_{0y} = 98.0 \text{ m/s}, \quad v_0 = 196.0 \text{ m/s}$$

,

b. $h = 490.0 \text{ m},$

c. $v_{0x} = 169.7 \text{ m/s}, \quad x = 3394.0 \text{ m},$

$$x = 169.7 \text{ m/s} (15.0 \text{ s}) = 2550 \text{ m}$$

d. $y = (98.0 \text{ m/s}) (15.0 \text{ s}) - 4.9(15.0\text{s})^2 = 368 \text{ m}$

$$\vec{r} = 2550 \text{ m}\hat{i} + 368 \text{ m}\hat{j}$$

Exercise:**Problem:**

A basketball player shoots toward a basket 6.1 m away and 3.0 m above the floor. If the ball is released 1.8 m above the floor at an angle of 60° above the horizontal, what must the initial speed be if it were to go through the basket?

Exercise:**Problem:**

At a particular instant, a hot air balloon is 100 m in the air and descending at a constant speed of 2.0 m/s. At this exact instant, a girl throws a ball horizontally, relative to herself, with an initial speed of 20 m/s. When she lands, where will she find the ball? Ignore air resistance.

Solution:

$$-100 \text{ m} = (-2.0 \text{ m/s})t - (4.9 \text{ m/s}^2)t^2, t = 4.3 \text{ s}, x = 86.0 \text{ m}$$

Exercise:

Problem:

A man on a motorcycle traveling at a uniform speed of 10 m/s throws an empty can straight upward relative to himself with an initial speed of 3.0 m/s. Find the equation of the trajectory as seen by a police officer on the side of the road. Assume the initial position of the can is the point where it is thrown. Ignore air resistance.

Exercise:

Problem:

An athlete can jump a distance of 8.0 m in the broad jump. What is the maximum distance the athlete can jump on the Moon, where the gravitational acceleration is one-sixth that of Earth?

Solution:

$$R_{Moon} = 48 \text{ m}$$

Exercise:

Problem:

The maximum horizontal distance a boy can throw a ball is 50 m. Assume he can throw with the same initial speed at all angles. How high does he throw the ball when he throws it straight upward?

Exercise:

Problem:

A rock is thrown off a cliff at an angle of 53° with respect to the horizontal. The cliff is 100 m high. The initial speed of the rock is 30 m/s. (a) How high above the edge of the cliff does the rock rise? (b) How far has it moved horizontally when it is at maximum altitude? (c) How long after the release does it hit the ground? (d) What is the range of the rock? (e) What are the horizontal and vertical positions of the rock relative to the edge of the cliff at $t = 2.0 \text{ s}$, $t = 4.0 \text{ s}$, and $t = 6.0 \text{ s}$?

Solution:

a. $v_{0y} = 24 \text{ m/s}$ $v_y^2 = v_{0y}^2 - 2gy \Rightarrow h = 29.3 \text{ m}$,

b. $t = 2.4 \text{ s}$ $v_{0x} = 18 \text{ m/s}$ $x = 43.2 \text{ m}$,

c. $y = -100 \text{ m}$ $y_0 = 0$ $y - y_0 = v_{0y}t - \frac{1}{2}gt^2$ $-100 = 24t - 4.9t^2$
 $\Rightarrow t = 7.58 \text{ s}$,

d. $x = 136.44 \text{ m}$,

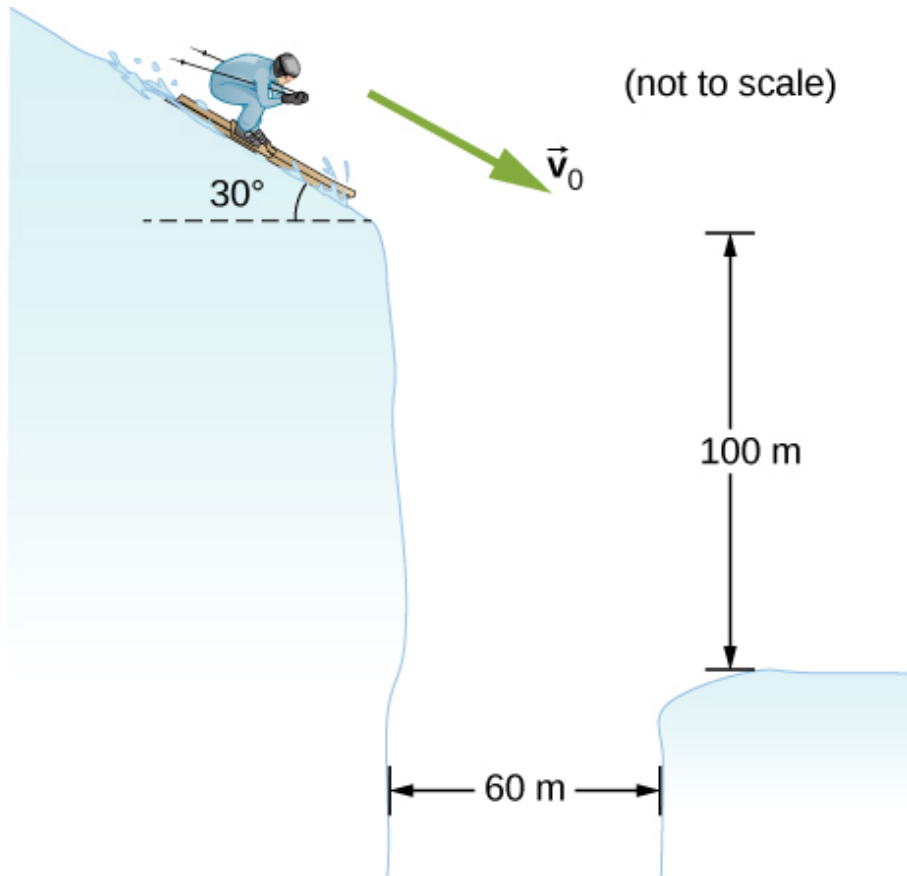
e. $t = 2.0 \text{ s}$ $y = 28.4 \text{ m}$ $x = 36 \text{ m}$

$t = 4.0 \text{ s}$ $y = 17.6 \text{ m}$ $x = 72 \text{ m}$

$t = 6.0 \text{ s}$ $y = -32.4 \text{ m}$ $x = 108 \text{ m}$

Exercise:**Problem:**

Trying to escape his pursuers, a secret agent skis off a slope inclined at 30° below the horizontal at 60 km/h . To survive and land on the snow 100 m below, he must clear a gorge 60 m wide. Does he make it? Ignore air resistance.



Exercise:

Problem:

A golfer on a fairway is 70 m away from the green, which sits below the level of the fairway by 20 m. If the golfer hits the ball at an angle of 40° with an initial speed of 20 m/s, how close to the green does she come?

Solution:

$$v_{0y} = 12.9 \text{ m/s} \quad y - y_0 = v_{0y}t - \frac{1}{2}gt^2 \quad -20.0 = 12.9t - 4.9t^2$$

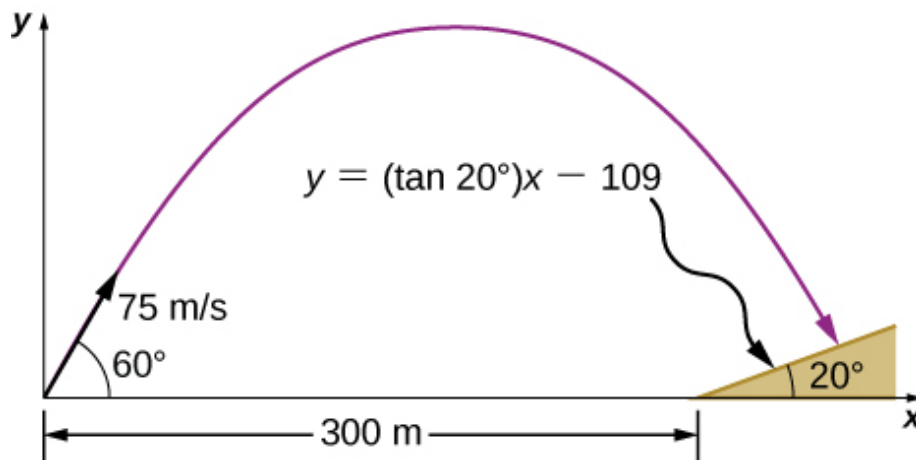
$$t = 3.7 \text{ s} \quad v_{0x} = 15.3 \text{ m/s} \Rightarrow x = 56.7 \text{ m}$$

So the golfer's shot lands 13.3 m short of the green.

Exercise:

Problem:

A projectile is shot at a hill, the base of which is 300 m away. The projectile is shot at 60° above the horizontal with an initial speed of 75 m/s. The hill can be approximated by a plane sloped at 20° to the horizontal. Relative to the coordinate system shown in the following figure, the equation of this straight line is $y = (\tan 20^\circ)x - 109$. Where on the hill does the projectile land?

**Exercise:****Problem:**

An astronaut on Mars kicks a soccer ball at an angle of 45° with an initial velocity of 15 m/s. If the acceleration of gravity on Mars is 3.7 m/s^2 , (a) what is the range of the soccer kick on a flat surface? (b) What would be the range of the same kick on the Moon, where gravity is one-sixth that of Earth?

Solution:

- a. $R = 60.8 \text{ m}$,
- b. $R = 137.8 \text{ m}$

Exercise:

Problem:

Mike Powell holds the record for the long jump of 8.95 m, established in 1991. If he left the ground at an angle of 15° , what was his initial speed?

Exercise:**Problem:**

MIT's robot cheetah can jump over obstacles 46 cm high and has speed of 12.0 km/h. (a) If the robot launches itself at an angle of 60° at this speed, what is its maximum height? (b) What would the launch angle have to be to reach a height of 46 cm?

Solution:

$$a. v_y^2 = v_{0y}^2 - 2gy \Rightarrow y = 2.9 \text{ m/s}$$

$$y = 3.3 \text{ m/s}$$

$$y = \frac{v_{0y}^2}{2g} = \frac{(v_0 \sin \theta)^2}{2g} \Rightarrow \sin \theta = 0.91 \Rightarrow \theta = 65.5^\circ$$

Exercise:**Problem:**

Mt. Asama, Japan, is an active volcano. In 2009, an eruption threw solid volcanic rocks that landed 1 km horizontally from the crater. If the volcanic rocks were launched at an angle of 40° with respect to the horizontal and landed 900 m below the crater, (a) what would be their initial velocity and (b) what is their time of flight?

Exercise:**Problem:**

Drew Brees of the New Orleans Saints can throw a football 23.0 m/s (50 mph). If he angles the throw at 10° from the horizontal, what distance does it go if it is to be caught at the same elevation as it was thrown?

Solution:

$$R = 18.5 \text{ m}$$

Exercise:**Problem:**

The Lunar Roving Vehicle used in NASA's late *Apollo* missions reached an unofficial lunar land speed of 5.0 m/s by astronaut Eugene Cernan. If the rover was moving at this speed on a flat lunar surface and hit a small bump that projected it off the surface at an angle of 20° , how long would it be "airborne" on the Moon?

Exercise:**Problem:**

A soccer goal is 2.44 m high. A player kicks the ball at a distance 10 m from the goal at an angle of 25° . The ball hits the crossbar at the top of the goal. What is the initial speed of the soccer ball?

Solution:

$$y = (\tan \theta_0)x - \left[\frac{g}{2(v_0 \cos \theta_0)^2} \right] x^2 \Rightarrow v_0 = 16.4 \text{ m/s}$$

Exercise:**Problem:**

Olympus Mons on Mars is the largest volcano in the solar system, at a height of 25 km and with a radius of 312 km. If you are standing on the summit, with what initial velocity would you have to fire a projectile from a cannon horizontally to clear the volcano and land on the surface of Mars? Note that Mars has an acceleration of gravity of 3.7 m/s^2 .

Exercise:**Problem:**

In 1999, Robbie Knievel was the first to jump the Grand Canyon on a motorcycle. At a narrow part of the canyon (69.0 m wide) and traveling 35.8 m/s off the takeoff ramp, he reached the other side. What was his launch angle?

Solution:

$$R = \frac{v_0^2 \sin 2\theta_0}{g} \Rightarrow \theta_0 = 15.9^\circ$$

Exercise:**Problem:**

You throw a baseball at an initial speed of 15.0 m/s at an angle of 30° with respect to the horizontal. What would the ball's initial speed have to be at 30° on a planet that has twice the acceleration of gravity as Earth to achieve the same range? Consider launch and impact on a horizontal surface.

Exercise:**Problem:**

Aaron Rodgers throws a football at 20.0 m/s to his wide receiver, who is running straight down the field at 9.4 m/s. If Aaron throws the football when the wide receiver is 10.0 m in front of him, what angle does Aaron have to launch the ball at so the receiver catches it 20.0 m in front of Aaron?

Solution:

It takes the wide receiver 1.1 s to cover the last 10 m of his run.

$$T_{\text{tof}} = \frac{2(v_0 \sin \theta)}{g} \Rightarrow \sin \theta = 0.27 \Rightarrow \theta = 15.6^\circ$$

Glossary

projectile motion

motion of an object subject only to the acceleration of gravity

range

maximum horizontal distance a projectile travels

time of flight

elapsed time a projectile is in the air

trajectory

path of a projectile through the air

Uniform Circular Motion

By the end of this section, you will be able to:

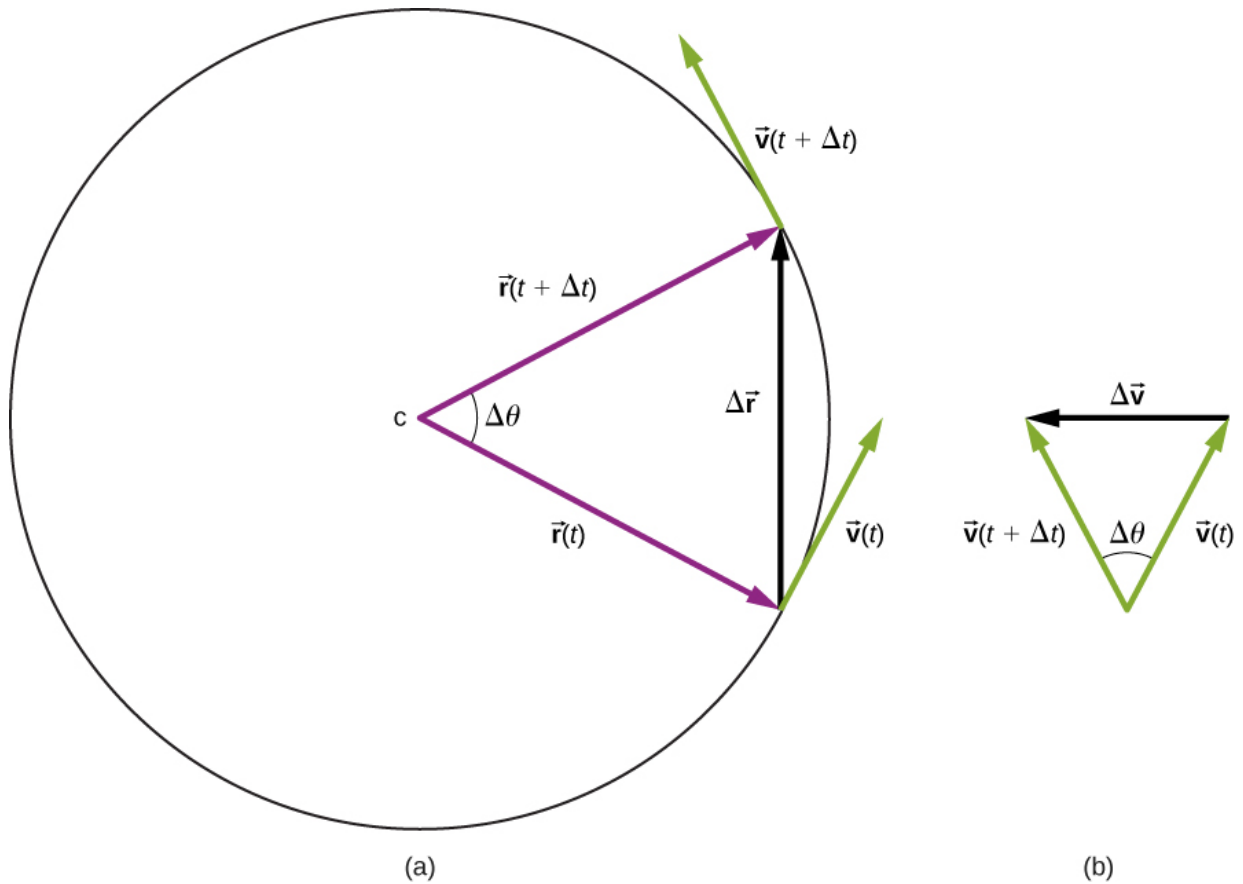
- Solve for the centripetal acceleration of an object moving on a circular path.
- Use the equations of circular motion to find the position, velocity, and acceleration of a particle executing circular motion.
- Explain the differences between centripetal acceleration and tangential acceleration resulting from nonuniform circular motion.
- Evaluate centripetal and tangential acceleration in nonuniform circular motion, and find the total acceleration vector.

Uniform circular motion is a specific type of motion in which an object travels in a circle with a constant speed. For example, any point on a propeller spinning at a constant rate is executing uniform circular motion. Other examples are the second, minute, and hour hands of a watch. It is remarkable that points on these rotating objects are actually accelerating, although the rotation rate is a constant. To see this, we must analyze the motion in terms of vectors.

Centripetal Acceleration

In one-dimensional kinematics, objects with a constant speed have zero acceleration. However, in two- and three-dimensional kinematics, even if the speed is a constant, a particle can have acceleration if it moves along a curved trajectory such as a circle. In this case the velocity vector is changing, or $d\vec{v}/dt \neq 0$. This is shown in [\[link\]](#). As the particle moves counterclockwise in time Δt on the circular path, its position vector moves from $\vec{r}(t)$ to $\vec{r}(t + \Delta t)$. The velocity vector has constant magnitude and is tangent to the path as it changes from $\vec{v}(t)$ to $\vec{v}(t + \Delta t)$, changing its direction only. Since the velocity vector $\vec{v}(t)$ is perpendicular to the position vector $\vec{r}(t)$, the triangles formed by the position vectors and $\Delta\vec{r}$, and the velocity vectors and $\Delta\vec{v}$ are similar. Furthermore, since $|\vec{r}(t)| = |\vec{r}(t + \Delta t)|$ and $|\vec{v}(t)| = |\vec{v}(t + \Delta t)|$, the two triangles are isosceles. From these facts we can make the assertion

$$\frac{\Delta v}{v} = \frac{\Delta r}{r} \text{ or } \Delta v = \frac{v}{r} \Delta r.$$



(a) A particle is moving in a circle at a constant speed, with position and velocity vectors at times t and $t + \Delta t$. (b) Velocity vectors forming a triangle. The two triangles in the figure are similar. The vector $\Delta \vec{v}$ points toward the center of the circle in the limit $\Delta t \rightarrow 0$.

We can find the magnitude of the acceleration from

Equation:

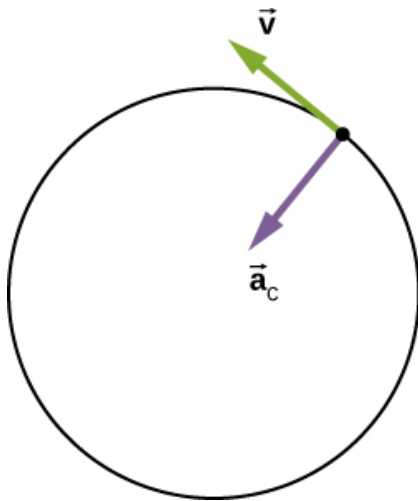
$$a = \lim_{\Delta t \rightarrow 0} \left(\frac{\Delta v}{\Delta t} \right) = \frac{v}{r} \left(\lim_{\Delta t \rightarrow 0} \frac{\Delta r}{\Delta t} \right) = \frac{v^2}{r}.$$

The direction of the acceleration can also be found by noting that as Δt and therefore $\Delta \theta$ approach zero, the vector $\Delta \vec{v}$ approaches a direction perpendicular to \vec{v} . In the limit $\Delta t \rightarrow 0$, $\Delta \vec{v}$ is perpendicular to \vec{v} . Since \vec{v} is tangent to the circle, the acceleration $d\vec{v}/dt$ points toward the center of the circle. Summarizing, a particle moving in a circle at a constant speed has an acceleration with magnitude

Note:
Equation:

$$a_c = \frac{v^2}{r}.$$

The direction of the acceleration vector is toward the center of the circle ([link](#)). This is a radial acceleration and is called the **centripetal acceleration**, which is why we give it the subscript c. The word *centripetal* comes from the Latin words *centrum* (meaning “center”) and *petere* (meaning “to seek”), and thus takes the meaning “center seeking.”



The centripetal acceleration vector points toward the center of the circular path of motion and is an acceleration in the radial direction. The velocity vector is also shown and is tangent to the circle.

Let's investigate some examples that illustrate the relative magnitudes of the velocity, radius, and centripetal acceleration.

Example:**Creating an Acceleration of 1 g**

A jet is flying at 134.1 m/s along a straight line and makes a turn along a circular path level with the ground. What does the radius of the circle have to be to produce a centripetal acceleration of 1 g on the pilot and jet toward the center of the circular trajectory?

Strategy

Given the speed of the jet, we can solve for the radius of the circle in the expression for the centripetal acceleration.

Solution

Set the centripetal acceleration equal to the acceleration of gravity: $9.8 \text{ m/s}^2 = v^2/r$.

Solving for the radius, we find

Equation:

$$r = \frac{(134.1 \text{ m/s})^2}{9.8 \text{ m/s}^2} = 1835 \text{ m} = 1.835 \text{ km}.$$

Significance

To create a greater acceleration than g on the pilot, the jet would either have to decrease the radius of its circular trajectory or increase its speed on its existing trajectory or both.

Note:**Exercise:****Problem:**

Check Your Understanding A flywheel has a radius of 20.0 cm. What is the speed of a point on the edge of the flywheel if it experiences a centripetal acceleration of 900.0 cm/s^2 ?

Solution:

134.0 cm/s

Centripetal acceleration can have a wide range of values, depending on the speed and radius of curvature of the circular path. Typical centripetal accelerations are given in the following table.

Object	Centripetal Acceleration (m/s ² or factors of <i>g</i>)
Earth around the Sun	5.93×10^{-3}
Moon around the Earth	2.73×10^{-3}
Satellite in geosynchronous orbit	0.233
Outer edge of a CD when playing	5.78
Jet in a barrel roll	(2–3 <i>g</i>)
Roller coaster	(5 <i>g</i>)
Electron orbiting a proton in a simple Bohr model of the atom	9.0×10^{22}

Typical Centripetal Accelerations

Equations of Motion for Uniform Circular Motion

A particle executing circular motion can be described by its position vector $\vec{r}(t)$. [\[link\]](#) shows a particle executing circular motion in a counterclockwise direction. As the particle moves on the circle, its position vector sweeps out the angle θ with the *x*-axis. Vector $\vec{r}(t)$ making an angle θ with the *x*-axis is shown with its components along the *x*- and *y*-axes. The magnitude of the position vector is $A = |\vec{r}(t)|$ and is also the radius of the circle, so that in terms of its components,

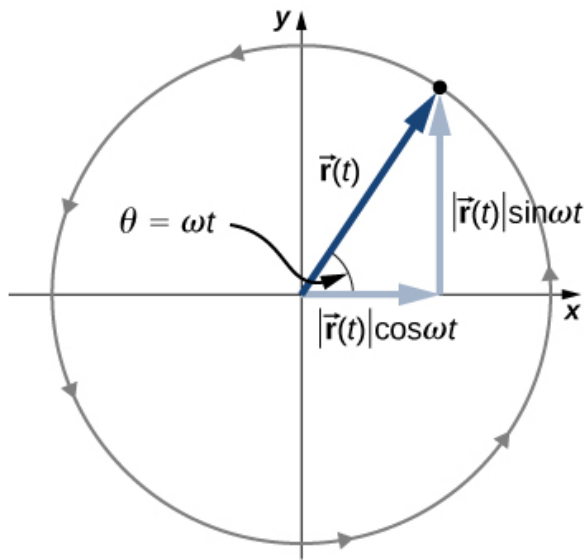
Note:
Equation:

$$\vec{r}(t) = A \cos \omega t \hat{i} + A \sin \omega t \hat{j}.$$

Here, ω is a constant called the **angular frequency** of the particle. The angular frequency has units of radians (rad) per second and is simply the number of radians of angular measure through which the particle passes per second. The angle θ that the position vector has at any particular time is ωt .

If T is the period of motion, or the time to complete one revolution (2π rad), then
Equation:

$$\omega = \frac{2\pi}{T}.$$



The position vector for a particle in circular motion with its components along the x - and y -axes. The particle moves counterclockwise. Angle θ is the angular frequency ω in radians per second multiplied by t .

Velocity and acceleration can be obtained from the position function by differentiation:

Note:

Equation:

$$\vec{v}(t) = \frac{d\vec{r}(t)}{dt} = -A\omega \sin \omega t \hat{i} + A\omega \cos \omega t \hat{j}.$$

It can be shown from [\[link\]](#) that the velocity vector is tangential to the circle at the location of the particle, with magnitude $A\omega$. Similarly, the acceleration vector is found by differentiating the velocity:

Note:

Equation:

$$\vec{a}(t) = \frac{d\vec{v}(t)}{dt} = -A\omega^2 \cos \omega t \hat{i} - A\omega^2 \sin \omega t \hat{j}.$$

From this equation we see that the acceleration vector has magnitude $A\omega^2$ and is directed opposite the position vector, toward the origin, because $\vec{a}(t) = -\omega^2 \vec{r}(t)$.

Example:

Circular Motion of a Proton

A proton has speed 5×10^6 m/s and is moving in a circle in the xy plane of radius $r = 0.175$ m. What is its position in the xy plane at time $t = 2.0 \times 10^{-7}$ s = 200 ns? At $t = 0$, the position of the proton is $0.175 \hat{i}$ m and it circles counterclockwise. Sketch the trajectory.

Solution

From the given data, the proton has period and angular frequency:

Equation:

$$T = \frac{2\pi r}{v} = \frac{2\pi(0.175 \text{ m})}{5.0 \times 10^6 \text{ m/s}} = 2.20 \times 10^{-7} \text{ s}$$

Equation:

$$\omega = \frac{2\pi}{T} = \frac{2\pi}{2.20 \times 10^{-7} \text{ s}} = 2.856 \times 10^7 \text{ rad/s}.$$

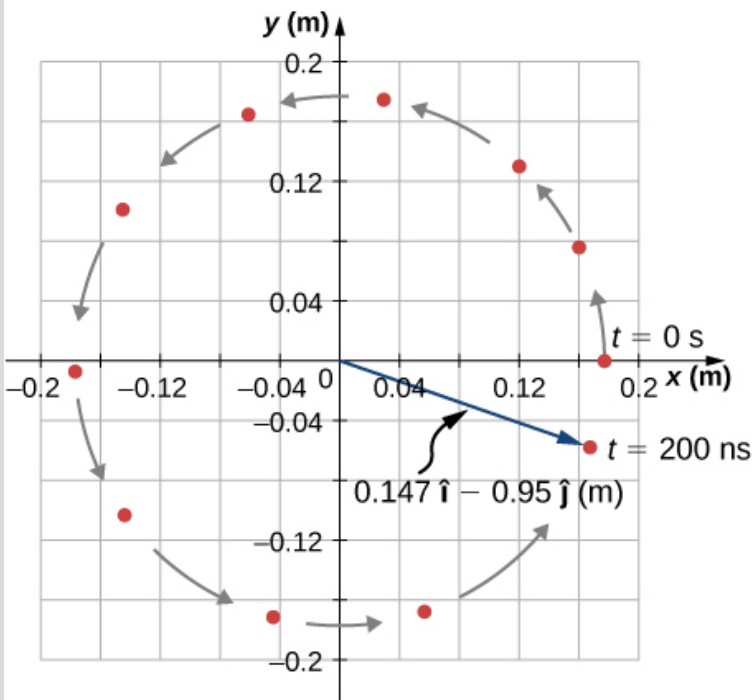
The position of the particle at $t = 2.0 \times 10^{-7}$ s with $A = 0.175$ m is

Equation:

$$\begin{aligned} \vec{r}(2.0 \times 10^{-7} \text{ s}) &= A \cos \omega(2.0 \times 10^{-7} \text{ s}) \hat{i} + A \sin \omega(2.0 \times 10^{-7} \text{ s}) \hat{j} \text{ m} \\ &= 0.175 \cos[(2.856 \times 10^7 \text{ rad/s})(2.0 \times 10^{-7} \text{ s})] \hat{i} \\ &\quad + 0.175 \sin[(2.856 \times 10^7 \text{ rad/s})(2.0 \times 10^{-7} \text{ s})] \hat{j} \text{ m} \\ &= 0.175 \cos(5.712 \text{ rad}) \hat{i} + 0.175 \sin(5.712 \text{ rad}) \hat{j} = 0.147 \hat{i} - 0.095 \hat{j} \text{ m}. \end{aligned}$$

From this result we see that the proton is located slightly below the x -axis. This is shown in [\[link\]](#).

Position Vector at $t = 200 \text{ ns}$



Position vector of the proton at $t = 2.0 \times 10^{-7} \text{ s} = 200 \text{ ns}$. The trajectory of the proton is shown. The angle through which the proton travels along the circle is 5.712 rad , which is a little less than one complete revolution.

Significance

We picked the initial position of the particle to be on the x -axis. This was completely arbitrary. If a different starting position were given, we would have a different final position at $t = 200 \text{ ns}$.

Nonuniform Circular Motion

Circular motion does not have to be at a constant speed. A particle can travel in a circle and speed up or slow down, showing an acceleration in the direction of the motion.

In uniform circular motion, the particle executing circular motion has a constant speed and the circle is at a fixed radius. If the speed of the particle is changing as well, then we

introduce an additional acceleration in the direction tangential to the circle. Such accelerations occur at a point on a top that is changing its spin rate, or any accelerating rotor. In [Displacement and Velocity Vectors](#) we showed that centripetal acceleration is the time rate of change of the direction of the velocity vector. If the speed of the particle is changing, then it has a **tangential acceleration** that is the time rate of change of the magnitude of the velocity:

Note:

Equation:

$$a_T = \frac{d|\vec{v}|}{dt}.$$

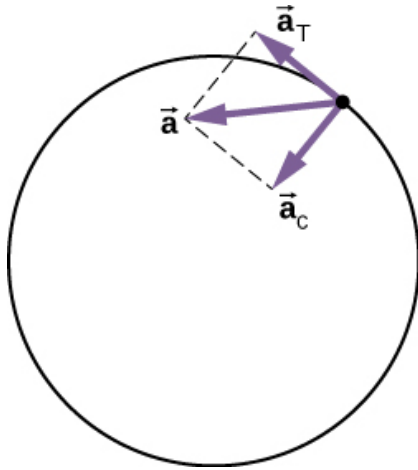
The direction of tangential acceleration is tangent to the circle whereas the direction of centripetal acceleration is radially inward toward the center of the circle. Thus, a particle in circular motion with a tangential acceleration has a **total acceleration** that is the vector sum of the centripetal and tangential accelerations:

Note:

Equation:

$$\vec{a} = \vec{a}_c + \vec{a}_T.$$

The acceleration vectors are shown in [\[link\]](#). Note that the two acceleration vectors \vec{a}_c and \vec{a}_T are perpendicular to each other, with \vec{a}_c in the radial direction and \vec{a}_T in the tangential direction. The total acceleration \vec{a} points at an angle between \vec{a}_c and \vec{a}_T .



The centripetal acceleration points toward the center of the circle. The tangential acceleration is tangential to the circle at the particle's position. The total acceleration is the vector sum of the tangential and centripetal accelerations, which are perpendicular.

Example:

Total Acceleration during Circular Motion

A particle moves in a circle of radius $r = 2.0$ m. During the time interval from $t = 1.5$ s to $t = 4.0$ s its speed varies with time according to

Equation:

$$v(t) = c_1 - \frac{c_2}{t^2}, \quad c_1 = 4.0 \text{ m/s}, \quad c_2 = 6.0 \text{ m} \cdot \text{s}.$$

What is the total acceleration of the particle at $t = 2.0$ s?

Strategy

We are given the speed of the particle and the radius of the circle, so we can calculate centripetal acceleration easily. The direction of the centripetal acceleration is toward the center of the circle. We find the magnitude of the tangential acceleration by taking the

derivative with respect to time of $|v(t)|$ using [\[link\]](#) and evaluating it at $t = 2.0$ s. We use this and the magnitude of the centripetal acceleration to find the total acceleration.

Solution

Centripetal acceleration is

Equation:

$$v(2.0\text{s}) = \left(4.0 - \frac{6.0}{(2.0)^2} \right) \text{m/s} = 2.5 \text{ m/s}$$

Equation:

$$a_c = \frac{v^2}{r} = \frac{(2.5 \text{ m/s})^2}{2.0 \text{ m}} = 3.1 \text{ m/s}^2$$

directed toward the center of the circle. Tangential acceleration is

Equation:

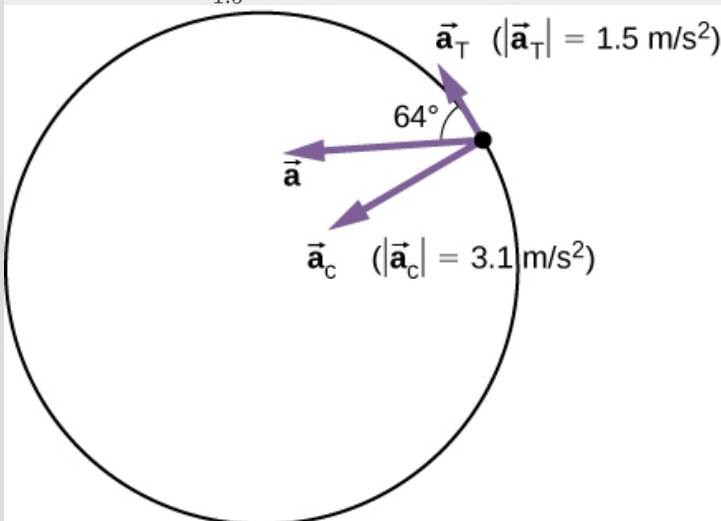
$$a_T = \left| \frac{d\vec{v}}{dt} \right| = \frac{2c_2}{t^3} = \frac{12.0}{(2.0)^3} \text{m/s}^2 = 1.5 \text{ m/s}^2.$$

Total acceleration is

Equation:

$$|\vec{a}| = \sqrt{3.1^2 + 1.5^2} \text{m/s}^2 = 3.44 \text{ m/s}^2$$

and $\theta = \tan^{-1} \frac{3.1}{1.5} = 64^\circ$ from the tangent to the circle. See [\[link\]](#).



The tangential and centripetal acceleration vectors. The net acceleration \vec{a} is the vector sum of the two accelerations.

Significance

The directions of centripetal and tangential accelerations can be described more conveniently in terms of a polar coordinate system, with unit vectors in the radial and tangential directions. This coordinate system, which is used for motion along curved paths, is discussed in detail later in the book.

Summary

- Uniform circular motion is motion in a circle at constant speed.
- Centripetal acceleration \vec{a}_C is the acceleration a particle must have to follow a circular path. Centripetal acceleration always points toward the center of rotation and has magnitude $a_C = v^2/r$.
- Nonuniform circular motion occurs when there is tangential acceleration of an object executing circular motion such that the speed of the object is changing. This acceleration is called tangential acceleration \vec{a}_T . The magnitude of tangential acceleration is the time rate of change of the magnitude of the velocity. The tangential acceleration vector is tangential to the circle, whereas the centripetal acceleration vector points radially inward toward the center of the circle. The total acceleration is the vector sum of tangential and centripetal accelerations.
- An object executing uniform circular motion can be described with equations of motion. The position vector of the object is $\vec{r}(t) = A \cos \omega t \hat{i} + A \sin \omega t \hat{j}$, where A is the magnitude $|\vec{r}(t)|$, which is also the radius of the circle, and ω is the angular frequency.

Conceptual Questions

Exercise:

Problem:

Can centripetal acceleration change the speed of a particle undergoing circular motion?

Exercise:

Problem:

Can tangential acceleration change the speed of a particle undergoing circular motion?

Solution:

yes

Problems

Exercise:

Problem:

A flywheel is rotating at 30 rev/s. What is the total angle, in radians, through which a point on the flywheel rotates in 40 s?

Exercise:

Problem:

A particle travels in a circle of radius 10 m at a constant speed of 20 m/s. What is the magnitude of the acceleration?

Solution:

$$a_C = 40 \text{ m/s}^2$$

Exercise:

Problem:

Cam Newton of the Carolina Panthers throws a perfect football spiral at 8.0 rev/s. The radius of a pro football is 8.5 cm at the middle of the short side. What is the centripetal acceleration of the laces on the football?

Exercise:

Problem:

A fairground ride spins its occupants inside a flying saucer-shaped container. If the horizontal circular path the riders follow has an 8.00-m radius, at how many revolutions per minute are the riders subjected to a centripetal acceleration equal to that of gravity?

Solution:

$$a_C = \frac{v^2}{r} \Rightarrow v^2 = r a_C = 78.4, \quad v = 8.85 \text{ m/s}$$
$$T = 5.68 \text{ s, which is } 0.176 \text{ rev/s} = 10.6 \text{ rev/min}$$

Exercise:

Problem:

A runner taking part in the 200-m dash must run around the end of a track that has a circular arc with a radius of curvature of 30.0 m. The runner starts the race at a constant speed. If she completes the 200-m dash in 23.2 s and runs at constant speed throughout the race, what is her centripetal acceleration as she runs the curved portion of the track?

Exercise:

Problem: What is the acceleration of Venus toward the Sun, assuming a circular orbit?

Solution:

Venus is 108.2 million km from the Sun and has an orbital period of 0.6152 y.

$$r = 1.082 \times 10^{11} \text{ m} \quad T = 1.94 \times 10^7 \text{ s}$$

$$v = 3.5 \times 10^4 \text{ m/s}, \quad a_C = 1.135 \times 10^{-2} \text{ m/s}^2$$

Exercise:

Problem:

An experimental jet rocket travels around Earth along its equator just above its surface. At what speed must the jet travel if the magnitude of its acceleration is g ?

Exercise:

Problem:

A fan is rotating at a constant 360.0 rev/min. What is the magnitude of the acceleration of a point on one of its blades 10.0 cm from the axis of rotation?

Solution:

$$360 \text{ rev/min} = 6 \text{ rev/s}$$

$$v = 3.8 \text{ m/s} \quad a_C = 144. \text{ m/s}^2$$

Exercise:

Problem:

A point located on the second hand of a large clock has a radial acceleration of 0.1 cm/s^2 . How far is the point from the axis of rotation of the second hand?

Glossary

angular frequency

ω , rate of change of an angle with which an object that is moving on a circular path

centripetal acceleration

component of acceleration of an object moving in a circle that is directed radially inward toward the center of the circle

tangential acceleration

magnitude of which is the time rate of change of speed. Its direction is tangent to the circle.

total acceleration

vector sum of centripetal and tangential accelerations

Relative Motion in One and Two Dimensions

By the end of this section, you will be able to:

- Explain the concept of reference frames.
- Write the position and velocity vector equations for relative motion.
- Draw the position and velocity vectors for relative motion.
- Analyze one-dimensional and two-dimensional relative motion problems using the position and velocity vector equations.

Motion does not happen in isolation. If you're riding in a train moving at 10 m/s east, this velocity is measured relative to the ground on which you're traveling. However, if another train passes you at 15 m/s east, your velocity relative to this other train is different from your velocity relative to the ground. Your velocity relative to the other train is 5 m/s west. To explore this idea further, we first need to establish some terminology.

Reference Frames

To discuss relative motion in one or more dimensions, we first introduce the concept of **reference frames**. When we say an object has a certain velocity, we must state it has a velocity with respect to a given reference frame. In most examples we have examined so far, this reference frame has been Earth. If you say a person is sitting in a train moving at 10 m/s east, then you imply the person on the train is moving relative to the surface of Earth at this velocity, and Earth is the reference frame. We can expand our view of the motion of the person on the train and say Earth is spinning in its orbit around the Sun, in which case the motion becomes more complicated. In this case, the solar system is the reference frame. In summary, all discussion of relative motion must define the reference frames involved. We now develop a method to refer to reference frames in relative motion.

Relative Motion in One Dimension

We introduce relative motion in one dimension first, because the velocity vectors simplify to having only two possible directions. Take the example of the person sitting in a train moving east. If we choose east as the positive direction and Earth as the reference frame, then we can write the velocity of the train with respect to the Earth as $\vec{v}_{TE} = 10 \text{ m/s } \hat{i}$ east, where the subscripts TE refer to train and Earth. Let's now say the person gets up out of her seat and walks toward the back of the train at 2 m/s. This tells us she has a velocity relative to the reference frame of the train. Since the person is walking west, in the negative direction, we write her velocity with respect to the train as $\vec{v}_{PT} = -2 \text{ m/s } \hat{i}$. We can add the two velocity vectors to find the velocity of the person with respect to Earth. This **relative velocity** is written as

Equation:

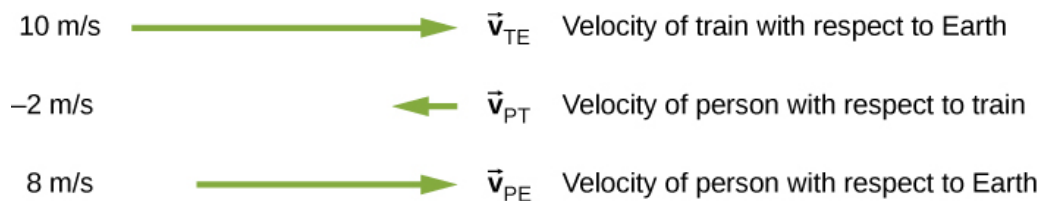
$$\vec{v}_{PE} = \vec{v}_{PT} + \vec{v}_{TE}.$$

Note the ordering of the subscripts for the various reference frames in [\[link\]](#). The subscripts for the coupling reference frame, which is the train, appear consecutively in the right-hand side of the equation. [\[link\]](#) shows the correct order of subscripts when forming the vector equation.

$$\vec{v}_{PE} = \underbrace{\vec{v}_{PT} + \vec{v}_{TE}}$$

When constructing the vector equation, the subscripts for the coupling reference frame appear consecutively on the inside. The subscripts on the left-hand side of the equation are the same as the two outside subscripts on the right-hand side of the equation.

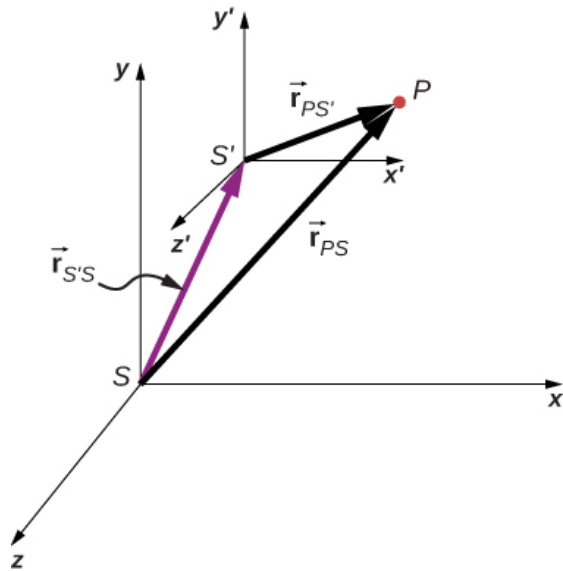
Adding the vectors, we find $\vec{v}_{PE} = 8 \text{ m/s } \hat{i}$, so the person is moving 8 m/s east with respect to Earth. Graphically, this is shown in [\[link\]](#).



Velocity vectors of the train with respect to Earth, person with respect to the train, and person with respect to Earth.

Relative Velocity in Two Dimensions

We can now apply these concepts to describing motion in two dimensions. Consider a particle P and reference frames S and S' , as shown in [\[link\]](#). The position of the origin of S' as measured in S is $\vec{r}_{S'S}$, the position of P as measured in S' is $\vec{r}_{PS'}$, and the position of P as measured in S is \vec{r}_{PS} .



The positions of particle P relative to frames S and S' are \vec{r}_{PS} and $\vec{r}_{PS'}$, respectively.

From [\[link\]](#) we see that

Note:

Equation:

$$\vec{r}_{PS} = \vec{r}_{PS'} + \vec{r}_{S'S}.$$

The relative velocities are the time derivatives of the position vectors. Therefore,

Note:

Equation:

$$\vec{v}_{PS} = \vec{v}_{PS'} + \vec{v}_{S'S}.$$

The velocity of a particle relative to S is equal to its velocity relative to S' plus the velocity of S' relative to S .

We can extend [\[link\]](#) to any number of reference frames. For particle P with velocities \vec{v}_{PA} , \vec{v}_{PB} , and \vec{v}_{PC} in frames A , B , and C ,

Note:

Equation:

$$\vec{v}_{PC} = \vec{v}_{PA} + \vec{v}_{AB} + \vec{v}_{BC}.$$

We can also see how the accelerations are related as observed in two reference frames by differentiating [\[link\]](#):

Note:

Equation:

$$\vec{a}_{PS} = \vec{a}_{PS'} + \vec{a}_{S'S}.$$

We see that if the velocity of S' relative to S is a constant, then $\vec{a}_{S'S} = 0$ and

Equation:

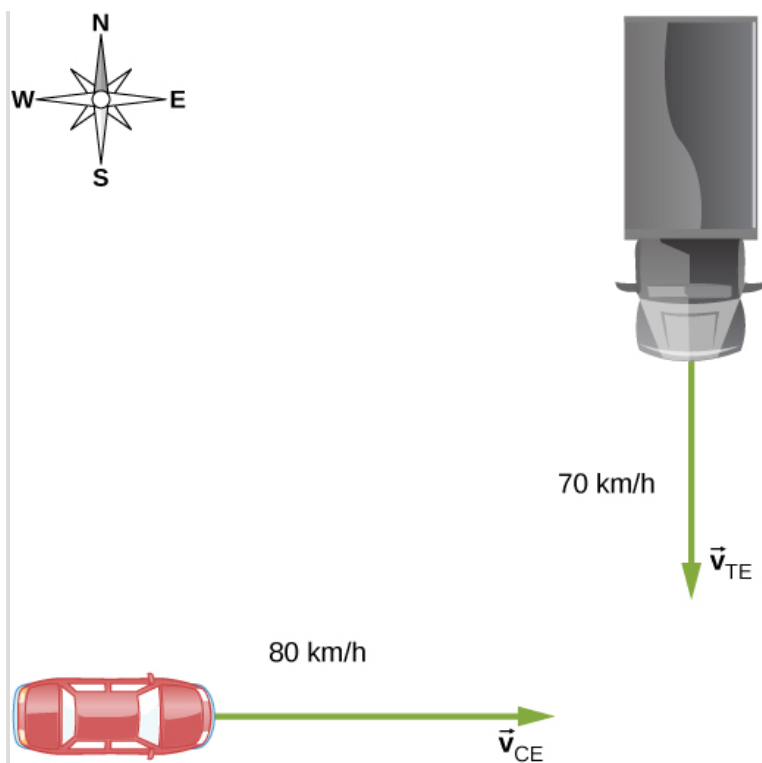
$$\vec{a}_{PS} = \vec{a}_{PS'}.$$

This says the acceleration of a particle is the same as measured by two observers moving at a constant velocity relative to each other.

Example:

Motion of a Car Relative to a Truck

A truck is traveling south at a speed of 70 km/h toward an intersection. A car is traveling east toward the intersection at a speed of 80 km/h ([\[link\]](#)). What is the velocity of the car relative to the truck?



A car travels east toward an intersection while a truck travels south toward the same intersection.

Strategy

First, we must establish the reference frame common to both vehicles, which is Earth. Then, we write the velocities of each with respect to the reference frame of Earth, which enables us to form a vector equation that links the car, the truck, and Earth to solve for the velocity of the car with respect to the truck.

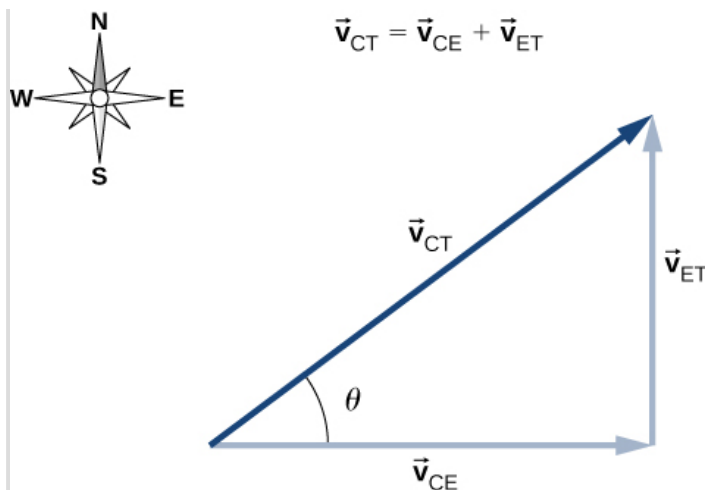
Solution

The velocity of the car with respect to Earth is $\vec{v}_{CE} = 80 \text{ km/h } \hat{i}$. The velocity of the truck with respect to Earth is $\vec{v}_{TE} = -70 \text{ km/h } \hat{j}$. Using the velocity addition rule, the relative motion equation we are seeking is

Equation:

$$\vec{v}_{CT} = \vec{v}_{CE} + \vec{v}_{ET}.$$

Here, \vec{v}_{CT} is the velocity of the car with respect to the truck, and Earth is the connecting reference frame. Since we have the velocity of the truck with respect to Earth, the negative of this vector is the velocity of Earth with respect to the truck: $\vec{v}_{ET} = -\vec{v}_{TE}$. The vector diagram of this equation is shown in [\[link\]](#).



Vector diagram of the vector equation

$$\vec{v}_{CT} = \vec{v}_{CE} + \vec{v}_{ET}.$$

We can now solve for the velocity of the car with respect to the truck:

Equation:

$$|\vec{v}_{CT}| = \sqrt{(80.0 \text{ km/h})^2 + (70.0 \text{ km/h})^2} = 106. \text{ km/h}$$

and

Equation:

$$\theta = \tan^{-1} \left(\frac{70.0}{80.0} \right) = 41.2^\circ \text{ north of east.}$$

Significance

Drawing a vector diagram showing the velocity vectors can help in understanding the relative velocity of the two objects.

Note:

Exercise:

Problem:

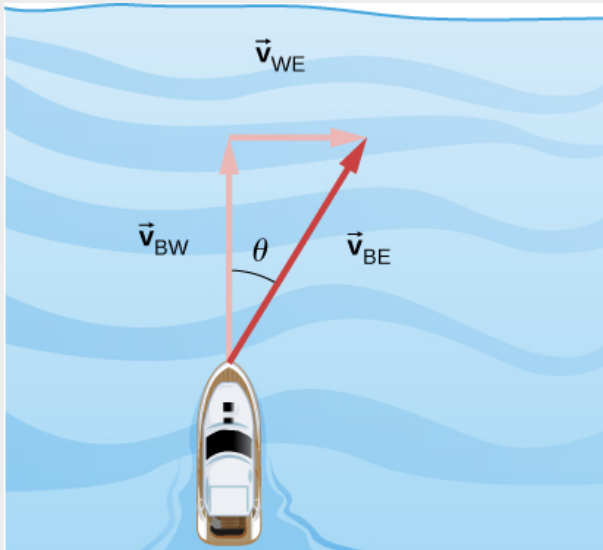
Check Your Understanding A boat heads north in still water at 4.5 m/s directly across a river that is running east at 3.0 m/s. What is the velocity of the boat with respect to Earth?

Solution:

Labeling subscripts for the vector equation, we have B = boat, R = river, and E = Earth. The vector equation becomes $\vec{v}_{BE} = \vec{v}_{BR} + \vec{v}_{RE}$. We have right triangle geometry shown in Figure 04_05_BoatRiv_img. Solving for \vec{v}_{BE} , we have

$$v_{BE} = \sqrt{v_{BR}^2 + v_{RE}^2} = \sqrt{4.5^2 + 3.0^2}$$

$$v_{BE} = 5.4 \text{ m/s}, \quad \theta = \tan^{-1} \left(\frac{3.0}{4.5} \right) = 33.7^\circ.$$



Example:

Flying a Plane in a Wind

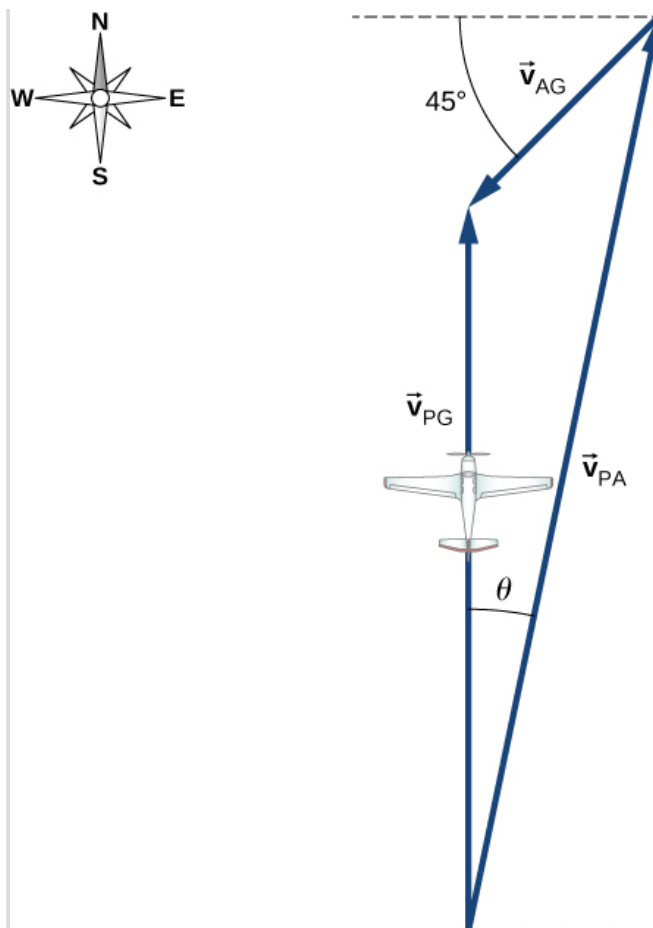
A pilot must fly his plane due north to reach his destination. The plane can fly at 300 km/h in still air. A wind is blowing out of the northeast at 90 km/h. (a) What is the speed of the plane relative to the ground? (b) In what direction must the pilot head her plane to fly due north?

Strategy

The pilot must point her plane somewhat east of north to compensate for the wind velocity. We need to construct a vector equation that contains the velocity of the plane with respect to the ground, the velocity of the plane with respect to the air, and the velocity of the air with respect to the ground. Since these last two quantities are known, we can solve for the velocity of the plane with respect to the ground. We can graph the vectors and use this diagram to evaluate the magnitude of the plane's velocity with respect to the ground. The diagram will also tell us the angle the plane's velocity makes with north with respect to the air, which is the direction the pilot must head her plane.

Solution

The vector equation is $\vec{v}_{PG} = \vec{v}_{PA} + \vec{v}_{AG}$, where P = plane, A = air, and G = ground. From the geometry in [\[link\]](#), we can solve easily for the magnitude of the velocity of the plane with respect to the ground and the angle of the plane's heading, θ .



Vector diagram for [\[link\]](#) showing the vectors \vec{v}_{PA} , \vec{v}_{AG} , and \vec{v}_{PG} .

(a) Known quantities:

Equation:

$$|\vec{v}_{PA}| = 300 \text{ km/h}$$

Equation:

$$|\vec{v}_{AG}| = 90 \text{ km/h}$$

Substituting into the equation of motion, we obtain $|\vec{v}_{PG}| = 230 \text{ km/h}$.

(b) The angle $\theta = \tan^{-1} \frac{63.64}{300} = 12^\circ$ east of north.

Summary

- When analyzing motion of an object, the reference frame in terms of position, velocity, and acceleration needs to be specified.

- Relative velocity is the velocity of an object as observed from a particular reference frame, and it varies with the choice of reference frame.
- If S and S' are two reference frames moving relative to each other at a constant velocity, then the velocity of an object relative to S is equal to its velocity relative to S' plus the velocity of S' relative to S .
- If two reference frames are moving relative to each other at a constant velocity, then the accelerations of an object as observed in both reference frames are equal.

Key Equations

Position vector	$\vec{\mathbf{r}}(t) = x(t)\hat{\mathbf{i}} + y(t)\hat{\mathbf{j}} + z(t)\hat{\mathbf{k}}$
Displacement vector	$\Delta\vec{\mathbf{r}} = \vec{\mathbf{r}}(t_2) - \vec{\mathbf{r}}(t_1)$
Velocity vector	$\vec{\mathbf{v}}(t) = \lim_{\Delta t \rightarrow 0} \frac{\vec{\mathbf{r}}(t+\Delta t) - \vec{\mathbf{r}}(t)}{\Delta t} = \frac{d\vec{\mathbf{r}}}{dt}$
Velocity in terms of components	$\vec{\mathbf{v}}(t) = v_x(t)\hat{\mathbf{i}} + v_y(t)\hat{\mathbf{j}} + v_z(t)\hat{\mathbf{k}}$
Velocity components	$v_x(t) = \frac{dx(t)}{dt} \quad v_y(t) = \frac{dy(t)}{dt} \quad v_z(t) = \frac{dz(t)}{dt}$
Average velocity	$\vec{\mathbf{v}}_{\text{avg}} = \frac{\vec{\mathbf{r}}(t_2) - \vec{\mathbf{r}}(t_1)}{t_2 - t_1}$
Instantaneous acceleration	$\vec{\mathbf{a}}(t) = \lim_{\Delta t \rightarrow 0} \frac{\vec{\mathbf{v}}(t+\Delta t) - \vec{\mathbf{v}}(t)}{\Delta t} = \frac{d\vec{\mathbf{v}}(t)}{dt}$
Instantaneous acceleration, component form	$\vec{\mathbf{a}}(t) = \frac{dv_x(t)}{dt}\hat{\mathbf{i}} + \frac{dv_y(t)}{dt}\hat{\mathbf{j}} + \frac{dv_z(t)}{dt}\hat{\mathbf{k}}$
Instantaneous acceleration as second derivatives of position	$\vec{\mathbf{a}}(t) = \frac{d^2x(t)}{dt^2}\hat{\mathbf{i}} + \frac{d^2y(t)}{dt^2}\hat{\mathbf{j}} + \frac{d^2z(t)}{dt^2}\hat{\mathbf{k}}$
Time of flight	$T_{\text{tof}} = \frac{2(v_0 \sin \theta_0)}{g}$
Trajectory	$y = (\tan \theta_0)x - \left[\frac{g}{2(v_0 \cos \theta_0)^2} \right] x^2$
Range	$R = \frac{v_0^2 \sin 2\theta_0}{g}$
Centripetal acceleration	$a_C = \frac{v^2}{r}$
Position vector, uniform circular motion	$\vec{\mathbf{r}}(t) = A \cos \omega t \hat{\mathbf{i}} + A \sin \omega t \hat{\mathbf{j}}$
Velocity vector, uniform circular motion	

	$\vec{v}(t) = \frac{d\vec{r}(t)}{dt} = -A\omega \sin \omega t \hat{i} + A\omega \cos \omega t \hat{j}$
Acceleration vector, uniform circular motion	$\vec{a}(t) = \frac{d\vec{v}(t)}{dt} = -A\omega^2 \cos \omega t \hat{i} - A\omega^2 \sin \omega t \hat{j}$
Tangential acceleration	$a_T = \frac{d \vec{v} }{dt}$
Total acceleration	$\vec{a} = \vec{a}_C + \vec{a}_T$
Position vector in frame S is the position vector in frame S' plus the vector from the origin of S to the origin of S'	$\vec{r}_{PS} = \vec{r}_{PS'} + \vec{r}_{S'S}$
Relative velocity equation connecting two reference frames	$\vec{v}_{PS} = \vec{v}_{PS'} + \vec{v}_{S'S}$
Relative velocity equation connecting more than two reference frames	$\vec{v}_{PC} = \vec{v}_{PA} + \vec{v}_{AB} + \vec{v}_{BC}$
Relative acceleration equation	$\vec{a}_{PS} = \vec{a}_{PS'} + \vec{a}_{S'S}$

Conceptual Questions

Exercise:

Problem:

What frame or frames of reference do you use instinctively when driving a car? When flying in a commercial jet?

Exercise:

Problem:

A basketball player dribbling down the court usually keeps his eyes fixed on the players around him. He is moving fast. Why doesn't he need to keep his eyes on the ball?

Solution:

If he is going to pass the ball to another player, he needs to keep his eyes on the reference frame in which the other players on the team are located.

Exercise:

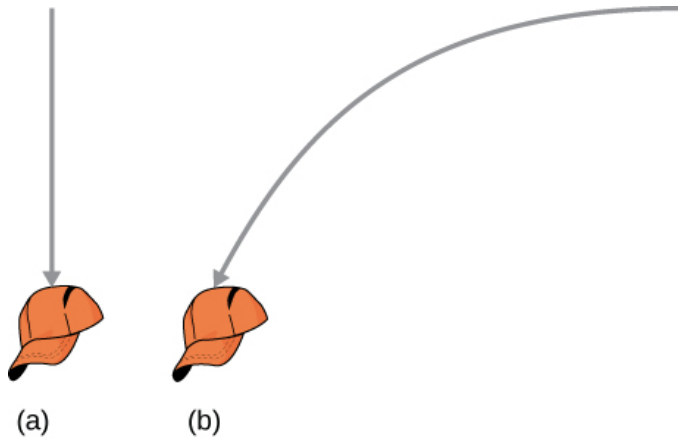
Problem:

If someone is riding in the back of a pickup truck and throws a softball straight backward, is it possible for the ball to fall straight down as viewed by a person standing at the side of the road? Under what condition would this occur? How would the motion of the ball appear to the person who threw it?

Exercise:

Problem:

The hat of a jogger running at constant velocity falls off the back of his head. Draw a sketch showing the path of the hat in the jogger's frame of reference. Draw its path as viewed by a stationary observer. Neglect air resistance.

Solution:**Exercise:****Problem:**

A clod of dirt falls from the bed of a moving truck. It strikes the ground directly below the end of the truck. (a) What is the direction of its velocity relative to the truck just before it hits? (b) Is this the same as the direction of its velocity relative to ground just before it hits? Explain your answers.

Problems**Exercise:****Problem:**

The coordinate axes of the reference frame S' remain parallel to those of S , as S' moves away from S at a constant velocity $\vec{v}_{S'}^S = (4.0\hat{i} + 3.0\hat{j} + 5.0\hat{k})$ m/s. (a) If at time $t = 0$ the origins coincide, what is the position of the origin O' in the S frame as a function of time? (b) How is particle position for $\vec{r}(t)$ and $\vec{r}'(t)$, as measured in S and S' , respectively, related? (c) What is the relationship between particle velocities $\vec{v}(t)$ and $\vec{v}'(t)$? (d) How are accelerations $\vec{a}(t)$ and $\vec{a}'(t)$ related?

Solution:

- $O'(t) = (4.0\hat{i} + 3.0\hat{j} + 5.0\hat{k})t$ m,
- $\vec{r}_{PS} = \vec{r}_{PS'} + \vec{r}_{S'S}$, $\vec{r}(t) = \vec{r}'(t) + (4.0\hat{i} + 3.0\hat{j} + 5.0\hat{k})t$ m,
- $\vec{v}(t) = \vec{v}'(t) + (4.0\hat{i} + 3.0\hat{j} + 5.0\hat{k})$ m/s, d. The accelerations are the same.

Exercise:

Problem:

The coordinate axes of the reference frame S' remain parallel to those of S , as S' moves away from S at a constant velocity $\vec{v}_{S'S} = (1.0\hat{i} + 2.0\hat{j} + 3.0\hat{k})t$ m/s. (a) If at time $t = 0$ the origins coincide, what is the position of origin O' in the S frame as a function of time? (b) How is particle position for $\vec{r}(t)$ and $\vec{r}'(t)$, as measured in S and S' , respectively, related? (c) What is the relationship between particle velocities $\vec{v}(t)$ and $\vec{v}'(t)$? (d) How are accelerations $\vec{a}(t)$ and $\vec{a}'(t)$ related?

Exercise:**Problem:**

The velocity of a particle in reference frame A is $(2.0\hat{i} + 3.0\hat{j})$ m/s. The velocity of reference frame A with respect to reference frame B is $4.0\hat{k}$ m/s, and the velocity of reference frame B with respect to C is $2.0\hat{j}$ m/s. What is the velocity of the particle in reference frame C ?

Solution:

$$\vec{v}_{PC} = (2.0\hat{i} + 5.0\hat{j} + 4.0\hat{k})\text{m/s}$$

Exercise:**Problem:**

Raindrops fall vertically at 4.5 m/s relative to the earth. What does an observer in a car moving at 22.0 m/s in a straight line measure as the velocity of the raindrops?

Exercise:**Problem:**

A seagull can fly at a velocity of 9.00 m/s in still air. (a) If it takes the bird 20.0 min to travel 6.00 km straight into an oncoming wind, what is the velocity of the wind? (b) If the bird turns around and flies with the wind, how long will it take the bird to return 6.00 km?

Solution:

a. A = air, S = seagull, G = ground

$\vec{v}_{SA} = 9.0$ m/s velocity of seagull with respect to still air

$$\vec{v}_{AG} = ? \quad \vec{v}_{SG} = 5 \text{ m/s} \quad \vec{v}_{SG} = \vec{v}_{SA} + \vec{v}_{AG} \Rightarrow \vec{v}_{AG} = \vec{v}_{SG} - \vec{v}_{SA}$$

$$\vec{v}_{AG} = -4.0 \text{ m/s}$$

$$\text{b. } \vec{v}_{SG} = \vec{v}_{SA} + \vec{v}_{AG} \Rightarrow \vec{v}_{SG} = -13.0 \text{ m/s}$$

$$\frac{-6000 \text{ m}}{-13.0 \text{ m/s}} = 7 \text{ min } 42 \text{ s}$$

Exercise:**Problem:**

A ship sets sail from Rotterdam, heading due north at 7.00 m/s relative to the water. The local ocean current is 1.50 m/s in a direction 40.0° north of east. What is the velocity of the ship relative to Earth?

Exercise:

Problem:

A boat can be rowed at 8.0 km/h in still water. (a) How much time is required to row 1.5 km downstream in a river moving 3.0 km/h relative to the shore? (b) How much time is required for the return trip? (c) In what direction must the boat be aimed to row straight across the river? (d) Suppose the river is 0.8 km wide. What is the velocity of the boat with respect to Earth and how much time is required to get to the opposite shore? (e) Suppose, instead, the boat is aimed straight across the river. How much time is required to get across and how far downstream is the boat when it reaches the opposite shore?

Solution:

Take the positive direction to be the same direction that the river is flowing, which is east. S = shore/Earth, W = water, and B = boat.

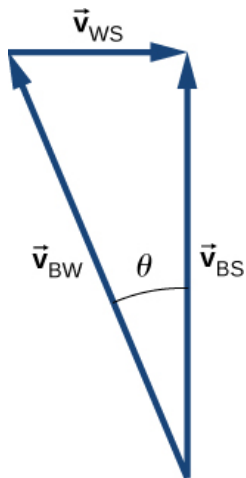
a. $\vec{v}_{BS} = 11 \text{ km/h}$

$t = 8.2 \text{ min}$

b. $\vec{v}_{BS} = -5 \text{ km/h}$

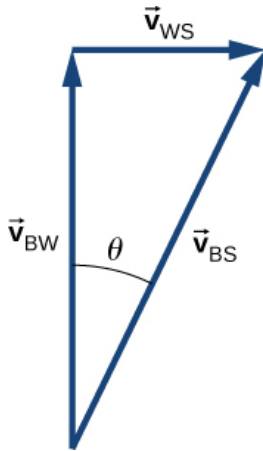
$t = 18 \text{ min}$

c. $\vec{v}_{BS} = \vec{v}_{BW} + \vec{v}_{WS}$ $\theta = 22^\circ$ west of north



d. $|\vec{v}_{BS}| = 7.4 \text{ km/h}$ $t = 6.5 \text{ min}$

e. $\vec{v}_{BS} = 8.54 \text{ km/h}$, but only the component of the velocity straight across the river is used to get the time



$$t = 6.0 \text{ min}$$

$$\text{Downstream} = 0.3 \text{ km}$$

Exercise:

Problem:

A small plane flies at 200 km/h in still air. If the wind blows directly out of the west at 50 km/h, (a) in what direction must the pilot head her plane to move directly north across land and (b) how long does it take her to reach a point 300 km directly north of her starting point?

Exercise:

Problem:

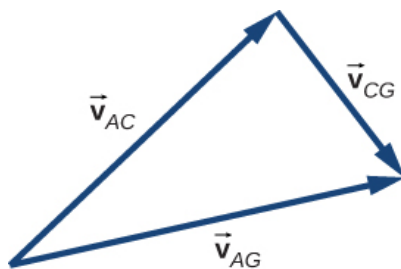
A cyclist traveling southeast along a road at 15 km/h feels a wind blowing from the southwest at 25 km/h. To a stationary observer, what are the speed and direction of the wind?

Solution:

$$\vec{v}_{AG} = \vec{v}_{AC} + \vec{v}_{CG}$$

$$|\vec{v}_{AC}| = 25 \text{ km/h} \quad |\vec{v}_{CG}| = 15 \text{ km/h} \quad |\vec{v}_{AG}| = 29.15 \text{ km/h} \quad \vec{v}_{AG} = \vec{v}_{AC} + \vec{v}_{CG}$$

The angle between \vec{v}_{AC} and \vec{v}_{AG} is 31° , so the direction of the wind is 14° north of east.



Exercise:

Problem:

A river is moving east at 4 m/s. A boat starts from the dock heading 30° north of west at 7 m/s. If the river is 1800 m wide, (a) what is the velocity of the boat with respect to Earth and (b) how long does it take the boat to cross the river?

Additional Problems**Exercise:****Problem:**

A Formula One race car is traveling at 89.0 m/s along a straight track enters a turn on the race track with radius of curvature of 200.0 m. What centripetal acceleration must the car have to stay on the track?

Solution:

$$a_C = 39.6 \text{ m/s}^2$$

Exercise:**Problem:**

A particle travels in a circular orbit of radius 10 m. Its speed is changing at a rate of 15.0 m/s^2 at an instant when its speed is 40.0 m/s. What is the magnitude of the acceleration of the particle?

Exercise:**Problem:**

The driver of a car moving at 90.0 km/h presses down on the brake as the car enters a circular curve of radius 150.0 m. If the speed of the car is decreasing at a rate of 9.0 km/h each second, what is the magnitude of the acceleration of the car at the instant its speed is 60.0 km/h?

Solution:

$$90.0 \text{ km/h} = 25.0 \text{ m/s}, \quad 9.0 \text{ km/h} = 2.5 \text{ m/s}, \quad 60.0 \text{ km/h} = 16.7 \text{ m/s}$$
$$a_T = -2.5 \text{ m/s}^2, \quad a_C = 1.86 \text{ m/s}^2, \quad a = 3.1 \text{ m/s}^2$$

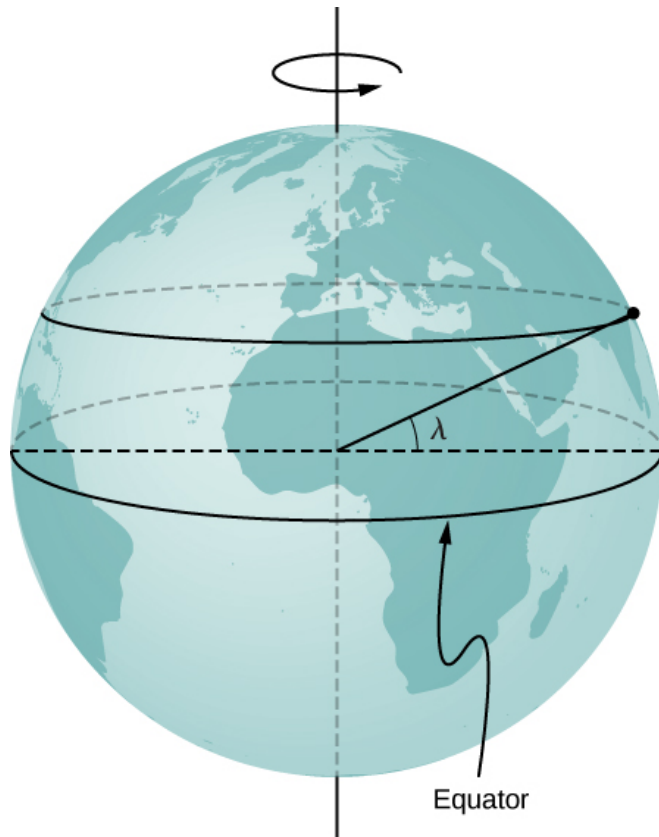
Exercise:**Problem:**

A race car entering the curved part of the track at the Daytona 500 drops its speed from 85.0 m/s to 80.0 m/s in 2.0 s. If the radius of the curved part of the track is 316.0 m, calculate the total acceleration of the race car at the beginning and ending of reduction of speed.

Exercise:

Problem:

An elephant is located on Earth's surface at a latitude λ . Calculate the centripetal acceleration of the elephant resulting from the rotation of Earth around its polar axis. Express your answer in terms of λ , the radius R_E of Earth, and time T for one rotation of Earth. Compare your answer with g for $\lambda = 40^\circ$.



Solution:

The radius of the circle of revolution at latitude λ is $R_E \cos \lambda$. The velocity of the body is $\frac{2\pi r}{T}$. $a_C = \frac{4\pi^2 R_E \cos \lambda}{T^2}$ for $\lambda = 40^\circ$, $a_C = 0.26\% g$

Exercise:**Problem:**

A proton in a synchrotron is moving in a circle of radius 1 km and increasing its speed by $v(t) = c_1 + c_2 t^2$, where $c_1 = 2.0 \times 10^5 \text{ m/s}$, $c_2 = 10^5 \text{ m/s}^3$. (a) What is the proton's total acceleration at $t = 5.0 \text{ s}$? (b) At what time does the expression for the velocity become unphysical?

Exercise:

Problem:

A propeller blade at rest starts to rotate from $t = 0$ s to $t = 5.0$ s with a tangential acceleration of the tip of the blade at 3.00 m/s^2 . The tip of the blade is 1.5 m from the axis of rotation. At $t = 5.0$ s, what is the total acceleration of the tip of the blade?

Solution:

$$a_T = 3.00 \text{ m/s}^2$$

$$v(5 \text{ s}) = 15.00 \text{ m/s} \quad a_C = 150.00 \text{ m/s}^2 \quad \theta = 88.8^\circ \text{ with respect to the tangent to the circle of revolution directed inward. } |\vec{a}| = 150.03 \text{ m/s}^2$$

Exercise:**Problem:**

A particle is executing circular motion with a constant angular frequency of $\omega = 4.00 \text{ rad/s}$. If time $t = 0$ corresponds to the position of the particle being located at $y = 0$ m and $x = 5$ m, (a) what is the position of the particle at $t = 10$ s? (b) What is its velocity at this time? (c) What is its acceleration?

Exercise:**Problem:**

A particle's centripetal acceleration is $a_C = 4.0 \text{ m/s}^2$ at $t = 0$ s where it is on the x -axis and moving counterclockwise in the xy plane. It is executing uniform circular motion about an axis at a distance of 5.0 m. What is its velocity at $t = 10$ s?

Solution:

$$\vec{a}(t) = -A\omega^2 \cos \omega t \hat{i} - A\omega^2 \sin \omega t \hat{j}$$

$$a_C = 5.0 \text{ m}\omega^2 \quad \omega = 0.89 \text{ rad/s}$$

$$\vec{v}(t) = -2.24 \text{ m/s} \hat{i} - 3.87 \text{ m/s} \hat{j}$$

Exercise:**Problem:**

A rod 3.0 m in length is rotating at 2.0 rev/s about an axis at one end. Compare the centripetal accelerations at radii of (a) 1.0 m, (b) 2.0 m, and (c) 3.0 m.

Exercise:**Problem:**

A particle located initially at $(1.5\hat{j} + 4.0\hat{k})\text{m}$ undergoes a displacement of $(2.5\hat{i} + 3.2\hat{j} - 1.2\hat{k})\text{m}$. What is the final position of the particle?

Solution:

$$\vec{r}_1 = 1.5\hat{j} + 4.0\hat{k} \quad \vec{r}_2 = \Delta\vec{r} + \vec{r}_1 = 2.5\hat{i} + 4.7\hat{j} + 2.8\hat{k}$$

Exercise:

Problem:

The position of a particle is given by $\vec{r}(t) = (50 \text{ m/s})t\hat{i} - (4.9 \text{ m/s}^2)t^2\hat{j}$. (a) What are the particle's velocity and acceleration as functions of time? (b) What are the initial conditions to produce the motion?

Exercise:**Problem:**

A spaceship is traveling at a constant velocity of $\vec{v}(t) = 250.0\hat{i}\text{m/s}$ when its rockets fire, giving it an acceleration of $\vec{a}(t) = (3.0\hat{i} + 4.0\hat{k})\text{m/s}^2$. What is its velocity 5 s after the rockets fire?

Solution:

$$v_x(t) = 265.0 \text{ m/s}$$

$$v_y(t) = 20.0 \text{ m/s}$$

$$\vec{v}(5.0 \text{ s}) = (265.0\hat{i} + 20.0\hat{j})\text{m/s}$$

Exercise:**Problem:**

A crossbow is aimed horizontally at a target 40 m away. The arrow hits 30 cm below the spot at which it was aimed. What is the initial velocity of the arrow?

Exercise:**Problem:**

A long jumper can jump a distance of 8.0 m when he takes off at an angle of 45° with respect to the horizontal. Assuming he can jump with the same initial speed at all angles, how much distance does he lose by taking off at 30° ?

Solution:

$$R = 1.07 \text{ m}$$

Exercise:**Problem:**

On planet Arcon, the maximum horizontal range of a projectile launched at 10 m/s is 20 m. What is the acceleration of gravity on this planet?

Exercise:**Problem:**

A mountain biker encounters a jump on a race course that sends him into the air at 60° to the horizontal. If he lands at a horizontal distance of 45.0 m and 20 m below his launch point, what is his initial speed?

Solution:

$$v_0 = 20.1 \text{ m/s}$$

Exercise:**Problem:**

Which has the greater centripetal acceleration, a car with a speed of 15.0 m/s along a circular track of radius 100.0 m or a car with a speed of 12.0 m/s along a circular track of radius 75.0 m?

Exercise:**Problem:**

A geosynchronous satellite orbits Earth at a distance of 42,250.0 km and has a period of 1 day. What is the centripetal acceleration of the satellite?

Solution:

$$v = 3072.5 \text{ m/s}$$

$$a_C = 0.223 \text{ m/s}^2$$

Exercise:**Problem:**

Two speedboats are traveling at the same speed relative to the water in opposite directions in a moving river. An observer on the riverbank sees the boats moving at 4.0 m/s and 5.0 m/s. (a) What is the speed of the boats relative to the river? (b) How fast is the river moving relative to the shore?

Challenge Problems**Exercise:****Problem:**

World's Longest Par 3. The tee of the world's longest par 3 sits atop South Africa's Hanglip Mountain at 400.0 m above the green and can only be reached by helicopter. The horizontal distance to the green is 359.0 m. Neglect air resistance and answer the following questions. (a) If a golfer launches a shot that is 40° with respect to the horizontal, what initial velocity must she give the ball? (b) What is the time to reach the green?

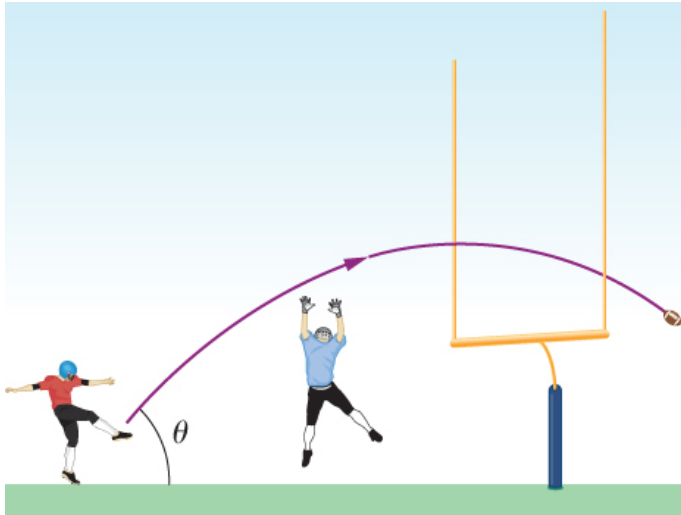
Solution:

$$\begin{aligned} \text{a. } -400.0 \text{ m} &= v_{0y}t - 4.9t^2 & 359.0 \text{ m} &= v_{0x}t & t &= \frac{359.0}{v_{0x}} & -400.0 &= 359.0 \frac{v_{0y}}{v_{0x}} - 4.9 \left(\frac{359.0}{v_{0x}} \right)^2 \\ -400.0 &= 359.0 \tan 40^\circ - \frac{631,516.9}{v_{0x}^2} & \Rightarrow v_{0x}^2 &= 900.6 & v_{0x} &= 30.0 \text{ m/s} & v_{0y} &= v_{0x} \tan 40^\circ = 25.2 \text{ m/s} \\ v &= 39.2 \text{ m/s, b. } t &= 12.0 \text{ s} \end{aligned}$$

Exercise:

Problem:

When a field goal kicker kicks a football as hard as he can at 45° to the horizontal, the ball just clears the 3-m-high crossbar of the goalposts 45.7 m away. (a) What is the maximum speed the kicker can impart to the football? (b) In addition to clearing the crossbar, the football must be high enough in the air early during its flight to clear the reach of the onrushing defensive lineman. If the lineman is 4.6 m away and has a vertical reach of 2.5 m, can he block the 45.7-m field goal attempt? (c) What if the lineman is 1.0 m away?

**Exercise:****Problem:**

A truck is traveling east at 80 km/h. At an intersection 32 km ahead, a car is traveling north at 50 km/h. (a) How long after this moment will the vehicles be closest to each other? (b) How far apart will they be at that point?

Solution:

$$\begin{aligned} \text{a. } \vec{r}_{TC} &= (-32 + 80t)\hat{i} + 50t\hat{j}, \quad |\vec{r}_{TC}|^2 = (-32 + 80t)^2 + (50t)^2 \\ 2r \frac{dr}{dt} &= 2(-32 + 80t)(80) + 5000t \frac{dr}{dt} = \frac{160(-32 + 80t) + 5000t}{2r} = 0 \\ 17800t &= 5184 \Rightarrow t = 0.29 \text{ hr,} \\ \text{b. } |\vec{r}_{TC}| &= 17 \text{ km} \end{aligned}$$

Glossary**reference frame**

coordinate system in which the position, velocity, and acceleration of an object at rest or moving is measured

relative velocity

velocity of an object as observed from a particular reference frame, or the velocity of one reference frame with respect to another reference frame

Introduction

class="introduction"

The Golden Gate Bridge, one of the greatest works of modern engineering, was the longest suspension bridge in the world in the year it opened, 1937. It is still among the 10 longest suspension bridges as of this writing. In designing and building a bridge, what physics must we consider? What forces act on the bridge?

What forces
keep the
bridge from
falling?
How do the
towers,
cables, and
ground
interact to
maintain
stability?



When you drive across a bridge, you expect it to remain stable. You also expect to speed up or slow your car in response to traffic changes. In both cases, you deal with forces. The forces on the bridge are in equilibrium, so it stays in place. In contrast, the force produced by your car engine causes a change in motion. Isaac Newton discovered the laws of motion that describe these situations.

Forces affect every moment of your life. Your body is held to Earth by force and held together by the forces of charged particles. When you open a door, walk down a street, lift your fork, or touch a baby's face, you are applying forces. Zooming in deeper, your body's atoms are held together by electrical

forces, and the core of the atom, called the nucleus, is held together by the strongest force we know—strong nuclear force.

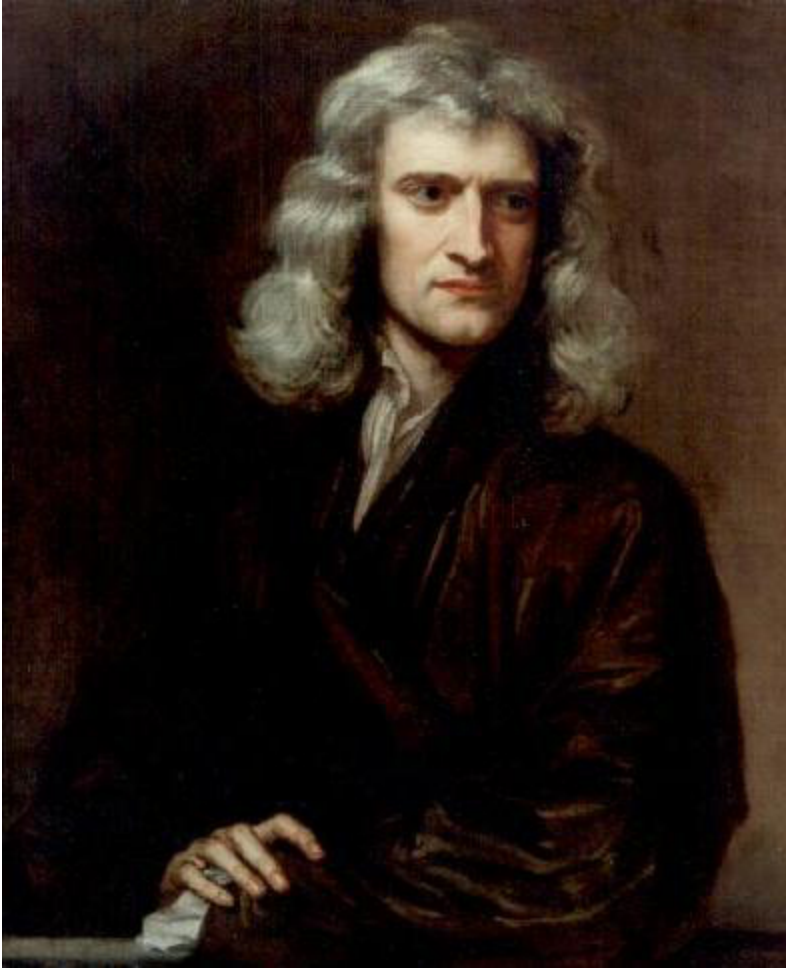
Forces

By the end of the section, you will be able to:

- Distinguish between kinematics and dynamics
- Understand the definition of force
- Identify simple free-body diagrams
- Define the SI unit of force, the newton
- Describe force as a vector

The study of motion is called *kinematics*, but kinematics only describes the way objects move—their velocity and their acceleration. **Dynamics** is the study of how forces affect the motion of objects and systems. It considers the causes of motion of objects and systems of interest, where a system is anything being analyzed. The foundation of dynamics are the laws of motion stated by Isaac Newton (1642–1727). These laws provide an example of the breadth and simplicity of principles under which nature functions. They are also universal laws in that they apply to situations on Earth and in space.

Newton's laws of motion were just one part of the monumental work that has made him legendary ([\[link\]](#)). The development of Newton's laws marks the transition from the Renaissance to the modern era. Not until the advent of modern physics was it discovered that Newton's laws produce a good description of motion only when the objects are moving at speeds much less than the speed of light and when those objects are larger than the size of most molecules (about 10^{-9} m in diameter). These constraints define the realm of Newtonian mechanics. At the beginning of the twentieth century, Albert Einstein (1879–1955) developed the theory of relativity and, along with many other scientists, quantum mechanics. Quantum mechanics does not have the constraints present in Newtonian physics. All of the situations we consider in this chapter, and all those preceding the introduction of relativity in [Relativity](#), are in the realm of Newtonian physics.

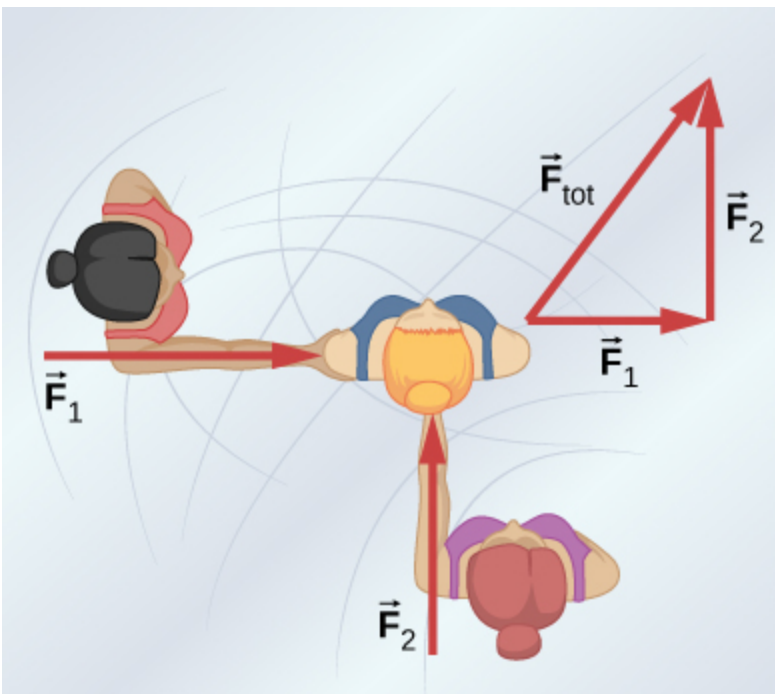


Isaac Newton (1642–1727) published his amazing work, *Philosophiae Naturalis Principia Mathematica*, in 1687. It proposed scientific laws that still apply today to describe the motion of objects (the laws of motion). Newton also discovered the law of gravity, invented calculus, and made great contributions to the theories of light and color.

Working Definition of Force

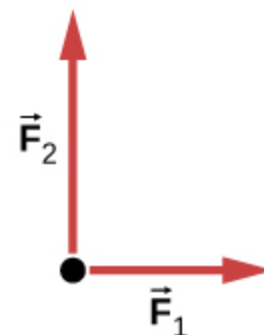
Dynamics is the study of the forces that cause objects and systems to move. To understand this, we need a working definition of force. An intuitive definition of **force**—that is, a push or a pull—is a good place to start. We know that a push or a pull has both magnitude and direction (therefore, it is a vector quantity), so we can define force as the push or pull on an object with a specific magnitude and direction. Force can be represented by vectors or expressed as a multiple of a standard force.

The push or pull on an object can vary considerably in either magnitude or direction. For example, a cannon exerts a strong force on a cannonball that is launched into the air. In contrast, Earth exerts only a tiny downward pull on a flea. Our everyday experiences also give us a good idea of how multiple forces add. If two people push in different directions on a third person, as illustrated in [\[link\]](#), we might expect the total force to be in the direction shown. Since force is a vector, it adds just like other vectors. Forces, like other vectors, are represented by arrows and can be added using the familiar head-to-tail method or trigonometric methods. These ideas were developed in [Vectors](#).



(a)

Free-body diagram



(b)

(a) An overhead view of two ice skaters pushing on a third skater. Forces are vectors and add like other vectors, so the total force on the third skater is in the direction shown. (b) A free-body diagram representing the forces acting on the third skater.

[\[link\]](#)(b) is our first example of a **free-body diagram**, which is a sketch showing all external forces acting on an object or system. The object or system is represented by a single isolated point (or free body), and only those forces acting *on* it that originate outside of the object or system—that is, **external forces**—are shown. (These forces are the only ones shown because only external forces acting on the free body affect its motion. We can ignore any internal forces within the body.) The forces are represented by vectors extending outward from the free body.

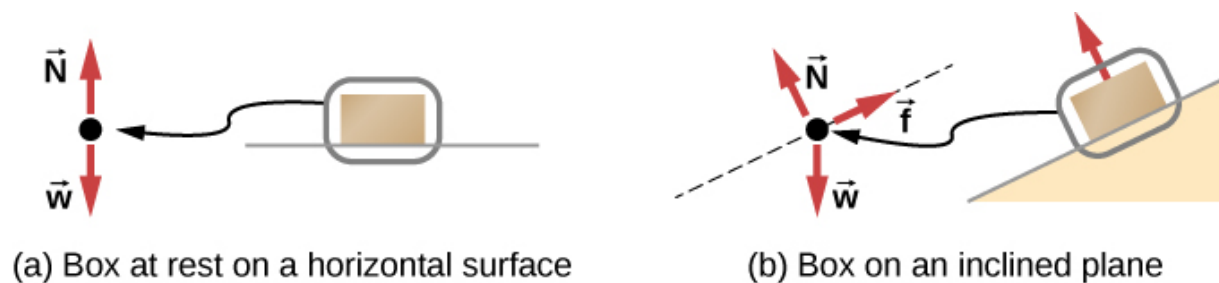
Free-body diagrams are useful in analyzing forces acting on an object or system, and are employed extensively in the study and application of Newton's laws of motion. You will see them throughout this text and in all your studies of physics. The following steps briefly explain how a free-body diagram is created; we examine this strategy in more detail in [Drawing Free-Body Diagrams](#).

Note:

Drawing Free-Body Diagrams

1. Draw the object under consideration. If you are treating the object as a particle, represent the object as a point. Place this point at the origin of an xy -coordinate system.
2. Include all forces that act on the object, representing these forces as vectors. However, do not include the net force on the object or the forces that the object exerts on its environment.
3. Resolve all force vectors into x - and y -components.
4. Draw a separate free-body diagram for each object in the problem.

We illustrate this strategy with two examples of free-body diagrams ([\[link\]](#)). The terms used in this figure are explained in more detail later in the chapter.

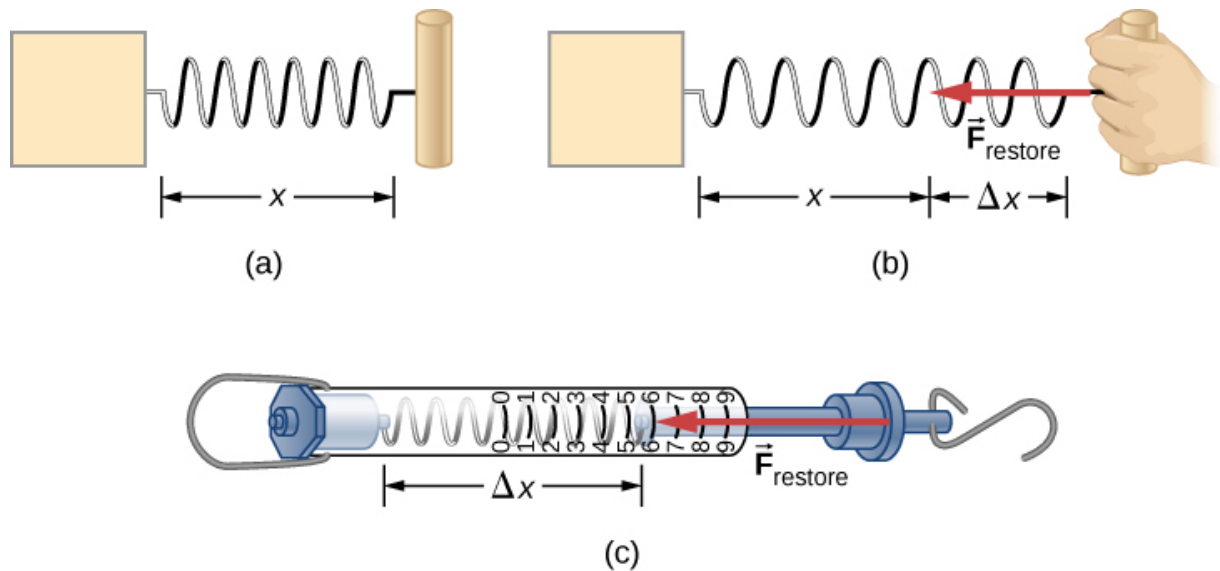


In these free-body diagrams, \vec{N} is the normal force, \vec{w} is the weight of the object, and \vec{f} is the friction.

The steps given here are sufficient to guide you in this important problem-solving strategy. The final section of this chapter explains in more detail how to draw free-body diagrams when working with the ideas presented in this chapter.

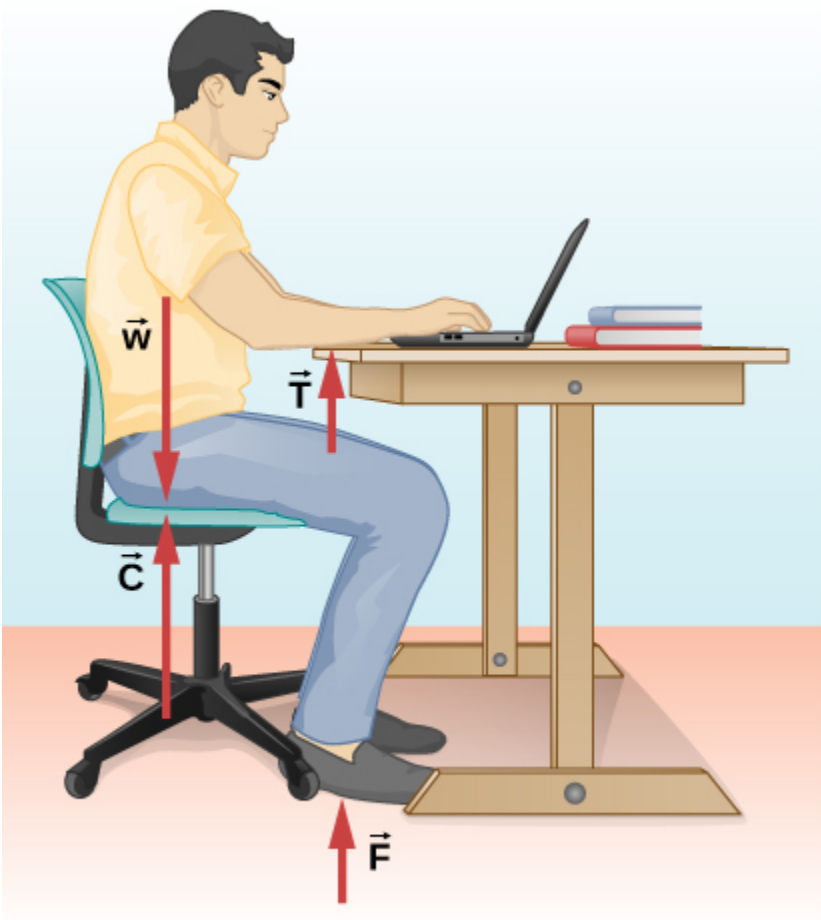
Development of the Force Concept

A quantitative definition of force can be based on some standard force, just as distance is measured in units relative to a standard length. One possibility is to stretch a spring a certain fixed distance ([\[link\]](#)) and use the force it exerts to pull itself back to its relaxed shape—called a *restoring force*—as a standard. The magnitude of all other forces can be considered as multiples of this standard unit of force. Many other possibilities exist for standard forces. Some alternative definitions of force will be given later in this chapter.



The force exerted by a stretched spring can be used as a standard unit of force. (a) This spring has a length x when undistorted. (b) When stretched a distance Δx , the spring exerts a restoring force \vec{F}_{restore} , which is reproducible. (c) A spring scale is one device that uses a spring to measure force. The force \vec{F}_{restore} is exerted on whatever is attached to the hook. Here, this force has a magnitude of six units of the force standard being employed.

Let's analyze force more deeply. Suppose a physics student sits at a table, working diligently on his homework ([\[link\]](#)). What external forces act on him? Can we determine the origin of these forces?



(a)



(b)

(a) The forces acting on the student are due to the chair, the table, the floor, and Earth's gravitational attraction. (b) In solving a problem involving the student, we may want to consider only the forces acting along the line running through his torso. A free-body diagram for this situation is shown.

In most situations, forces are grouped into two categories: *contact forces* and *field forces*. As you might guess, contact forces are due to direct physical contact between objects. For example, the student in [\[link\]](#) experiences the contact forces \vec{C} , \vec{F} , and \vec{T} , which are exerted by the chair on his posterior, the floor on his feet, and the table on his forearms,

respectively. Field forces, however, act without the necessity of physical contact between objects. They depend on the presence of a “field” in the region of space surrounding the body under consideration. Since the student is in Earth’s gravitational field, he feels a gravitational force \vec{w} ; in other words, he has weight.

You can think of a field as a property of space that is detectable by the forces it exerts. Scientists think there are only four fundamental force fields in nature. These are the gravitational, electromagnetic, strong nuclear, and weak fields (we consider these four forces in nature later in this text). As noted for \vec{w} in [\[link\]](#), the gravitational field is responsible for the weight of a body. The forces of the electromagnetic field include those of static electricity and magnetism; they are also responsible for the attraction among atoms in bulk matter. Both the strong nuclear and the weak force fields are effective only over distances roughly equal to a length of scale no larger than an atomic nucleus (10^{-15} m). Their range is so small that neither field has influence in the macroscopic world of Newtonian mechanics.

Contact forces are fundamentally electromagnetic. While the elbow of the student in [\[link\]](#) is in contact with the tabletop, the atomic charges in his skin interact electromagnetically with the charges in the surface of the table. The net (total) result is the force \vec{T} . Similarly, when adhesive tape sticks to a piece of paper, the atoms of the tape are intermingled with those of the paper to cause a net electromagnetic force between the two objects. However, in the context of Newtonian mechanics, the electromagnetic origin of contact forces is not an important concern.

Vector Notation for Force

As previously discussed, force is a vector; it has both magnitude and direction. The SI unit of force is called the **newton** (abbreviated N), and 1 N is the force needed to accelerate an object with a mass of 1 kg at a rate of 1 m/s^2 : $1 \text{ N} = 1 \text{ kg} \cdot \text{m/s}^2$. An easy way to remember the size of a newton is to imagine holding a small apple; it has a weight of about 1 N.

We can thus describe a two-dimensional force in the form $\vec{\mathbf{F}} = a\hat{\mathbf{i}} + b\hat{\mathbf{j}}$ (the unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ indicate the direction of these forces along the x-axis and the y-axis, respectively) and a three-dimensional force in the form $\vec{\mathbf{F}} = a\hat{\mathbf{i}} + b\hat{\mathbf{j}} + c\hat{\mathbf{k}}$. In [\[link\]](#), let's suppose that ice skater 1, on the left side of the figure, pushes horizontally with a force of 30.0 N to the right; we represent this as $\vec{\mathbf{F}}_1 = 30.0\hat{\mathbf{i}}$ N. Similarly, if ice skater 2 pushes with a force of 40.0 N in the positive vertical direction shown, we would write $\vec{\mathbf{F}}_2 = 40.0\hat{\mathbf{j}}$ N. The resultant of the two forces causes a mass to accelerate—in this case, the third ice skater. This resultant is called the **net external force** $\vec{\mathbf{F}}_{\text{net}}$ and is found by taking the vector sum of all external forces acting on an object or system (thus, we can also represent net external force as $\sum \vec{\mathbf{F}}$):

Note:
Equation:

$$\vec{\mathbf{F}}_{\text{net}} = \sum \vec{\mathbf{F}} = \vec{\mathbf{F}}_1 + \vec{\mathbf{F}}_2 + \cdots$$

This equation can be extended to any number of forces.

In this example, we have $\vec{\mathbf{F}}_{\text{net}} = \sum \vec{\mathbf{F}} = \vec{\mathbf{F}}_1 + \vec{\mathbf{F}}_2 = 30.0\hat{\mathbf{i}} + 40.0\hat{\mathbf{j}}$ N. The hypotenuse of the triangle shown in [\[link\]](#) is the resultant force, or net force. It is a vector. To find its magnitude (the size of the vector, without regard to direction), we use the rule given in [Vectors](#), taking the square root of the sum of the squares of the components:

Equation:

$$F_{\text{net}} = \sqrt{(30.0 \text{ N})^2 + (40.0 \text{ N})^2} = 50.0 \text{ N}.$$

The direction is given by

Equation:

$$\theta = \tan^{-1} \left(\frac{F_2}{F_1} \right) = \tan^{-1} \left(\frac{40.0}{30.0} \right) = 53.1^\circ,$$

measured from the positive x-axis, as shown in the free-body diagram in [\[link\]](#)(b).

Let's suppose the ice skaters now push the third ice skater with

$\vec{F}_1 = 3.0\hat{i} + 8.0\hat{j}$ N and $\vec{F}_2 = 5.0\hat{i} + 4.0\hat{j}$ N. What is the resultant of these two forces? We must recognize that force is a vector; therefore, we must add using the rules for vector addition:

Equation:

$$\vec{F}_{\text{net}} = \vec{F}_1 + \vec{F}_2 = (3.0\hat{i} + 8.0\hat{j}) + (5.0\hat{i} + 4.0\hat{j}) = 8.0\hat{i} + 12\hat{j} \text{ N}$$

Note:

Exercise:

Problem:

Check Your Understanding Find the magnitude and direction of the net force in the ice skater example just given.

Solution:

14 N, 56° measured from the positive x-axis

Note:

View this [interactive simulation](#) to learn how to add vectors. Drag vectors onto a graph, change their length and angle, and sum them together. The magnitude, angle, and components of each vector can be displayed in several formats.

Summary

- Dynamics is the study of how forces affect the motion of objects, whereas kinematics simply describes the way objects move.
- Force is a push or pull that can be defined in terms of various standards, and it is a vector that has both magnitude and direction.
- External forces are any outside forces that act on a body. A free-body diagram is a drawing of all external forces acting on a body.
- The SI unit of force is the newton (N).

Conceptual Questions

Exercise:

Problem:

What properties do forces have that allow us to classify them as vectors?

Solution:

Forces are directional and have magnitude.

Problems

Exercise:

Problem:

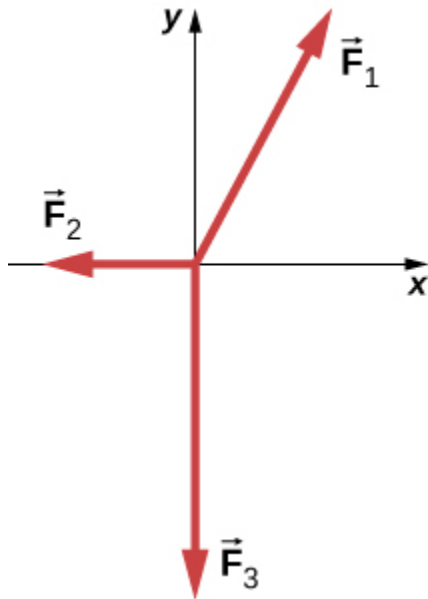
Two ropes are attached to a tree, and forces of $\vec{F}_1 = 2.0\hat{i} + 4.0\hat{j}$ N and $\vec{F}_2 = 3.0\hat{i} + 6.0\hat{j}$ N are applied. The forces are coplanar (in the same plane). (a) What is the resultant (net force) of these two force vectors? (b) Find the magnitude and direction of this net force.

Solution:

a. $\vec{F}_{\text{net}} = 5.0\hat{i} + 10.0\hat{j}$ N; b. the magnitude is $F_{\text{net}} = 11$ N, and the direction is $\theta = 63^\circ$

Exercise:**Problem:**

A telephone pole has three cables pulling as shown from above, with $\vec{F}_1 = (300.0\hat{i} + 500.0\hat{j})$, $\vec{F}_2 = -200.0\hat{i}$, and $\vec{F}_3 = -800.0\hat{j}$. (a) Find the net force on the telephone pole in component form. (b) Find the magnitude and direction of this net force.

**Exercise:**

Problem:

Two teenagers are pulling on ropes attached to a tree. The angle between the ropes is 30.0° . David pulls with a force of 400.0 N and Stephanie pulls with a force of 300.0 N. (a) Find the component form of the net force. (b) Find the magnitude of the resultant (net) force on the tree and the angle it makes with David's rope.

Solution:

a. $\vec{F}_{\text{net}} = 660.0\hat{i} + 150.0\hat{j}$ N; b. $F_{\text{net}} = 676.6$ N at $\theta = 12.8^\circ$ from David's rope

Glossary

dynamics

study of how forces affect the motion of objects and systems

external force

force acting on an object or system that originates outside of the object or system

force

push or pull on an object with a specific magnitude and direction; can be represented by vectors or expressed as a multiple of a standard force

free-body diagram

sketch showing all external forces acting on an object or system; the system is represented by a single isolated point, and the forces are represented by vectors extending outward from that point

net external force

vector sum of all external forces acting on an object or system; causes a mass to accelerate

newton

SI unit of force; 1 N is the force needed to accelerate an object with a mass of 1 kg at a rate of 1 m/s^2

Newton's First Law

By the end of the section, you will be able to:

- Describe Newton's first law of motion
- Recognize friction as an external force
- Define inertia
- Identify inertial reference frames
- Calculate equilibrium for a system

Experience suggests that an object at rest remains at rest if left alone and that an object in motion tends to slow down and stop unless some effort is made to keep it moving. However, **Newton's first law** gives a deeper explanation of this observation.

Note:

Newton's First Law of Motion

A body at rest remains at rest or, if in motion, remains in motion at constant velocity unless acted on by a net external force.

Note the repeated use of the verb “remains.” We can think of this law as preserving the status quo of motion. Also note the expression “constant velocity;” this means that the object maintains a path along a straight line, since neither the magnitude nor the direction of the velocity vector changes. We can use [\[link\]](#) to consider the two parts of Newton's first law.



(a)



(b)

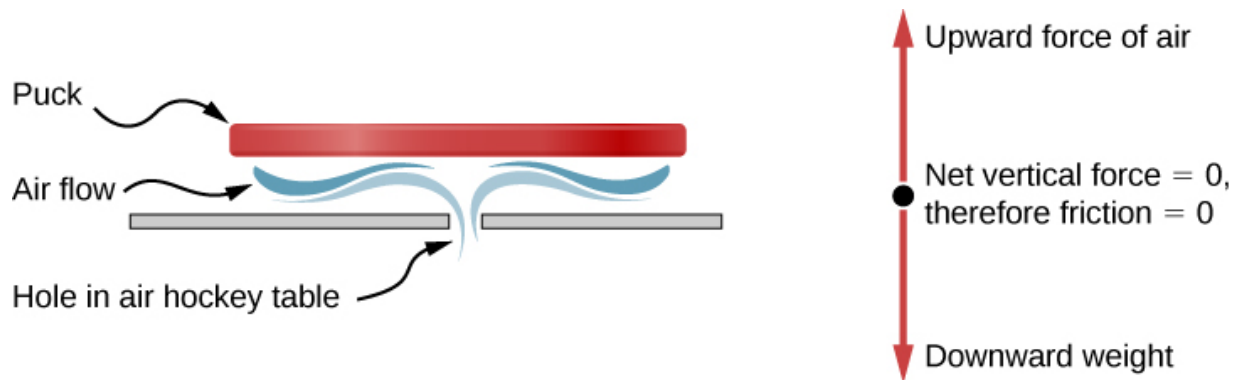
(a) A hockey puck is shown at rest; it remains at rest until an outside force such as a hockey stick changes its state of rest; (b) a hockey puck is shown in motion; it continues in motion in a straight line until an outside force causes it to change its state of motion. Although it is slick, an ice surface provides some friction that slows the puck.

Rather than contradicting our experience, Newton's first law says that there must be a cause for any change in velocity (a change in either magnitude or direction) to occur. This cause is a net external force, which we defined earlier in the chapter. An object sliding across a table or floor slows down due to the net force of friction acting on the object. If friction disappears, will the object still slow down?

The idea of cause and effect is crucial in accurately describing what happens in various situations. For example, consider what happens to an object sliding along a rough horizontal surface. The object quickly grinds to a halt. If we spray the surface with talcum powder to make the surface smoother, the object slides farther. If we make the surface even smoother by rubbing lubricating oil on it, the object slides farther yet. Extrapolating to a

frictionless surface and ignoring air resistance, we can imagine the object sliding in a straight line indefinitely. Friction is thus the cause of slowing (consistent with Newton's first law). The object would not slow down if friction were eliminated.

Consider an air hockey table ([link](#)). When the air is turned off, the puck slides only a short distance before friction slows it to a stop. However, when the air is turned on, it creates a nearly frictionless surface, and the puck glides long distances without slowing down. Additionally, if we know enough about the friction, we can accurately predict how quickly the object slows down.



An air hockey table is useful in illustrating Newton's laws. When the air is off, friction quickly slows the puck; but when the air is on, it minimizes contact between the puck and the hockey table, and the puck glides far down the table.

Newton's first law is general and can be applied to anything from an object sliding on a table to a satellite in orbit to blood pumped from the heart. Experiments have verified that any change in velocity (speed or direction) must be caused by an external force. The idea of *generally applicable or universal laws* is important—it is a basic feature of all laws of physics. Identifying these laws is like recognizing patterns in nature from which further patterns can be discovered. The genius of Galileo, who first developed the idea for the first law of motion, and Newton, who clarified it, was to ask the fundamental question: “What is the cause?” Thinking in terms of cause and effect is fundamentally different from the typical ancient

Greek approach, when questions such as “Why does a tiger have stripes?” would have been answered in Aristotelian fashion, such as “That is the nature of the beast.” The ability to think in terms of cause and effect is the ability to make a connection between an observed behavior and the surrounding world.

Gravitation and Inertia

Regardless of the scale of an object, whether a molecule or a subatomic particle, two properties remain valid and thus of interest to physics: gravitation and inertia. Both are connected to mass. Roughly speaking, *mass* is a measure of the amount of matter in something. *Gravitation* is the attraction of one mass to another, such as the attraction between yourself and Earth that holds your feet to the floor. The magnitude of this attraction is your weight, and it is a force.

Mass is also related to **inertia**, the ability of an object to resist changes in its motion—in other words, to resist acceleration. Newton’s first law is often called the **law of inertia**. As we know from experience, some objects have more inertia than others. It is more difficult to change the motion of a large boulder than that of a basketball, for example, because the boulder has more mass than the basketball. In other words, the inertia of an object is measured by its mass. The relationship between mass and weight is explored later in this chapter.

Inertial Reference Frames

Earlier, we stated Newton’s first law as “A body at rest remains at rest or, if in motion, remains in motion at constant velocity unless acted on by a net external force.” It can also be stated as “Every body remains in its state of uniform motion in a straight line unless it is compelled to change that state by forces acting on it.” To Newton, “uniform motion in a straight line” meant constant velocity, which includes the case of zero velocity, or rest. Therefore, the first law says that the velocity of an object remains constant if the net force on it is zero.

Newton's first law is usually considered to be a statement about reference frames. It provides a method for identifying a special type of reference frame: the **inertial reference frame**. In principle, we can make the net force on a body zero. If its velocity relative to a given frame is constant, then that frame is said to be inertial. So by definition, an inertial reference frame is a reference frame in which Newton's first law is valid. Newton's first law applies to objects with constant velocity. From this fact, we can infer the following statement.

Note:

Inertial Reference Frame

A reference frame moving at constant velocity relative to an inertial frame is also inertial. A reference frame accelerating relative to an inertial frame is not inertial.

Are inertial frames common in nature? It turns out that well within experimental error, a reference frame at rest relative to the most distant, or "fixed," stars is inertial. All frames moving uniformly with respect to this fixed-star frame are also inertial. For example, a nonrotating reference frame attached to the Sun is, for all practical purposes, inertial, because its velocity relative to the fixed stars does not vary by more than one part in 10^{10} . Earth accelerates relative to the fixed stars because it rotates on its axis and revolves around the Sun; hence, a reference frame attached to its surface is not inertial. For most problems, however, such a frame serves as a sufficiently accurate approximation to an inertial frame, because the acceleration of a point on Earth's surface relative to the fixed stars is rather small ($< 3.4 \times 10^{-2} \text{ m/s}^2$). Thus, unless indicated otherwise, we consider reference frames fixed on Earth to be inertial.

Finally, no particular inertial frame is more special than any other. As far as the laws of nature are concerned, all inertial frames are equivalent. In analyzing a problem, we choose one inertial frame over another simply on the basis of convenience.

Newton's First Law and Equilibrium

Newton's first law tells us about the equilibrium of a system, which is the state in which the forces on the system are balanced. Returning to [Forces](#) and the ice skaters in [\[link\]](#), we know that the forces \vec{F}_1 and \vec{F}_2 combine to form a resultant force, or the net external force: $\vec{F}_R = \vec{F}_{\text{net}} = \vec{F}_1 + \vec{F}_2$. To create equilibrium, we require a balancing force that will produce a net force of zero. This force must be equal in magnitude but opposite in direction to \vec{F}_R , which means the vector must be $-\vec{F}_R$. Referring to the ice skaters, for which we found \vec{F}_R to be $30.0\hat{i} + 40.0\hat{j}$ N, we can determine the balancing force by simply finding $-\vec{F}_R = -30.0\hat{i} - 40.0\hat{j}$ N. See the free-body diagram in [\[link\]](#)(b).

We can give Newton's first law in vector form:

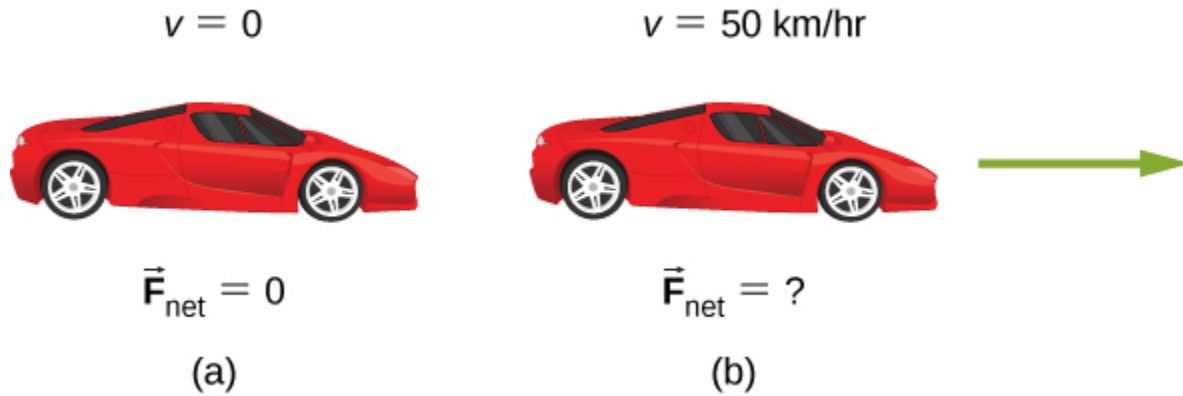
Note:

Equation:

$$\vec{v} = \text{constant when } \vec{F}_{\text{net}} = \vec{0} \text{ N.}$$

This equation says that a net force of zero implies that the velocity \vec{v} of the object is constant. (The word “constant” can indicate zero velocity.)

Newton's first law is deceptively simple. If a car is at rest, the only forces acting on the car are weight and the contact force of the pavement pushing up on the car ([\[link\]](#)). It is easy to understand that a nonzero net force is required to change the state of motion of the car. As a car moves with constant velocity, the friction force propels the car forward and opposes the drag force against it.



A car is shown (a) parked and (b) moving at constant velocity. How do Newton's laws apply to the parked car? What does the knowledge that the car is moving at constant velocity tell us about the net horizontal force on the car?

Example:

When Does Newton's First Law Apply to Your Car?

Newton's laws can be applied to all physical processes involving force and motion, including something as mundane as driving a car.

(a) Your car is parked outside your house. Does Newton's first law apply in this situation? Why or why not?

(b) Your car moves at constant velocity down the street. Does Newton's first law apply in this situation? Why or why not?

Strategy

In (a), we are considering the first part of Newton's first law, dealing with a body at rest; in (b), we look at the second part of Newton's first law for a body in motion.

Solution

- a. When your car is parked, all forces on the car must be balanced; the vector sum is 0 N. Thus, the net force is zero, and Newton's first law applies. The acceleration of the car is zero, and in this case, the velocity is also zero.

b. When your car is moving at constant velocity down the street, the net force must also be zero according to Newton's first law. The car's frictional force between the road and tires opposes the drag force on the car with the same magnitude, producing a net force of zero. The body continues in its state of constant velocity until the net force becomes nonzero. Realize that *a net force of zero means that an object is either at rest or moving with constant velocity, that is, it is not accelerating*. What do you suppose happens when the car accelerates? We explore this idea in the next section.

Significance

As this example shows, there are two kinds of equilibrium. In (a), the car is at rest; we say it is in *static equilibrium*. In (b), the forces on the car are balanced, but the car is moving; we say that it is in *dynamic equilibrium*. (We examine this idea in more detail in [Static Equilibrium and Elasticity](#).) Again, it is possible for two (or more) forces to act on an object yet for the object to move. In addition, a net force of zero cannot produce acceleration.

Note:

Exercise:

Problem:

Check Your Understanding A skydiver opens his parachute, and shortly thereafter, he is moving at constant velocity. (a) What forces are acting on him? (b) Which force is bigger?

Solution:

a. His weight acts downward, and the force of air resistance with the parachute acts upward. b. neither; the forces are equal in magnitude

Note:

Engage this [simulation](#) to predict, qualitatively, how an external force will affect the speed and direction of an object's motion. Explain the effects with the help of a free-body diagram. Use free-body diagrams to draw position, velocity, acceleration, and force graphs, and vice versa. Explain how the graphs relate to one another. Given a scenario or a graph, sketch all four graphs.

Summary

- According to Newton's first law, there must be a cause for any change in velocity (a change in either magnitude or direction) to occur. This law is also known as the law of inertia.
- Friction is an external force that causes an object to slow down.
- Inertia is the tendency of an object to remain at rest or remain in motion. Inertia is related to an object's mass.
- If an object's velocity relative to a given frame is constant, then the frame is inertial. This means that for an inertial reference frame, Newton's first law is valid.
- Equilibrium is achieved when the forces on a system are balanced.
- A net force of zero means that an object is either at rest or moving with constant velocity; that is, it is not accelerating.

Conceptual Questions

Exercise:

Problem:

Taking a frame attached to Earth as inertial, which of the following objects cannot have inertial frames attached to them, and which are inertial reference frames?

- (a) A car moving at constant velocity
- (b) A car that is accelerating

- (c) An elevator in free fall
- (d) A space capsule orbiting Earth
- (e) An elevator descending uniformly

Exercise:

Problem:

A woman was transporting an open box of cupcakes to a school party. The car in front of her stopped suddenly; she applied her brakes immediately. She was wearing her seat belt and suffered no physical harm (just a great deal of embarrassment), but the cupcakes flew into the dashboard and became “smushcakes.” Explain what happened.

Solution:

The cupcake velocity before the braking action was the same as that of the car. Therefore, the cupcakes were unrestricted bodies in motion, and when the car suddenly stopped, the cupcakes kept moving forward according to Newton’s first law.

Problems

Exercise:

Problem:

Two forces of $\vec{\mathbf{F}}_1 = \frac{75.0}{\sqrt{2}} (\hat{\mathbf{i}} - \hat{\mathbf{j}})$ N and $\vec{\mathbf{F}}_2 = \frac{150.0}{\sqrt{2}} (\hat{\mathbf{i}} - \hat{\mathbf{j}})$ N act on an object. Find the third force $\vec{\mathbf{F}}_3$ that is needed to balance the first two forces.

Exercise:

Problem:

While sliding a couch across a floor, Andrea and Jennifer exert forces \vec{F}_A and \vec{F}_J on the couch. Andrea's force is due north with a magnitude of 130.0 N and Jennifer's force is 32° east of north with a magnitude of 180.0 N. (a) Find the net force in component form. (b) Find the magnitude and direction of the net force. (c) If Andrea and Jennifer's housemates, David and Stephanie, disagree with the move and want to prevent its relocation, with what combined force \vec{F}_{DS} should they push so that the couch does not move?

Solution:

a. $\vec{F}_{\text{net}} = 95.0\hat{i} + 283\hat{j}\text{N}$; b. 299 N at 71° north of east; c.
 $\vec{F}_{DS} = - (95.0\hat{i} + 283\hat{j}) \text{ N}$

Glossary

inertia

ability of an object to resist changes in its motion

inertial reference frame

reference frame moving at constant velocity relative to an inertial frame is also inertial; a reference frame accelerating relative to an inertial frame is not inertial

law of inertia

see Newton's first law of motion

Newton's first law of motion

body at rest remains at rest or, if in motion, remains in motion at constant velocity unless acted on by a net external force; also known as the law of inertia

Newton's Second Law

By the end of the section, you will be able to:

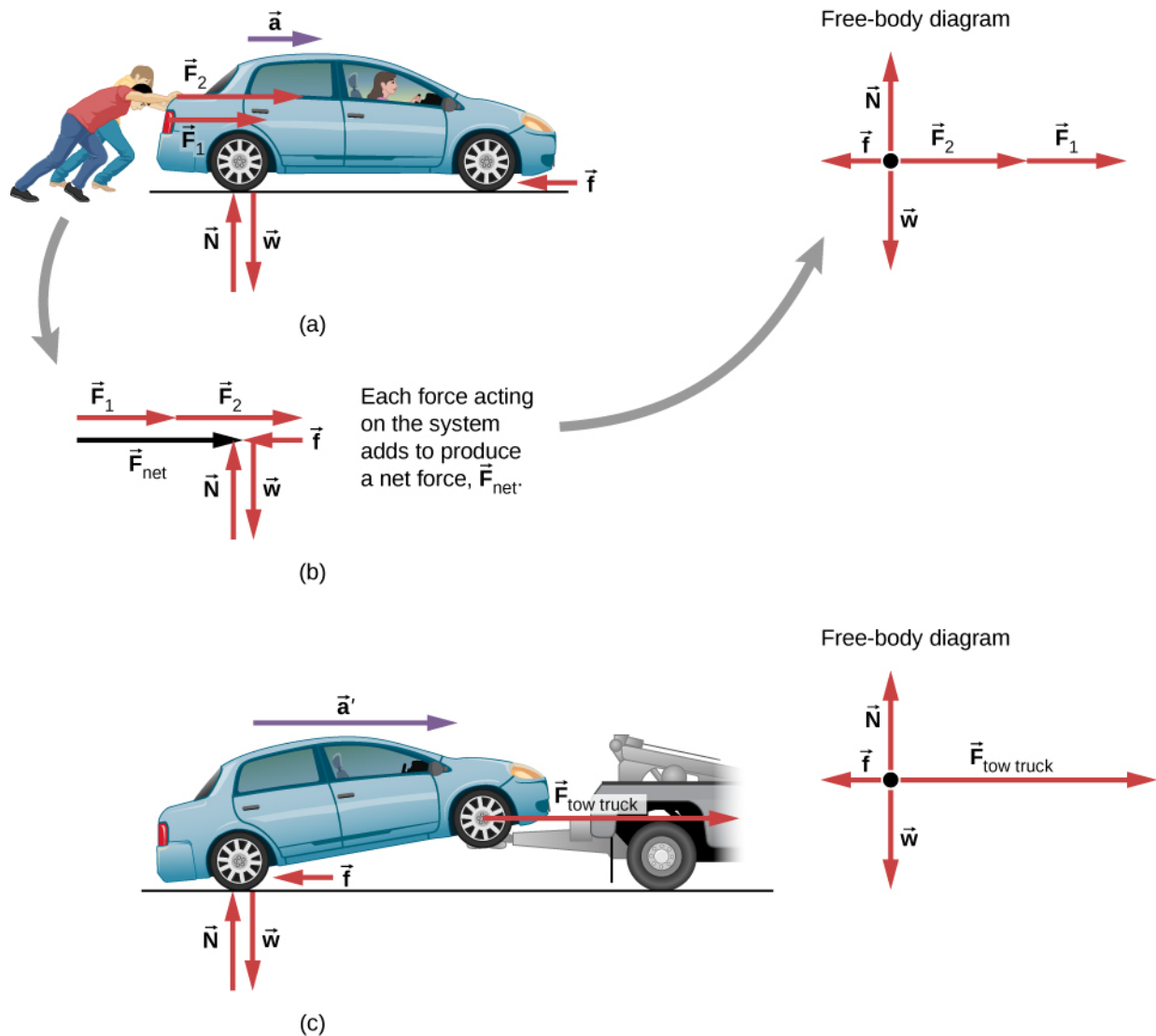
- Distinguish between external and internal forces
- Describe Newton's second law of motion
- Explain the dependence of acceleration on net force and mass

Newton's second law is closely related to his first law. It mathematically gives the cause-and-effect relationship between force and changes in motion. Newton's second law is quantitative and is used extensively to calculate what happens in situations involving a force. Before we can write down Newton's second law as a simple equation that gives the exact relationship of force, mass, and acceleration, we need to sharpen some ideas we mentioned earlier.

Force and Acceleration

First, what do we mean by a change in motion? The answer is that a change in motion is equivalent to a change in velocity. A change in velocity means, by definition, that there is acceleration. Newton's first law says that a net external force causes a change in motion; thus, we see that a *net external force causes nonzero acceleration*.

We defined external force in [Forces](#) as force acting on an object or system that originates outside of the object or system. Let's consider this concept further. An intuitive notion of *external* is correct—it is outside the system of interest. For example, in [\[link\]\(a\)](#), the system of interest is the car plus the person within it. The two forces exerted by the two students are external forces. In contrast, an internal force acts between elements of the system. Thus, the force the person in the car exerts to hang on to the steering wheel is an internal force between elements of the system of interest. Only external forces affect the motion of a system, according to Newton's first law. (The internal forces cancel each other out, as explained in the next section.) Therefore, we must define the boundaries of the system before we can determine which forces are external. Sometimes, the system is obvious, whereas at other times, identifying the boundaries of a system is more subtle. The concept of a system is fundamental to many areas of physics, as is the correct application of Newton's laws. This concept is revisited many times in the study of physics.



Different forces exerted on the same mass produce different accelerations. (a) Two students push a stalled car. All external forces acting on the car are shown. (b) The forces acting on the car are transferred to a coordinate plane (free-body diagram) for simpler analysis. (c) The tow truck can produce greater external force on the same mass, and thus greater acceleration.

From this example, you can see that different forces exerted on the same mass produce different accelerations. In [\[link\]](#)(a), the two students push a car with a driver in it. Arrows representing all external forces are shown. The system of interest is the car and its driver. The weight \vec{w} of the system and the support of the ground \vec{N} are also shown for completeness and are assumed to cancel (because there was no vertical motion and no imbalance of forces in the vertical direction to create a change in motion). The vector \vec{f} represents the friction acting on the car, and it acts to the left, opposing the motion of the car. (We discuss friction in more detail in the next chapter.) In [\[link\]](#)(b), all external forces acting on the system add together to produce the net force \vec{F}_{net} . The free-body diagram shows all of the forces acting on the system of interest. The dot represents the center of mass of the system.

Each force vector extends from this dot. Because there are two forces acting to the right, the vectors are shown collinearly. Finally, in [\[link\]](#)(c), a larger net external force produces a larger acceleration ($\vec{a}' > \vec{a}$) when the tow truck pulls the car.

It seems reasonable that acceleration would be directly proportional to and in the same direction as the net external force acting on a system. This assumption has been verified experimentally and is illustrated in [\[link\]](#). To obtain an equation for Newton's second law, we first write the relationship of acceleration \vec{a} and net external force \vec{F}_{net} as the proportionality

Equation:

$$\vec{a} \propto \vec{F}_{\text{net}}$$

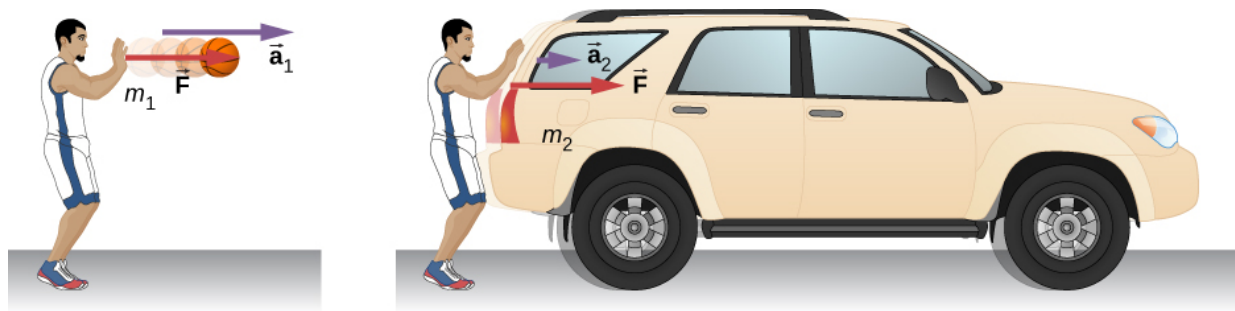
where the symbol \propto means “proportional to.” (Recall from [Forces](#) that the net external force is the vector sum of all external forces and is sometimes indicated as $\sum \vec{F}$.) This proportionality shows what we have said in words—acceleration is directly proportional to net external force. Once the system of interest is chosen, identify the external forces and ignore the internal ones. It is a tremendous simplification to disregard the numerous internal forces acting between objects within the system, such as muscular forces within the students' bodies, let alone the myriad forces between the atoms in the objects. Still, this simplification helps us solve some complex problems.

It also seems reasonable that acceleration should be inversely proportional to the mass of the system. In other words, the larger the mass (the inertia), the smaller the acceleration produced by a given force. As illustrated in [\[link\]](#), the same net external force applied to a basketball produces a much smaller acceleration when it is applied to an SUV. The proportionality is written as

Equation:

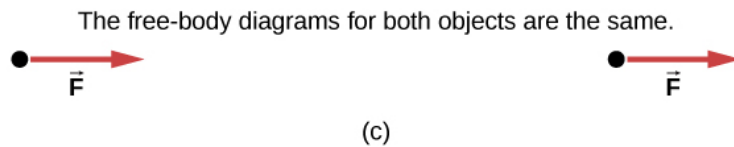
$$a \propto \frac{1}{m},$$

where m is the mass of the system and a is the magnitude of the acceleration. Experiments have shown that acceleration is exactly inversely proportional to mass, just as it is directly proportional to net external force.



(a)

(b)



The same force exerted on systems of different masses produces different accelerations. (a) A basketball player pushes on a basketball to make a pass. (Ignore the effect of gravity on the ball.) (b) The same player exerts an identical force on a stalled SUV and produces far less acceleration. (c) The free-body diagrams are identical, permitting direct comparison of the two situations. A series of patterns for free-body diagrams will emerge as you do more problems and learn how to draw them in [Drawing Free-Body Diagrams](#).

It has been found that the acceleration of an object depends only on the net external force and the mass of the object. Combining the two proportionalities just given yields **Newton's second law**.

Note:

Newton's Second Law of Motion

The acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system and is inversely proportion to its mass. In equation form, Newton's second law is

Equation:

$$\vec{a} = \frac{\vec{F}_{\text{net}}}{m},$$

where \vec{a} is the acceleration, \vec{F}_{net} is the net force, and m is the mass. This is often written in the more familiar form

Equation:

$$\vec{F}_{\text{net}} = \sum \vec{F} = m\vec{a},$$

but the first equation gives more insight into what Newton's second law means. When only the magnitude of force and acceleration are considered, this equation can be written in the simpler scalar

form:

Equation:

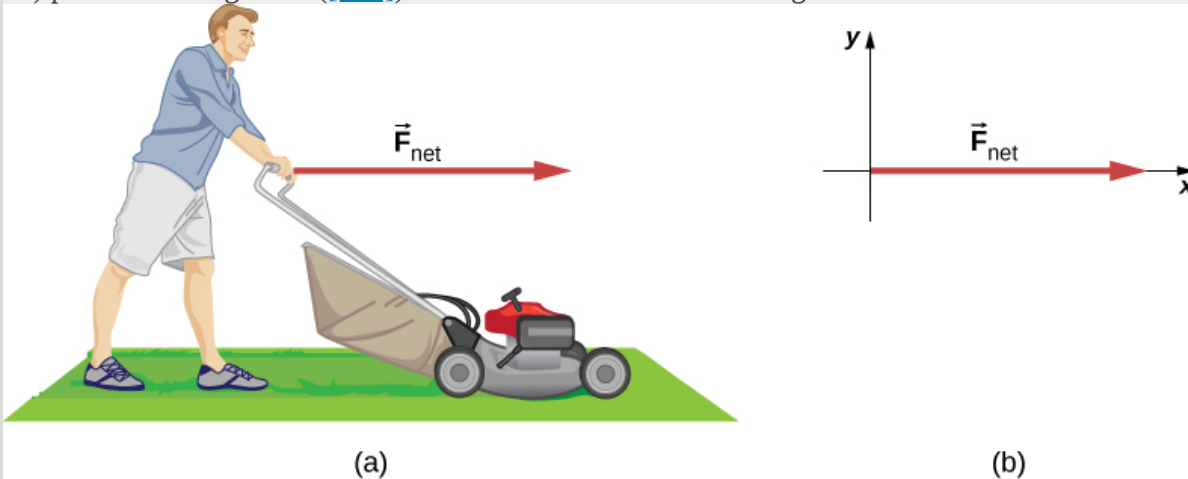
$$F_{\text{net}} = ma.$$

The law is a cause-and-effect relationship among three quantities that is not simply based on their definitions. The validity of the second law is based on experimental verification. The free-body diagram, which you will learn to draw in [Drawing Free-Body Diagrams](#), is the basis for writing Newton's second law.

Example:

What Acceleration Can a Person Produce When Pushing a Lawn Mower?

Suppose that the net external force (push minus friction) exerted on a lawn mower is 51 N (about 11 lb.) parallel to the ground ([link](#)). The mass of the mower is 24 kg. What is its acceleration?



(a) The net force on a lawn mower is 51 N to the right. At what rate does the lawn mower accelerate to the right? (b) The free-body diagram for this problem is shown.

Strategy

This problem involves only motion in the horizontal direction; we are also given the net force, indicated by the single vector, but we can suppress the vector nature and concentrate on applying Newton's second law. Since F_{net} and m are given, the acceleration can be calculated directly from Newton's second law as $F_{\text{net}} = ma$.

Solution

The magnitude of the acceleration a is $a = F_{\text{net}}/m$. Entering known values gives

Equation:

$$a = \frac{51 \text{ N}}{24 \text{ kg}}.$$

Substituting the unit of kilograms times meters per square second for newtons yields

Equation:

$$a = \frac{51 \text{ kg} \cdot \text{m/s}^2}{24 \text{ kg}} = 2.1 \text{ m/s}^2.$$

Significance

The direction of the acceleration is the same direction as that of the net force, which is parallel to the ground. This is a result of the vector relationship expressed in Newton's second law, that is, the vector representing net force is the scalar multiple of the acceleration vector. There is no information given in this example about the individual external forces acting on the system, but we can say something about their relative magnitudes. For example, the force exerted by the person pushing the mower must be greater than the friction opposing the motion (since we know the mower moved forward), and the vertical forces must cancel because no acceleration occurs in the vertical direction (the mower is moving only horizontally). The acceleration found is small enough to be reasonable for a person pushing a mower. Such an effort would not last too long, because the person's top speed would soon be reached.

Note:

Exercise:

Problem:

Check Your Understanding At the time of its launch, the HMS *Titanic* was the most massive mobile object ever built, with a mass of $6.0 \times 10^7 \text{ kg}$. If a force of 6 MN ($6 \times 10^6 \text{ N}$) was applied to the ship, what acceleration would it experience?

Solution:

$$0.1 \text{ m/s}^2$$

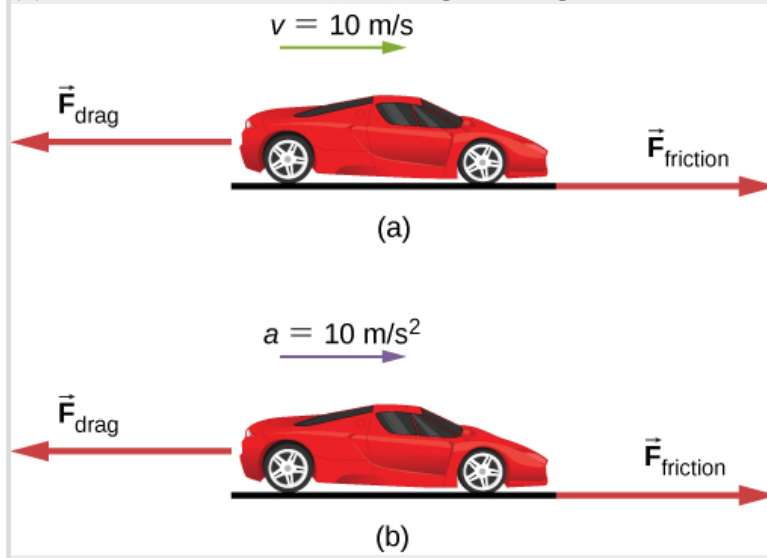
In the preceding example, we dealt with net force only for simplicity. However, several forces act on the lawn mower. The weight \vec{w} (discussed in detail in [Mass and Weight](#)) pulls down on the mower, toward the center of Earth; this produces a contact force on the ground. The ground must exert an upward force on the lawn mower, known as the normal force \vec{N} , which we define in [Common Forces](#). These forces are balanced and therefore do not produce vertical acceleration. In the next example, we show both of these forces. As you continue to solve problems using Newton's second law, be sure to show multiple forces.

Example:

Which Force Is Bigger?

(a) The car shown in [\[link\]](#) is moving at a constant speed. Which force is bigger, $\vec{F}_{\text{friction}}$ or \vec{F}_{drag} ? Explain.

(b) The same car is now accelerating to the right. Which force is bigger, $\vec{F}_{\text{friction}}$ or \vec{F}_{drag} ? Explain.



A car is shown (a) moving at constant speed and (b) accelerating. How do the forces acting on the car compare in each case? (a) What does the knowledge that the car is moving at constant velocity tell us about the net horizontal force on the car compared to the friction force? (b) What does the knowledge that the car is accelerating tell us about the horizontal force on the car compared to the friction force?

Strategy

We must consider Newton's first and second laws to analyze the situation. We need to decide which law applies; this, in turn, will tell us about the relationship between the forces.

Solution

- The forces are equal. According to Newton's first law, if the net force is zero, the velocity is constant.
- In this case, $\vec{F}_{\text{friction}}$ must be larger than \vec{F}_{drag} . According to Newton's second law, a net force is required to cause acceleration.

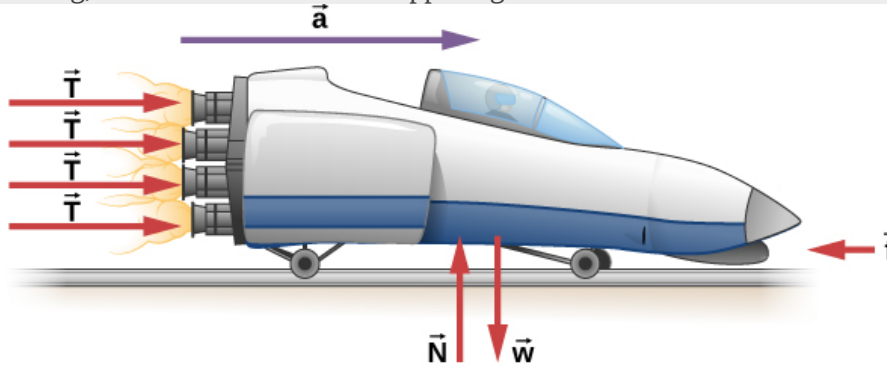
Significance

These questions may seem trivial, but they are commonly answered incorrectly. For a car or any other object to move, it must be accelerated from rest to the desired speed; this requires that the friction force be greater than the drag force. Once the car is moving at constant velocity, the net force must be zero; otherwise, the car will accelerate (gain speed). To solve problems involving Newton's laws, we must understand whether to apply Newton's first law (where $\sum \vec{F} = \vec{0}$) or Newton's second law (where $\sum \vec{F}$ is not zero). This will be apparent as you see more examples and attempt to solve problems on your own.

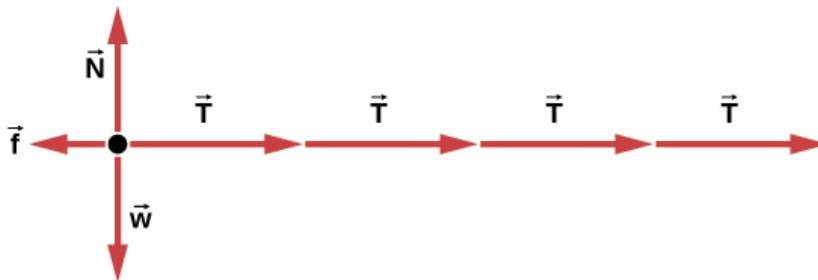
Example:**What Rocket Thrust Accelerates This Sled?**

Before space flights carrying astronauts, rocket sleds were used to test aircraft, missile equipment, and physiological effects on human subjects at high speeds. They consisted of a platform that was mounted on one or two rails and propelled by several rockets.

Calculate the magnitude of force exerted by each rocket, called its thrust T , for the four-rocket propulsion system shown in [\[link\]](#). The sled's initial acceleration is 49 m/s^2 , the mass of the system is 2100 kg , and the force of friction opposing the motion is 650 N .



Free-body diagram



A sled experiences a rocket thrust that accelerates it to the right. Each rocket creates an identical thrust T . The system here is the sled, its rockets, and its rider, so none of the forces between these objects are considered. The arrow representing friction (\vec{f}) is drawn larger than scale.

Strategy

Although forces are acting both vertically and horizontally, we assume the vertical forces cancel because there is no vertical acceleration. This leaves us with only horizontal forces and a simpler one-dimensional problem. Directions are indicated with plus or minus signs, with right taken as the positive direction. See the free-body diagram in [\[link\]](#).

Solution

Since acceleration, mass, and the force of friction are given, we start with Newton's second law and look for ways to find the thrust of the engines. We have defined the direction of the force and

acceleration as acting “to the right,” so we need to consider only the magnitudes of these quantities in the calculations. Hence we begin with

Equation:

$$F_{\text{net}} = ma$$

where F_{net} is the net force along the horizontal direction. We can see from the figure that the engine thrusts add, whereas friction opposes the thrust. In equation form, the net external force is

Equation:

$$F_{\text{net}} = 4T - f.$$

Substituting this into Newton’s second law gives us

Equation:

$$F_{\text{net}} = ma = 4T - f.$$

Using a little algebra, we solve for the total thrust $4T$:

Equation:

$$4T = ma + f.$$

Substituting known values yields

Equation:

$$4T = ma + f = (2100 \text{ kg})(49 \text{ m/s}^2) + 650 \text{ N}.$$

Therefore, the total thrust is

Equation:

$$4T = 1.0 \times 10^5 \text{ N},$$

and the individual thrusts are

Equation:

$$T = \frac{1.0 \times 10^5 \text{ N}}{4} = 2.5 \times 10^4 \text{ N}.$$

Significance

The numbers are quite large, so the result might surprise you. Experiments such as this were performed in the early 1960s to test the limits of human endurance, and the setup was designed to protect human subjects in jet fighter emergency ejections. Speeds of 1000 km/h were obtained, with accelerations of 45 g ’s. (Recall that g , acceleration due to gravity, is 9.80 m/s^2 . When we say that acceleration is 45 g ’s, it is $45 \times 9.8 \text{ m/s}^2$, which is approximately 440 m/s^2 .) Although living subjects are not used anymore, land speeds of 10,000 km/h have been obtained with a rocket sled. In this example, as in the preceding one, the system of interest is obvious. We see in later examples that choosing the system of interest is crucial—and the choice is not always obvious. Newton’s second law is more than a definition; it is a relationship among acceleration, force, and mass. It can help us make predictions. Each of those physical quantities can be defined independently, so the second law tells us something basic and universal about nature.

Note:

Exercise:

Problem:

Check Your Understanding A 550-kg sports car collides with a 2200-kg truck, and during the collision, the net force on each vehicle is the force exerted by the other. If the magnitude of the truck's acceleration is 10 m/s^2 , what is the magnitude of the sports car's acceleration?

Solution:

$$40 \text{ m/s}^2$$

Component Form of Newton's Second Law

We have developed Newton's second law and presented it as a vector equation in [\[link\]](#). This vector equation can be written as three component equations:

Note:

Equation:

$$\sum \vec{F}_x = m\vec{a}_x, \sum \vec{F}_y = m\vec{a}_y, \text{ and } \sum \vec{F}_z = m\vec{a}_z.$$

The second law is a description of how a body responds mechanically to its environment. The influence of the environment is the net force \vec{F}_{net} , the body's response is the acceleration \vec{a} , and the strength of the response is inversely proportional to the mass m . The larger the mass of an object, the smaller its response (its acceleration) to the influence of the environment (a given net force). Therefore, a body's mass is a measure of its inertia, as we explained in [Newton's First Law](#).

Example:

Force on a Soccer Ball

A 0.400-kg soccer ball is kicked across the field by a player; it undergoes acceleration given by $\vec{a} = 3.00\hat{i} + 7.00\hat{j} \text{ m/s}^2$. Find (a) the resultant force acting on the ball and (b) the magnitude and direction of the resultant force.

Strategy

The vectors in \hat{i} and \hat{j} format, which indicate force direction along the x -axis and the y -axis, respectively, are involved, so we apply Newton's second law in vector form.

Solution

- a. We apply Newton's second law:

Equation:

$$\vec{\mathbf{F}}_{\text{net}} = m\vec{\mathbf{a}} = (0.400 \text{ kg}) (3.00\hat{\mathbf{i}} + 7.00\hat{\mathbf{j}} \text{ m/s}^2) = 1.20\hat{\mathbf{i}} + 2.80\hat{\mathbf{j}} \text{ N}.$$

b. Magnitude and direction are found using the components of $\vec{\mathbf{F}}_{\text{net}}$:

Equation:

$$F_{\text{net}} = \sqrt{(1.20 \text{ N})^2 + (2.80 \text{ N})^2} = 3.05 \text{ N} \text{ and } \theta = \tan^{-1} \left(\frac{2.80}{1.20} \right) = 66.8^\circ.$$

Significance

We must remember that Newton's second law is a vector equation. In (a), we are multiplying a vector by a scalar to determine the net force in vector form. While the vector form gives a compact representation of the force vector, it does not tell us how "big" it is, or where it goes, in intuitive terms. In (b), we are determining the actual size (magnitude) of this force and the direction in which it travels.

Example:**Mass of a Car**

Find the mass of a car if a net force of $-600.0\hat{\mathbf{j}} \text{ N}$ produces an acceleration of $-0.2\hat{\mathbf{j}} \text{ m/s}^2$.

Strategy

Vector division is not defined, so $m = \vec{\mathbf{F}}_{\text{net}}/\vec{\mathbf{a}}$ cannot be performed. However, mass m is a scalar, so we can use the scalar form of Newton's second law, $m = F_{\text{net}}/a$.

Solution

We use $m = F_{\text{net}}/a$ and substitute the magnitudes of the two vectors: $F_{\text{net}} = 600.0 \text{ N}$ and $a = 0.2 \text{ m/s}^2$. Therefore,

Equation:

$$m = \frac{F_{\text{net}}}{a} = \frac{600.0 \text{ N}}{0.2 \text{ m/s}^2} = 3000 \text{ kg}.$$

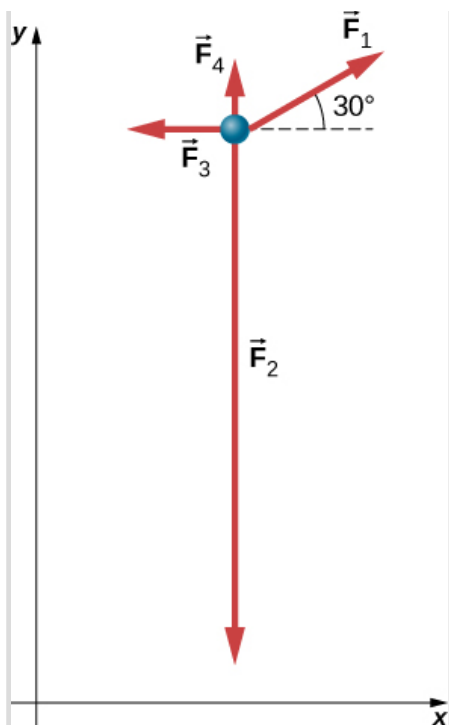
Significance

Force and acceleration were given in the $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ format, but the answer, mass m , is a scalar and thus is not given in $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ form.

Example:**Several Forces on a Particle**

A particle of mass $m = 4.0 \text{ kg}$ is acted upon by four forces of magnitudes.

$F_1 = 10.0 \text{ N}$, $F_2 = 40.0 \text{ N}$, $F_3 = 5.0 \text{ N}$, and $F_4 = 2.0 \text{ N}$, with the directions as shown in the free-body diagram in [\[link\]](#). What is the acceleration of the particle?



Four forces in the xy -plane are applied to a 4.0-kg particle.

Strategy

Because this is a two-dimensional problem, we must use a free-body diagram. First, \vec{F}_1 must be resolved into x - and y -components. We can then apply the second law in each direction.

Solution

We draw a free-body diagram as shown in [\[link\]](#). Now we apply Newton's second law. We consider all vectors resolved into x - and y -components:

Equation:

$$\sum F_x = ma_x$$

$$F_{1x} - F_{3x} = ma_x$$

$$F_1 \cos 30^\circ - F_{3x} = ma_x$$

$$(10.0 \text{ N}) (\cos 30^\circ) - 5.0 \text{ N} = (4.0 \text{ kg})a_x$$

$$a_x = 0.92 \text{ m/s}^2.$$

$$\sum F_y = ma_y$$

$$F_{1y} + F_{4y} - F_{2y} = ma_y$$

$$F_1 \sin 30^\circ + F_{4y} - F_{2y} = ma_y$$

$$(10.0 \text{ N}) (\sin 30^\circ) + 2.0 \text{ N} - 40.0 \text{ N} = (4.0 \text{ kg})a_y$$

$$a_y = -8.3 \text{ m/s}^2.$$

Thus, the net acceleration is

Equation:

$$\vec{a} = (0.92\hat{i} - 8.3\hat{j}) \text{ m/s}^2,$$

which is a vector of magnitude 8.4 m/s^2 directed at 276° to the positive x -axis.

Significance

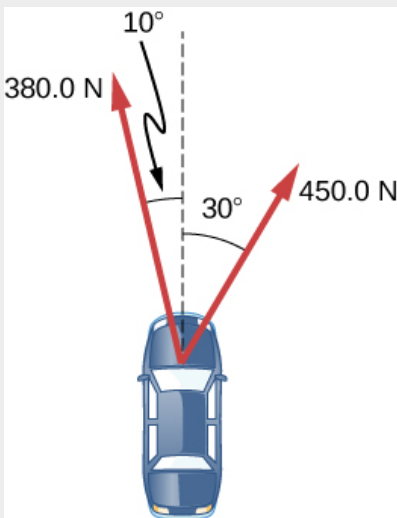
Numerous examples in everyday life can be found that involve three or more forces acting on a single object, such as cables running from the Golden Gate Bridge or a football player being tackled by three defenders. We can see that the solution of this example is just an extension of what we have already done.

Note:

Exercise:

Problem:

Check Your Understanding A car has forces acting on it, as shown below. The mass of the car is 1000.0 kg. The road is slick, so friction can be ignored. (a) What is the net force on the car? (b) What is the acceleration of the car?



Solution:

a. $159.0\hat{i} + 770.0\hat{j}$ N; b. $0.1590\hat{i} + 0.7700\hat{j}$ N

Newton's Second Law and Momentum

Newton actually stated his second law in terms of momentum: “The instantaneous rate at which a body’s momentum changes is equal to the net force acting on the body.” (“Instantaneous rate” implies that the derivative is involved.) This can be given by the vector equation

Note:

Equation:

$$\vec{\mathbf{F}}_{\text{net}} = \frac{d\vec{\mathbf{p}}}{dt}.$$

This means that Newton’s second law addresses the central question of motion: What causes a change in motion of an object? Momentum was described by Newton as “quantity of motion,” a way of combining both the velocity of an object and its mass. We devote [Linear Momentum and Collisions](#) to the study of momentum.

For now, it is sufficient to define *momentum* $\vec{\mathbf{p}}$ as the product of the mass of the object m and its velocity $\vec{\mathbf{v}}$:

Equation:

$$\vec{\mathbf{p}} = m\vec{\mathbf{v}}.$$

Since velocity is a vector, so is momentum.

It is easy to visualize momentum. A train moving at 10 m/s has more momentum than one that moves at 2 m/s. In everyday life, we speak of one sports team as “having momentum” when they score points against the opposing team.

If we substitute [\[link\]](#) into [\[link\]](#), we obtain

Equation:

$$\vec{\mathbf{F}}_{\text{net}} = \frac{d\vec{\mathbf{p}}}{dt} = \frac{d(m\vec{\mathbf{v}})}{dt}.$$

When m is constant, we have

Equation:

$$\vec{\mathbf{F}}_{\text{net}} = m \frac{d(\vec{\mathbf{v}})}{dt} = m\vec{\mathbf{a}}.$$

Thus, we see that the momentum form of Newton’s second law reduces to the form given earlier in this section.

Note:

Explore the [forces at work](#) when [pulling a cart](#) or pushing a refrigerator, crate, or person. Create an [applied force](#) and see how it makes objects move. Put [an object on a ramp](#) and see how it affects its motion.

Summary

- An external force acts on a system from outside the system, as opposed to internal forces, which act between components within the system.
- Newton's second law of motion says that the net external force on an object with a certain mass is directly proportional to and in the same direction as the acceleration of the object.
- Newton's second law can also describe net force as the instantaneous rate of change of momentum. Thus, a net external force causes nonzero acceleration.

Conceptual Questions

Exercise:

Problem:

Why can we neglect forces such as those holding a body together when we apply Newton's second law?

Exercise:

Problem:

A rock is thrown straight up. At the top of the trajectory, the velocity is momentarily zero. Does this imply that the force acting on the object is zero? Explain your answer.

Solution:

No. If the force were zero at this point, then there would be nothing to change the object's momentary zero velocity. Since we do not observe the object hanging motionless in the air, the force could not be zero.

Problems

Exercise:

Problem:

Andrea, a 63.0-kg sprinter, starts a race with an acceleration of 4.200 m/s^2 . What is the net external force on her?

Exercise:

Problem:

If the sprinter from the previous problem accelerates at that rate for 20.00 m and then maintains that velocity for the remainder of a 100.00-m dash, what will her time be for the race?

Solution:

Running from rest, the sprinter attains a velocity of $v = 12.96 \text{ m/s}$, at end of acceleration. We find the time for acceleration using $x = 20.00 \text{ m} = 0 + 0.5at_1^2$, or $t_1 = 3.086 \text{ s}$. For maintained velocity, $x_2 = vt_2$, or $t_2 = x_2/v = 80.00 \text{ m}/12.96 \text{ m/s} = 6.173 \text{ s}$.
Total time = 9.259 s.

Exercise:

Problem:

A cleaner pushes a 4.50-kg laundry cart in such a way that the net external force on it is 60.0 N. Calculate the magnitude of his cart's acceleration.

Exercise:**Problem:**

Astronauts in orbit are apparently weightless. This means that a clever method of measuring the mass of astronauts is needed to monitor their mass gains or losses, and adjust their diet. One way to do this is to exert a known force on an astronaut and measure the acceleration produced. Suppose a net external force of 50.0 N is exerted, and an astronaut's acceleration is measured to be 0.893 m/s^2 . (a) Calculate her mass. (b) By exerting a force on the astronaut, the vehicle in which she orbits experiences an equal and opposite force. Use this knowledge to find an equation for the acceleration of the system (astronaut and spaceship) that would be measured by a nearby observer. (c) Discuss how this would affect the measurement of the astronaut's acceleration. Propose a method by which recoil of the vehicle is avoided.

Solution:

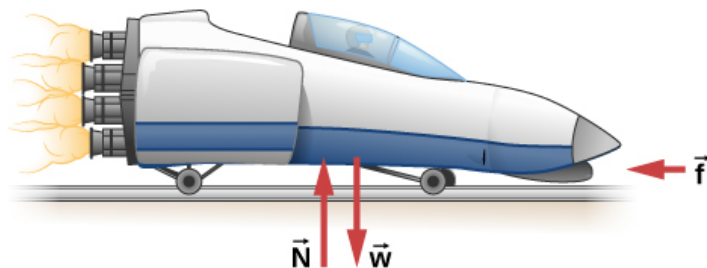
a. $m = 56.0 \text{ kg}$; b. $a_{\text{meas}} = a_{\text{astro}} + a_{\text{ship}}$, where $a_{\text{ship}} = \frac{m_{\text{astro}} a_{\text{astro}}}{m_{\text{ship}}}$; c. If the force could be exerted on the astronaut by another source (other than the spaceship), then the spaceship would not experience a recoil.

Exercise:**Problem:**

In [\[link\]](#), the net external force on the 24-kg mower is given as 51 N. If the force of friction opposing the motion is 24 N, what force F (in newtons) is the person exerting on the mower? Suppose the mower is moving at 1.5 m/s when the force F is removed. How far will the mower go before stopping?

Exercise:**Problem:**

The rocket sled shown below accelerates opposite to the motion at a rate of 196 m/s^2 . What force is necessary to produce this acceleration opposite to the motion? Assume that the rockets are off. The mass of the system is $2.10 \times 10^3 \text{ kg}$.

**Solution:**

$$F_{\text{net}} = 4.12 \times 10^5 \text{ N}$$

Exercise:

Problem:

If the rocket sled shown in the previous problem starts with only one rocket burning, what is the magnitude of this acceleration? Assume that the mass of the system is $2.10 \times 10^3 \text{ kg}$, the thrust T is $2.40 \times 10^4 \text{ N}$, and the force of friction opposing the motion is 650.0 N . (b) Why is the acceleration not one-fourth of what it is with all rockets burning?

Exercise:

Problem:

What is the acceleration opposite to the motion of the rocket sled if it comes to rest in 1.10 s from a speed of 1000.0 km/h ? (Such acceleration opposite to the motion caused one test subject to black out and have temporary blindness.)

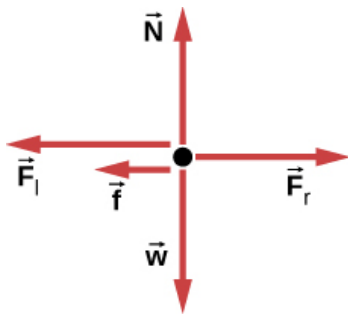
Solution:

$$a = 253 \text{ m/s}^2$$

Exercise:

Problem:

Suppose two children push horizontally, but in exactly opposite directions, on a third child in a wagon. The first child exerts a force of 75.0 N , the second exerts a force of 90.0 N , friction is 12.0 N , and the mass of the third child plus wagon is 23.0 kg . (a) What is the system of interest if the acceleration of the child in the wagon is to be calculated? (See the free-body diagram.) (b) Calculate the acceleration. (c) What would the acceleration be if friction were 15.0 N ?



Exercise:

Problem:

A powerful motorcycle can produce an acceleration of 3.50 m/s^2 while traveling at 90.0 km/h . At that speed, the forces resisting motion, including friction and air resistance, total 400.0 N . (Air resistance is analogous to air friction. It always opposes the motion of an object.) What is the magnitude of the force that motorcycle exerts backward on the ground to produce its acceleration if the mass of the motorcycle with rider is 245 kg ?

Solution:

$$F_{\text{net}} = F - f = ma \Rightarrow F = 1.26 \times 10^3 \text{ N}$$

Exercise:**Problem:**

A car with a mass of 1000.0 kg accelerates from 0 to 90.0 km/h in 10.0 s. (a) What is its acceleration? (b) What is the net force on the car?

Exercise:**Problem:**

The driver in the previous problem applies the brakes when the car is moving at 90.0 km/h, and the car comes to rest after traveling 40.0 m. What is the net force on the car during its acceleration opposite to the motion?

Solution:

$$v^2 = v_0^2 + 2ax \Rightarrow a = -7.80 \text{ m/s}^2$$

$$F_{\text{net}} = -7.80 \times 10^3 \text{ N}$$

Exercise:**Problem:**

An 80.0-kg passenger in an SUV traveling at 1.00×10^2 km/h is wearing a seat belt. The driver slams on the brakes and the SUV stops in 45.0 m. Find the force of the seat belt on the passenger.

Exercise:**Problem:**

A particle of mass 2.0 kg is acted on by a single force $\vec{F}_1 = 18\hat{i}$ N. (a) What is the particle's acceleration? (b) If the particle starts at rest, how far does it travel in the first 5.0 s?

Solution:

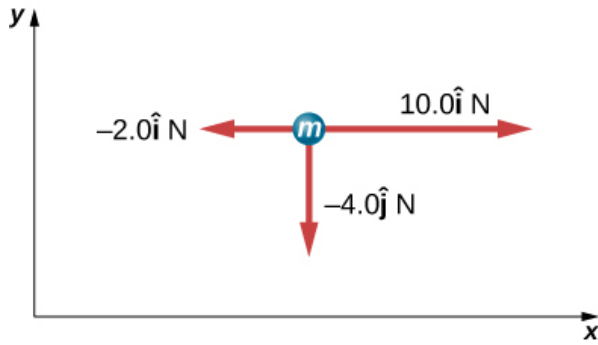
a. $\vec{F}_{\text{net}} = m\vec{a} \Rightarrow \vec{a} = 9.0\hat{i} \text{ m/s}^2$; b. The acceleration has magnitude 9.0 m/s^2 , so $x = 110 \text{ m}$.

Exercise:**Problem:**

Suppose that the particle of the previous problem also experiences forces $\vec{F}_2 = -15\hat{i}$ N and $\vec{F}_3 = 6.0\hat{j}$ N. What is its acceleration in this case?

Exercise:

Problem: Find the acceleration of the body of mass 5.0 kg shown below.



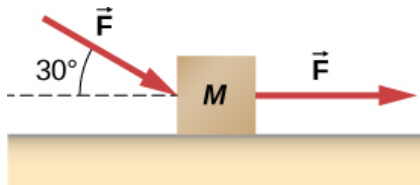
Solution:

$$1.6\hat{i} - 0.8\hat{j} \text{ m/s}^2$$

Exercise:

Problem:

In the following figure, the horizontal surface on which this block slides is frictionless. If the two forces acting on it each have magnitude $F = 30.0 \text{ N}$ and $M = 10.0 \text{ kg}$, what is the magnitude of the resulting acceleration of the block?



Glossary

Newton's second law of motion

acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system and is inversely proportional to its mass

Mass and Weight

By the end of the section, you will be able to:

- Explain the difference between mass and weight
- Explain why falling objects on Earth are never truly in free fall
- Describe the concept of weightlessness

Mass and weight are often used interchangeably in everyday conversation. For example, our medical records often show our weight in kilograms but never in the correct units of newtons. In physics, however, there is an important distinction. Weight is the pull of Earth on an object. It depends on the distance from the center of Earth. Unlike weight, mass does not vary with location. The mass of an object is the same on Earth, in orbit, or on the surface of the Moon.

Units of Force

The equation $F_{\text{net}} = ma$ is used to define net force in terms of mass, length, and time. As explained earlier, the SI unit of force is the newton. Since $F_{\text{net}} = ma$,

Equation:

$$1 \text{ N} = 1 \text{ kg} \cdot \text{m}/\text{s}^2.$$

Although almost the entire world uses the newton for the unit of force, in the United States, the most familiar unit of force is the pound (lb), where $1 \text{ N} = 0.225 \text{ lb}$. Thus, a 225-lb person weighs 1000 N.

Weight and Gravitational Force

When an object is dropped, it accelerates toward the center of Earth. Newton's second law says that a net force on an object is responsible for its acceleration. If air resistance is negligible, the net force on a falling object is the gravitational force, commonly called its **weight** \vec{w} , or its force due to gravity acting on an object of mass m . Weight can be denoted as a vector because it has a direction; *down* is, by definition, the direction of gravity,

and hence, weight is a downward force. The magnitude of weight is denoted as w . Galileo was instrumental in showing that, in the absence of air resistance, all objects fall with the same acceleration g . Using Galileo's result and Newton's second law, we can derive an equation for weight.

Consider an object with mass m falling toward Earth. It experiences only the downward force of gravity, which is the weight \vec{w} . Newton's second law says that the magnitude of the net external force on an object is $\vec{F}_{\text{net}} = m\vec{a}$. We know that the acceleration of an object due to gravity is \vec{g} , or $\vec{a} = \vec{g}$. Substituting these into Newton's second law gives us the following equations.

Note:

Weight

The gravitational force on a mass is its weight. We can write this in vector form, where \vec{w} is weight and m is mass, as

Equation:

$$\vec{w} = m\vec{g}.$$

In scalar form, we can write

Equation:

$$w = mg.$$

Since $g = 9.80 \text{ m/s}^2$ on Earth, the weight of a 1.00-kg object on Earth is 9.80 N:

Equation:

$$w = mg = (1.00 \text{ kg})(9.80 \text{ m/s}^2) = 9.80 \text{ N}.$$

When the net external force on an object is its weight, we say that it is in **free fall**, that is, the only force acting on the object is gravity. However, when objects on Earth fall downward, they are never truly in free fall because there is always some upward resistance force from the air acting on the object.

Acceleration due to gravity g varies slightly over the surface of Earth, so the weight of an object depends on its location and is not an intrinsic property of the object. Weight varies dramatically if we leave Earth's surface. On the Moon, for example, acceleration due to gravity is only 1.67 m/s^2 . A 1.0-kg mass thus has a weight of 9.8 N on Earth and only about 1.7 N on the Moon.

The broadest definition of weight in this sense is that the weight of an object is the gravitational force on it from the nearest large body, such as Earth, the Moon, or the Sun. This is the most common and useful definition of weight in physics. It differs dramatically, however, from the definition of weight used by NASA and the popular media in relation to space travel and exploration. When they speak of “weightlessness” and “microgravity,” they are referring to the phenomenon we call “free fall” in physics. We use the preceding definition of weight, force \vec{w} due to gravity acting on an object of mass m , and we make careful distinctions between free fall and actual weightlessness.

Be aware that weight and mass are different physical quantities, although they are closely related. Mass is an intrinsic property of an object: It is a quantity of matter. The quantity or amount of matter of an object is determined by the numbers of atoms and molecules of various types it contains. Because these numbers do not vary, in Newtonian physics, mass does not vary; therefore, its response to an applied force does not vary. In contrast, weight is the gravitational force acting on an object, so it does vary depending on gravity. For example, a person closer to the center of Earth, at a low elevation such as New Orleans, weighs slightly more than a person who is located in the higher elevation of Denver, even though they may have the same mass.

It is tempting to equate mass to weight, because most of our examples take place on Earth, where the weight of an object varies only a little with the location of the object. In addition, it is difficult to count and identify all of the atoms and molecules in an object, so mass is rarely determined in this manner. If we consider situations in which \vec{g} is a constant on Earth, we see that weight \vec{w} is directly proportional to mass m , since $\vec{w} = m\vec{g}$, that is, the more massive an object is, the more it weighs. Operationally, the masses of objects are determined by comparison with the standard kilogram, as we discussed in [Units and Measurement](#). But by comparing an object on Earth with one on the Moon, we can easily see a variation in weight but not in mass. For instance, on Earth, a 5.0-kg object weighs 49 N; on the Moon, where g is 1.67 m/s^2 , the object weighs 8.4 N. However, the mass of the object is still 5.0 kg on the Moon.

Example:**Clearing a Field**

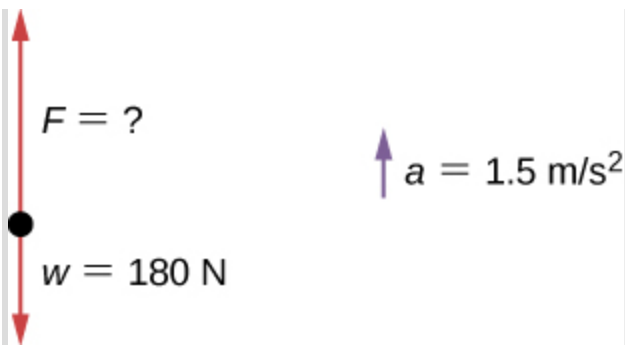
A farmer is lifting some moderately heavy rocks from a field to plant crops. He lifts a stone that weighs 40.0 lb. (about 180 N). What force does he apply if the stone accelerates at a rate of 1.5 m/s^2 ?

Strategy

We were given the weight of the stone, which we use in finding the net force on the stone. However, we also need to know its mass to apply Newton's second law, so we must apply the equation for weight, $w = mg$, to determine the mass.

Solution

No forces act in the horizontal direction, so we can concentrate on vertical forces, as shown in the following free-body diagram. We label the acceleration to the side; technically, it is not part of the free-body diagram, but it helps to remind us that the object accelerates upward (so the net force is upward).



Equation:

$$w = mg$$

$$m = \frac{w}{g} = \frac{180 \text{ N}}{9.8 \text{ m/s}^2} = 18 \text{ kg}$$

$$\sum F = ma$$

$$F - w = ma$$

$$F - 180 \text{ N} = (18 \text{ kg})(1.5 \text{ m/s}^2)$$

$$F - 180 \text{ N} = 27 \text{ N}$$

$$F = 207 \text{ N} = 210 \text{ N to two significant figures}$$

Significance

To apply Newton's second law as the primary equation in solving a problem, we sometimes have to rely on other equations, such as the one for weight or one of the kinematic equations, to complete the solution.

Note:

Exercise:

Problem:

Check Your Understanding For [\[link\]](#), find the acceleration when the farmer's applied force is 230.0 N.

Solution:

$$a = 2.78 \text{ m/s}^2$$

Note:

Can you avoid the boulder field and land safely just before your fuel runs out, as Neil Armstrong did in 1969? This [version of the classic video game](#) accurately simulates the real motion of the lunar lander, with the correct mass, thrust, fuel consumption rate, and lunar gravity. The real lunar lander is hard to control.

Summary

- Mass is the quantity of matter in a substance.
- The weight of an object is the net force on a falling object, or its gravitational force. The object experiences acceleration due to gravity.
- Some upward resistance force from the air acts on all falling objects on Earth, so they can never truly be in free fall.
- Careful distinctions must be made between free fall and weightlessness using the definition of weight as force due to gravity acting on an object of a certain mass.

Conceptual Questions

Exercise:**Problem:**

What is the relationship between weight and mass? Which is an intrinsic, unchanging property of a body?

Exercise:

Problem:

How much does a 70-kg astronaut weight in space, far from any celestial body? What is her mass at this location?

Solution:

The astronaut is truly weightless in the location described, because there is no large body (planet or star) nearby to exert a gravitational force. Her mass is 70 kg regardless of where she is located.

Exercise:

Problem: Which of the following statements is accurate?

- (a) Mass and weight are the same thing expressed in different units.
- (b) If an object has no weight, it must have no mass.
- (c) If the weight of an object varies, so must the mass.
- (d) Mass and inertia are different concepts.
- (e) Weight is always proportional to mass.

Exercise:**Problem:**

When you stand on Earth, your feet push against it with a force equal to your weight. Why doesn't Earth accelerate away from you?

Solution:

The force you exert (a contact force equal in magnitude to your weight) is small. Earth is extremely massive by comparison. Thus, the acceleration of Earth would be incredibly small. To see this, use Newton's second law to calculate the acceleration you would cause if your weight is 600.0 N and the mass of Earth is 6.00×10^{24} kg.

Exercise:

Problem: How would you give the value of \vec{g} in vector form?

Problems**Exercise:****Problem:**

The weight of an astronaut plus his space suit on the Moon is only 250 N. (a) How much does the suited astronaut weigh on Earth? (b) What is the mass on the Moon? On Earth?

Solution:

$$w_{\text{Moon}} = mg_{\text{Moon}}$$

a. $m = 150 \text{ kg}$; b. Mass does not change, so the suited

$$w_{\text{Earth}} = 1.5 \times 10^3 \text{ N}$$

astronaut's mass on both Earth and the Moon is 150 kg.

Exercise:**Problem:**

Suppose the mass of a fully loaded module in which astronauts take off from the Moon is $1.00 \times 10^4 \text{ kg}$. The thrust of its engines is $3.00 \times 10^4 \text{ N}$. (a) Calculate the module's magnitude of acceleration in a vertical takeoff from the Moon. (b) Could it lift off from Earth? If not, why not? If it could, calculate the magnitude of its acceleration.

Exercise:

Problem:

A rocket sled accelerates at a rate of 49.0 m/s^2 . Its passenger has a mass of 75.0 kg . (a) Calculate the horizontal component of the force the seat exerts against his body. Compare this with his weight using a ratio. (b) Calculate the direction and magnitude of the total force the seat exerts against his body.

Solution:

$$\begin{aligned} F_h &= 3.68 \times 10^3 \text{ N and} \\ \text{a. } w &= 7.35 \times 10^2 \text{ N} \quad ; \\ \frac{F_h}{w} &= 5.00 \text{ times greater than weight} \\ \text{b. } F_{\text{net}} &= 3750 \text{ N} \\ \theta &= 11.3^\circ \text{ from horizontal} \end{aligned}$$

Exercise:**Problem:**

Repeat the previous problem for a situation in which the rocket sled accelerates opposite to the motion at a rate of 201 m/s^2 . In this problem, the forces are exerted by the seat and the seat belt.

Exercise:**Problem:**

A body of mass 2.00 kg is pushed straight upward by a 25.0 N vertical force. What is its acceleration?

Solution:

$$\begin{aligned} w &= 19.6 \text{ N} \\ F_{\text{net}} &= 5.40 \text{ N} \\ F_{\text{net}} &= ma \Rightarrow a = 2.70 \text{ m/s}^2 \end{aligned}$$

Exercise:

Problem:

A car weighing 12,500 N starts from rest and accelerates to 83.0 km/h in 5.00 s. The friction force is 1350 N. Find the applied force produced by the engine.

Exercise:**Problem:**

A body with a mass of 10.0 kg is assumed to be in Earth's gravitational field with $g = 9.80 \text{ m/s}^2$. What is the net force on the body if there are no other external forces acting on the object?

Solution:

98 N

Exercise:**Problem:**

A fireman has mass m ; he hears the fire alarm and slides down the pole with acceleration a (which is less than g in magnitude). (a) Write an equation giving the vertical force he must apply to the pole. (b) If his mass is 90.0 kg and he accelerates at 5.00 m/s^2 , what is the magnitude of his applied force?

Exercise:**Problem:**

A baseball catcher is performing a stunt for a television commercial. He will catch a baseball (mass 145 g) dropped from a height of 60.0 m above his glove. His glove stops the ball in 0.0100 s. What is the force exerted by his glove on the ball?

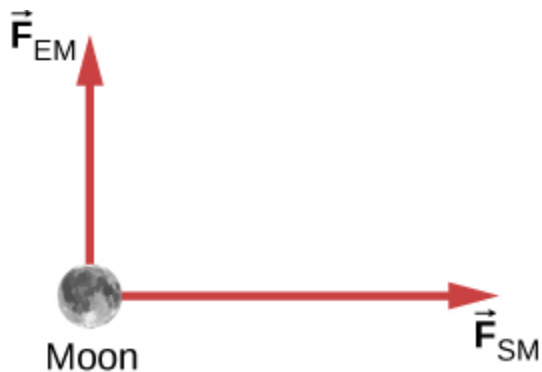
Solution:

497 N

Exercise:

Problem:

When the Moon is directly overhead at sunset, the force by Earth on the Moon, F_{EM} , is essentially at 90° to the force by the Sun on the Moon, F_{SM} , as shown below. Given that $F_{EM} = 1.98 \times 10^{20}$ N and $F_{SM} = 4.36 \times 10^{20}$ N, all other forces on the Moon are negligible, and the mass of the Moon is 7.35×10^{22} kg, determine the magnitude of the Moon's acceleration.



Glossary

free fall

situation in which the only force acting on an object is gravity

weight

force \vec{w} due to gravity acting on an object of mass m

Newton's Third Law

By the end of the section, you will be able to:

- State Newton's third law of motion
- Identify the action and reaction forces in different situations
- Apply Newton's third law to define systems and solve problems of motion

We have thus far considered force as a push or a pull; however, if you think about it, you realize that no push or pull ever occurs by itself. When you push on a wall, the wall pushes back on you. This brings us to **Newton's third law**.

Note:

Newton's Third Law of Motion

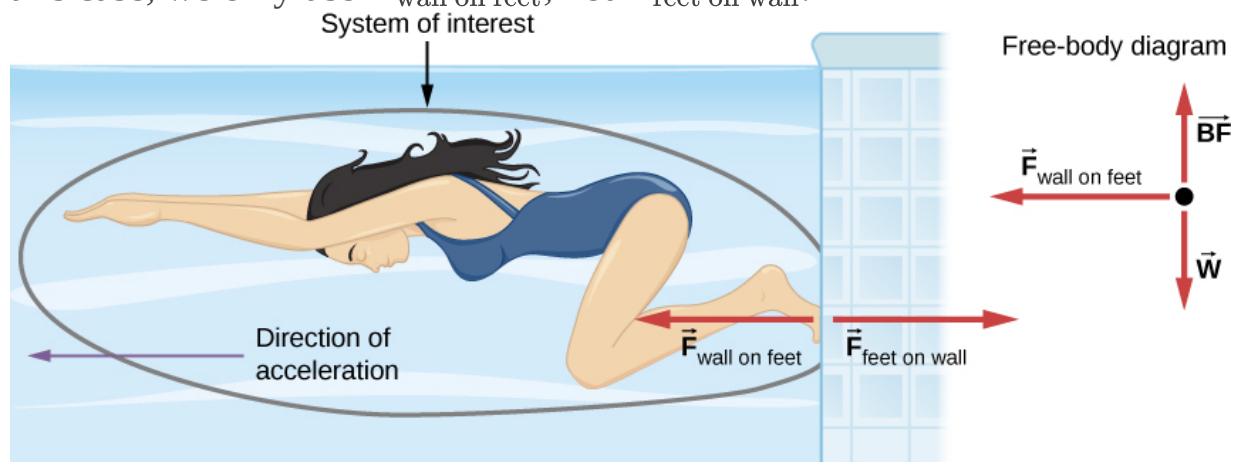
Whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts. Mathematically, if a body A exerts a force $\vec{\mathbf{F}}$ on body B , then B simultaneously exerts a force $-\vec{\mathbf{F}}$ on A , or in vector equation form,

Equation:

$$\vec{\mathbf{F}}_{AB} = -\vec{\mathbf{F}}_{BA}.$$

Newton's third law represents a certain symmetry in nature: Forces always occur in pairs, and one body cannot exert a force on another without experiencing a force itself. We sometimes refer to this law loosely as “action-reaction,” where the force exerted is the action and the force experienced as a consequence is the reaction. Newton's third law has practical uses in analyzing the origin of forces and understanding which forces are external to a system.

We can readily see Newton's third law at work by taking a look at how people move about. Consider a swimmer pushing off the side of a pool ([link](#)). She pushes against the wall of the pool with her feet and accelerates in the direction opposite that of her push. The wall has exerted an equal and opposite force on the swimmer. You might think that two equal and opposite forces would cancel, but they do not *because they act on different systems*. In this case, there are two systems that we could investigate: the swimmer and the wall. If we select the swimmer to be the system of interest, as in the figure, then $F_{\text{wall on feet}}$ is an external force on this system and affects its motion. The swimmer moves in the direction of this force. In contrast, the force $F_{\text{feet on wall}}$ acts on the wall, not on our system of interest. Thus, $F_{\text{feet on wall}}$ does not directly affect the motion of the system and does not cancel $F_{\text{wall on feet}}$. The swimmer pushes in the direction opposite that in which she wishes to move. The reaction to her push is thus in the desired direction. In a free-body diagram, such as the one shown in [link](#), we never include both forces of an action-reaction pair; in this case, we only use $F_{\text{wall on feet}}$, not $F_{\text{feet on wall}}$.

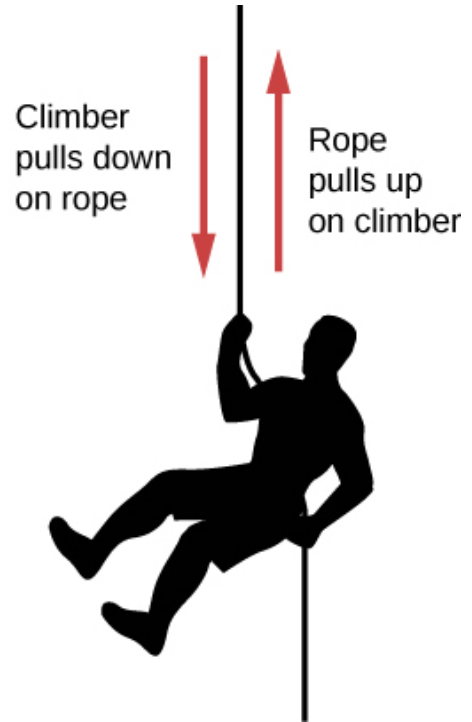


When the swimmer exerts a force on the wall, she accelerates in the opposite direction; in other words, the net external force on her is in the direction opposite of $F_{\text{feet on wall}}$. This opposition occurs because, in accordance with Newton's third law, the wall exerts a force $F_{\text{wall on feet}}$ on the swimmer that is equal in magnitude but in the direction opposite to the one she exerts on it. The line around the swimmer indicates the system of interest. Thus, the free-body diagram shows only $F_{\text{wall on feet}}$, w (the gravitational force), and BF , which is the buoyant force of the water supporting the swimmer's weight. The

vertical forces w and BF cancel because there is no vertical acceleration.

Other examples of Newton's third law are easy to find:

- As a professor paces in front of a whiteboard, he exerts a force backward on the floor. The floor exerts a reaction force forward on the professor that causes him to accelerate forward.
- A car accelerates forward because the ground pushes forward on the drive wheels, in reaction to the drive wheels pushing backward on the ground. You can see evidence of the wheels pushing backward when tires spin on a gravel road and throw the rocks backward.
- Rockets move forward by expelling gas backward at high velocity. This means the rocket exerts a large backward force on the gas in the rocket combustion chamber; therefore, the gas exerts a large reaction force forward on the rocket. This reaction force, which pushes a body forward in response to a backward force, is called **thrust**. It is a common misconception that rockets propel themselves by pushing on the ground or on the air behind them. They actually work better in a vacuum, where they can more readily expel the exhaust gases.
- Helicopters create lift by pushing air down, thereby experiencing an upward reaction force.
- Birds and airplanes also fly by exerting force on the air in a direction opposite that of whatever force they need. For example, the wings of a bird force air downward and backward to get lift and move forward.
- An octopus propels itself in the water by ejecting water through a funnel from its body, similar to a jet ski.
- When a person pulls down on a vertical rope, the rope pulls up on the person ([\[link\]](#)).



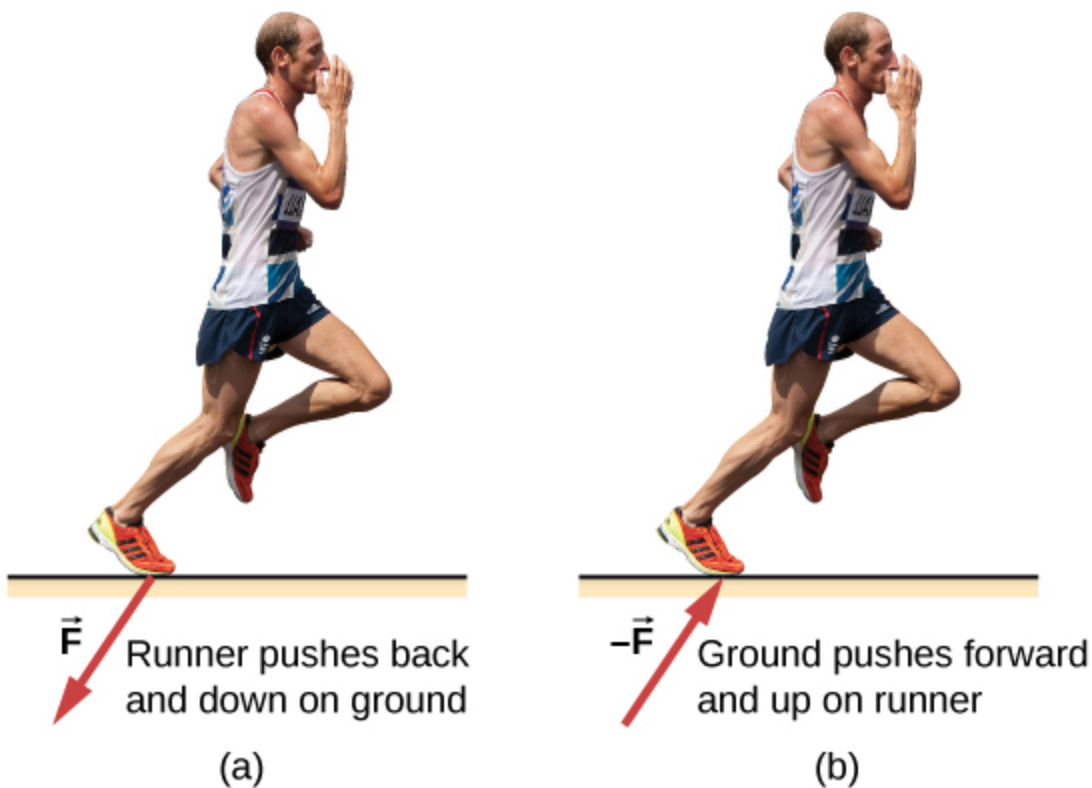
When the mountain climber pulls down on the rope, the rope pulls up on the mountain climber. (credit left: modification of work by Cristian Bortes)

There are two important features of Newton's third law. First, the forces exerted (the action and reaction) are always equal in magnitude but opposite in direction. Second, these forces are acting on different bodies or systems: A 's force acts on B and B 's force acts on A . In other words, the two forces are distinct forces that do not act on the same body. Thus, they do not cancel each other.

For the situation shown in [\[link\]](#), the third law indicates that because the chair is pushing upward on the boy with force \vec{C} , he is pushing downward on the chair with force $-\vec{C}$. Similarly, he is pushing downward with forces $-\vec{F}$ and $-\vec{T}$ on the floor and table, respectively. Finally, since Earth pulls downward on the boy with force \vec{w} , he pulls upward on Earth with force $-\vec{w}$. If that student were to angrily pound the table in frustration, he would

quickly learn the painful lesson (avoidable by studying Newton's laws) that the table hits back just as hard.

A person who is walking or running applies Newton's third law instinctively. For example, the runner in [\[link\]](#) pushes backward on the ground so that it pushes him forward.



The runner experiences Newton's third law. (a) A force is exerted by the runner on the ground. (b) The reaction force of the ground on the runner pushes him forward. (credit "runner": modification of work by "Greenwich Photography"/Flickr)

Example:
Forces on a Stationary Object

The package in [\[link\]](#) is sitting on a scale. The forces on the package are \vec{S} , which is due to the scale, and $-\vec{w}$, which is due to Earth's gravitational field. The reaction forces that the package exerts are $-\vec{S}$ on the scale and \vec{w} on Earth. Because the package is not accelerating, application of the second law yields

Equation:

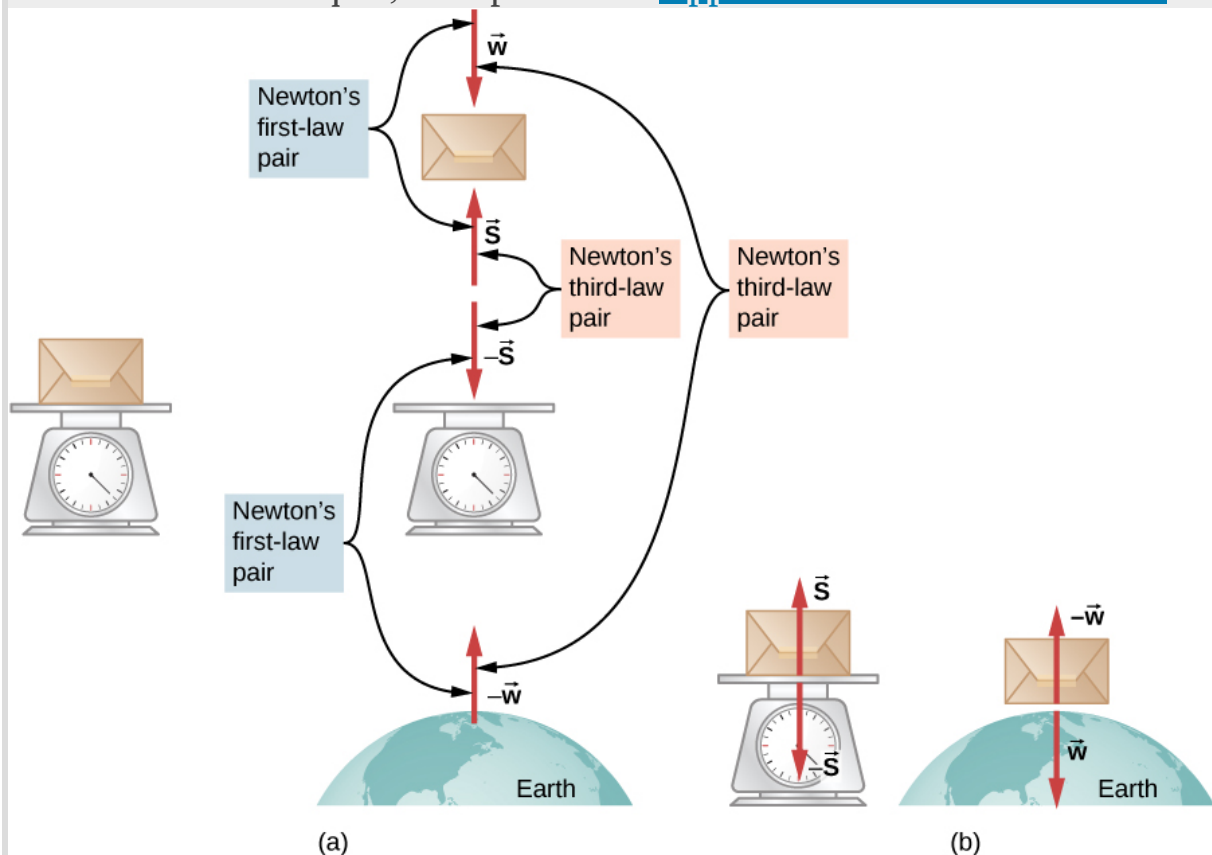
$$\vec{S} - \vec{w} = m\vec{a} = \vec{0},$$

so

Equation:

$$\vec{S} = \vec{w}.$$

Thus, the scale reading gives the magnitude of the package's weight. However, the scale does not measure the weight of the package; it measures the force $-\vec{S}$ on its surface. If the system is accelerating, \vec{S} and $-\vec{w}$ would not be equal, as explained in [Applications of Newton's Laws](#).

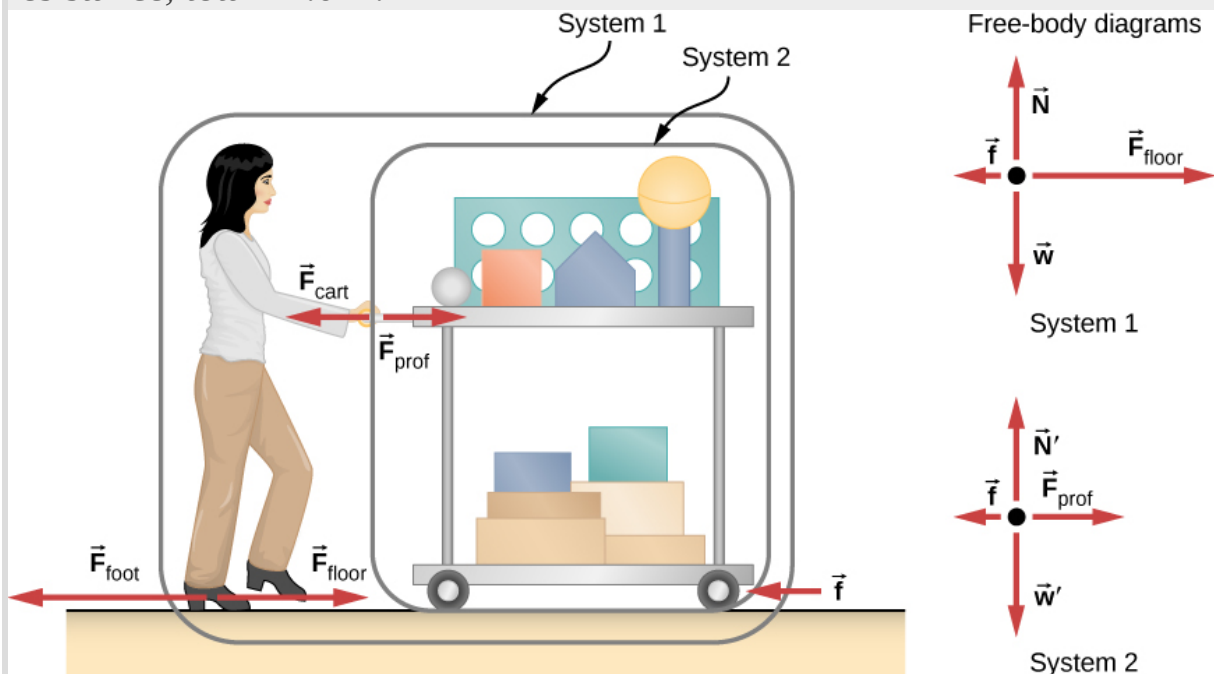


- (a) The forces on a package sitting on a scale, along with their reaction forces. The force \vec{w} is the weight of the package (the force due to Earth's gravity) and \vec{S} is the force of the scale on the package.
- (b) Isolation of the package-scale system and the package-Earth system makes the action and reaction pairs clear.

Example:

Getting Up to Speed: Choosing the Correct System

A physics professor pushes a cart of demonstration equipment to a lecture hall ([link](#)). Her mass is 65.0 kg, the cart's mass is 12.0 kg, and the equipment's mass is 7.0 kg. Calculate the acceleration produced when the professor exerts a backward force of 150 N on the floor. All forces opposing the motion, such as friction on the cart's wheels and air resistance, total 24.0 N.



A professor pushes the cart with her demonstration equipment. The lengths of the arrows are proportional to the magnitudes of the forces

(except for \vec{f} , because it is too small to draw to scale). System 1 is appropriate for this example, because it asks for the acceleration of the entire group of objects. Only \vec{F}_{floor} and \vec{f} are external forces acting on System 1 along the line of motion. All other forces either cancel or act on the outside world. System 2 is chosen for the next example so that \vec{F}_{prof} is an external force and enters into Newton's second law. The free-body diagrams, which serve as the basis for Newton's second law, vary with the system chosen.

Strategy

Since they accelerate as a unit, we define the system to be the professor, cart, and equipment. This is System 1 in [\[link\]](#). The professor pushes backward with a force F_{foot} of 150 N. According to Newton's third law, the floor exerts a forward reaction force F_{floor} of 150 N on System 1. Because all motion is horizontal, we can assume there is no net force in the vertical direction. Therefore, the problem is one-dimensional along the horizontal direction. As noted, friction f opposes the motion and is thus in the opposite direction of F_{floor} . We do not include the forces F_{prof} or F_{cart} because these are internal forces, and we do not include F_{foot} because it acts on the floor, not on the system. There are no other significant forces acting on System 1. If the net external force can be found from all this information, we can use Newton's second law to find the acceleration as requested. See the free-body diagram in the figure.

Solution

Newton's second law is given by

Equation:

$$a = \frac{F_{\text{net}}}{m}.$$

The net external force on System 1 is deduced from [\[link\]](#) and the preceding discussion to be

Equation:

$$F_{\text{net}} = F_{\text{floor}} - f = 150 \text{ N} - 24.0 \text{ N} = 126 \text{ N}.$$

The mass of System 1 is

Equation:

$$m = (65.0 + 12.0 + 7.0) \text{ kg} = 84 \text{ kg}.$$

These values of F_{net} and m produce an acceleration of

Equation:

$$a = \frac{F_{\text{net}}}{m} = \frac{126 \text{ N}}{84 \text{ kg}} = 1.5 \text{ m/s}^2.$$

Significance

None of the forces between components of System 1, such as between the professor's hands and the cart, contribute to the net external force because they are internal to System 1. Another way to look at this is that forces between components of a system cancel because they are equal in magnitude and opposite in direction. For example, the force exerted by the professor on the cart results in an equal and opposite force back on the professor. In this case, both forces act on the same system and therefore cancel. Thus, internal forces (between components of a system) cancel. Choosing System 1 was crucial to solving this problem.

Example:

Force on the Cart: Choosing a New System

Calculate the force the professor exerts on the cart in [\[link\]](#), using data from the previous example if needed.

Strategy

If we define the system of interest as the cart plus the equipment (System 2 in [\[link\]](#)), then the net external force on System 2 is the force the professor exerts on the cart minus friction. The force she exerts on the cart, F_{prof} , is an external force acting on System 2. F_{prof} was internal to System 1, but it is external to System 2 and thus enters Newton's second law for this system.

Solution

Newton's second law can be used to find F_{prof} . We start with

Equation:

$$a = \frac{F_{\text{net}}}{m}.$$

The magnitude of the net external force on System 2 is

Equation:

$$F_{\text{net}} = F_{\text{prof}} - f.$$

We solve for F_{prof} , the desired quantity:

Equation:

$$F_{\text{prof}} = F_{\text{net}} + f.$$

The value of f is given, so we must calculate net F_{net} . That can be done because both the acceleration and the mass of System 2 are known. Using Newton's second law, we see that

Equation:

$$F_{\text{net}} = ma,$$

where the mass of System 2 is 19.0 kg ($m = 12.0 \text{ kg} + 7.0 \text{ kg}$) and its acceleration was found to be $a = 1.5 \text{ m/s}^2$ in the previous example. Thus,

Equation:

$$F_{\text{net}} = ma = (19.0 \text{ kg}) (1.5 \text{ m/s}^2) = 29 \text{ N}.$$

Now we can find the desired force:

Equation:

$$F_{\text{prof}} = F_{\text{net}} + f = 29 \text{ N} + 24.0 \text{ N} = 53 \text{ N}.$$

Significance

This force is significantly less than the 150-N force the professor exerted backward on the floor. Not all of that 150-N force is transmitted to the cart; some of it accelerates the professor. The choice of a system is an important

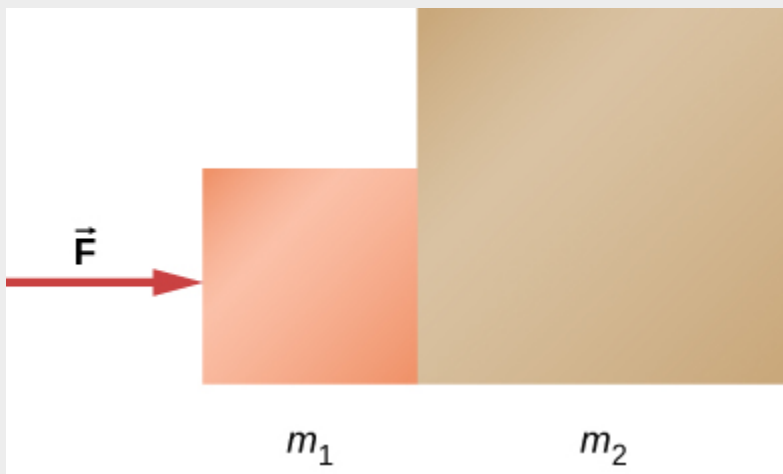
analytical step both in solving problems and in thoroughly understanding the physics of the situation (which are not necessarily the same things).

Note:

Exercise:

Problem:

Check Your Understanding Two blocks are at rest and in contact on a frictionless surface as shown below, with $m_1 = 2.0$ kg, $m_2 = 6.0$ kg, and applied force 24 N. (a) Find the acceleration of the system of blocks. (b) Suppose that the blocks are later separated. What force will give the second block, with the mass of 6.0 kg, the same acceleration as the system of blocks?



Solution:

a. 3.0 m/s^2 ; b. 18 N

Note:

View this [video](#) to watch examples of action and reaction.

Note:

View this [video](#) to watch examples of Newton's laws and internal and external forces.

Summary

- Newton's third law of motion represents a basic symmetry in nature, with an experienced force equal in magnitude and opposite in direction to an exerted force.
- Two equal and opposite forces do not cancel because they act on different systems.
- Action-reaction pairs include a swimmer pushing off a wall, helicopters creating lift by pushing air down, and an octopus propelling itself forward by ejecting water from its body. Rockets, airplanes, and cars are pushed forward by a thrust reaction force.
- Choosing a system is an important analytical step in understanding the physics of a problem and solving it.

Conceptual Questions

Exercise:**Problem:**

Identify the action and reaction forces in the following situations: (a) Earth attracts the Moon, (b) a boy kicks a football, (c) a rocket accelerates upward, (d) a car accelerates forward, (e) a high jumper leaps, and (f) a bullet is shot from a gun.

Solution:

a. action: Earth pulls on the Moon, reaction: Moon pulls on Earth; b. action: foot applies force to ball, reaction: ball applies force to foot; c. action: rocket pushes on gas, reaction: gas pushes back on rocket; d. action: car tires push backward on road, reaction: road pushes forward on tires; e. action: jumper pushes down on ground, reaction: ground

pushes up on jumper; f. action: gun pushes forward on bullet, reaction: bullet pushes backward on gun.

Exercise:

Problem:

Suppose that you are holding a cup of coffee in your hand. Identify all forces on the cup and the reaction to each force.

Exercise:

Problem:

(a) Why does an ordinary rifle recoil (kick backward) when fired? (b) The barrel of a recoilless rifle is open at both ends. Describe how Newton's third law applies when one is fired. (c) Can you safely stand close behind one when it is fired?

Solution:

a. The rifle (the shell supported by the rifle) exerts a force to expel the bullet; the reaction to this force is the force that the bullet exerts on the rifle (shell) in opposite direction. b. In a recoilless rifle, the shell is not secured in the rifle; hence, as the bullet is pushed to move forward, the shell is pushed to eject from the opposite end of the barrel. c. It is not safe to stand behind a recoilless rifle.

Problems

Exercise:

Problem:

(a) What net external force is exerted on a 1100.0-kg artillery shell fired from a battleship if the shell is accelerated at $2.40 \times 10^4 \text{ m/s}^2$?
(b) What is the magnitude of the force exerted on the ship by the artillery shell, and why?

Solution:

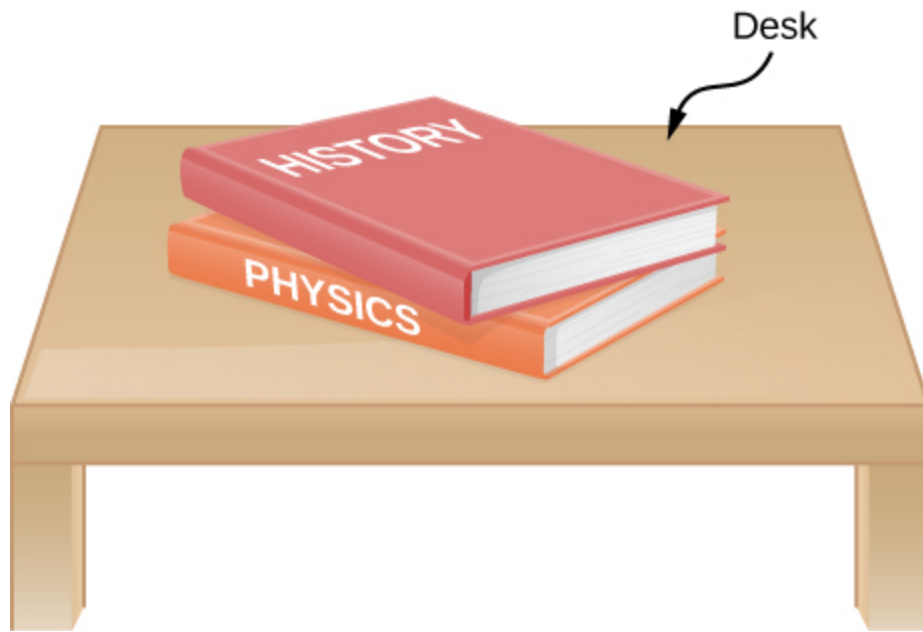
a. $F_{\text{net}} = 2.64 \times 10^7 \text{ N}$; b. The force exerted on the ship is also $2.64 \times 10^7 \text{ N}$ because it is opposite the shell's direction of motion.

Exercise:**Problem:**

A brave but inadequate rugby player is being pushed backward by an opposing player who is exerting a force of 800.0 N on him. The mass of the losing player plus equipment is 90.0 kg, and he is accelerating backward at 1.20 m/s^2 . (a) What is the force of friction between the losing player's feet and the grass? (b) What force does the winning player exert on the ground to move forward if his mass plus equipment is 110.0 kg?

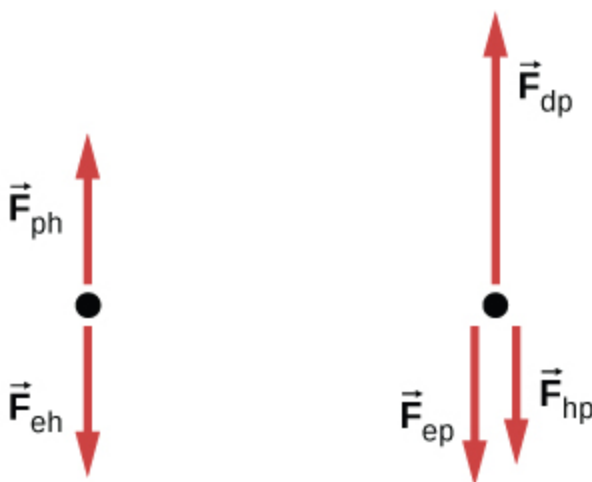
Exercise:**Problem:**

A history book is lying on top of a physics book on a desk, as shown below; a free-body diagram is also shown. The history and physics books weigh 14 N and 18 N, respectively. Identify each force on each book with a double subscript notation (for instance, the contact force of the history book pressing against physics book can be described as \vec{F}_{HP}), and determine the value of each of these forces, explaining the process used.



History book

Physics book



Solution:

Because the weight of the history book is the force exerted by Earth on the history book, we represent it as $\vec{\mathbf{F}}_{\text{EH}} = -14\hat{\mathbf{j}} \text{ N}$. Aside from this, the history book interacts only with the physics book. Because the acceleration of the history book is zero, the net force on it is zero by Newton's second law: $\vec{\mathbf{F}}_{\text{PH}} + \vec{\mathbf{F}}_{\text{EH}} = \vec{\mathbf{0}}$, where $\vec{\mathbf{F}}_{\text{PH}}$ is the force exerted by the physics book on the history book. Thus,

$\vec{\mathbf{F}}_{\text{PH}} = -\vec{\mathbf{F}}_{\text{EH}} = -(-14\hat{\mathbf{j}}) \text{ N} = 14\hat{\mathbf{j}} \text{ N}$. We find that the physics book exerts an upward force of magnitude 14 N on the history book. The physics book has three forces exerted on it: $\vec{\mathbf{F}}_{\text{EP}}$ due to Earth, $\vec{\mathbf{F}}_{\text{HP}}$ due to the history book, and $\vec{\mathbf{F}}_{\text{DP}}$ due to the desktop. Since the physics book weighs 18 N, $\vec{\mathbf{F}}_{\text{EP}} = -18\hat{\mathbf{j}} \text{ N}$. From Newton's third law, $\vec{\mathbf{F}}_{\text{HP}} = -\vec{\mathbf{F}}_{\text{PH}}$, so $\vec{\mathbf{F}}_{\text{HP}} = -14\hat{\mathbf{j}} \text{ N}$. Newton's second law applied to the physics book gives $\sum \vec{\mathbf{F}} = \vec{\mathbf{0}}$, or $\vec{\mathbf{F}}_{\text{DP}} + \vec{\mathbf{F}}_{\text{EP}} + \vec{\mathbf{F}}_{\text{HP}} = \vec{\mathbf{0}}$, so $\vec{\mathbf{F}}_{\text{DP}} = -(-18\hat{\mathbf{j}}) - (-14\hat{\mathbf{j}}) = 32\hat{\mathbf{j}} \text{ N}$. The desk exerts an upward force of 32 N on the physics book. To arrive at this solution, we apply Newton's second law twice and Newton's third law once.

Exercise:

Problem:

A truck collides with a car, and during the collision, the net force on each vehicle is essentially the force exerted by the other. Suppose the mass of the car is 550 kg, the mass of the truck is 2200 kg, and the magnitude of the truck's acceleration is 10 m/s^2 . Find the magnitude of the car's acceleration.

Glossary

Newton's third law of motion

whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts

thrust

reaction force that pushes a body forward in response to a backward force

Common Forces

By the end of the section, you will be able to:

- Define normal and tension forces
- Distinguish between real and fictitious forces
- Apply Newton's laws of motion to solve problems involving a variety of forces

Forces are given many names, such as push, pull, thrust, and weight. Traditionally, forces have been grouped into several categories and given names relating to their source, how they are transmitted, or their effects. Several of these categories are discussed in this section, together with some interesting applications. Further examples of forces are discussed later in this text.

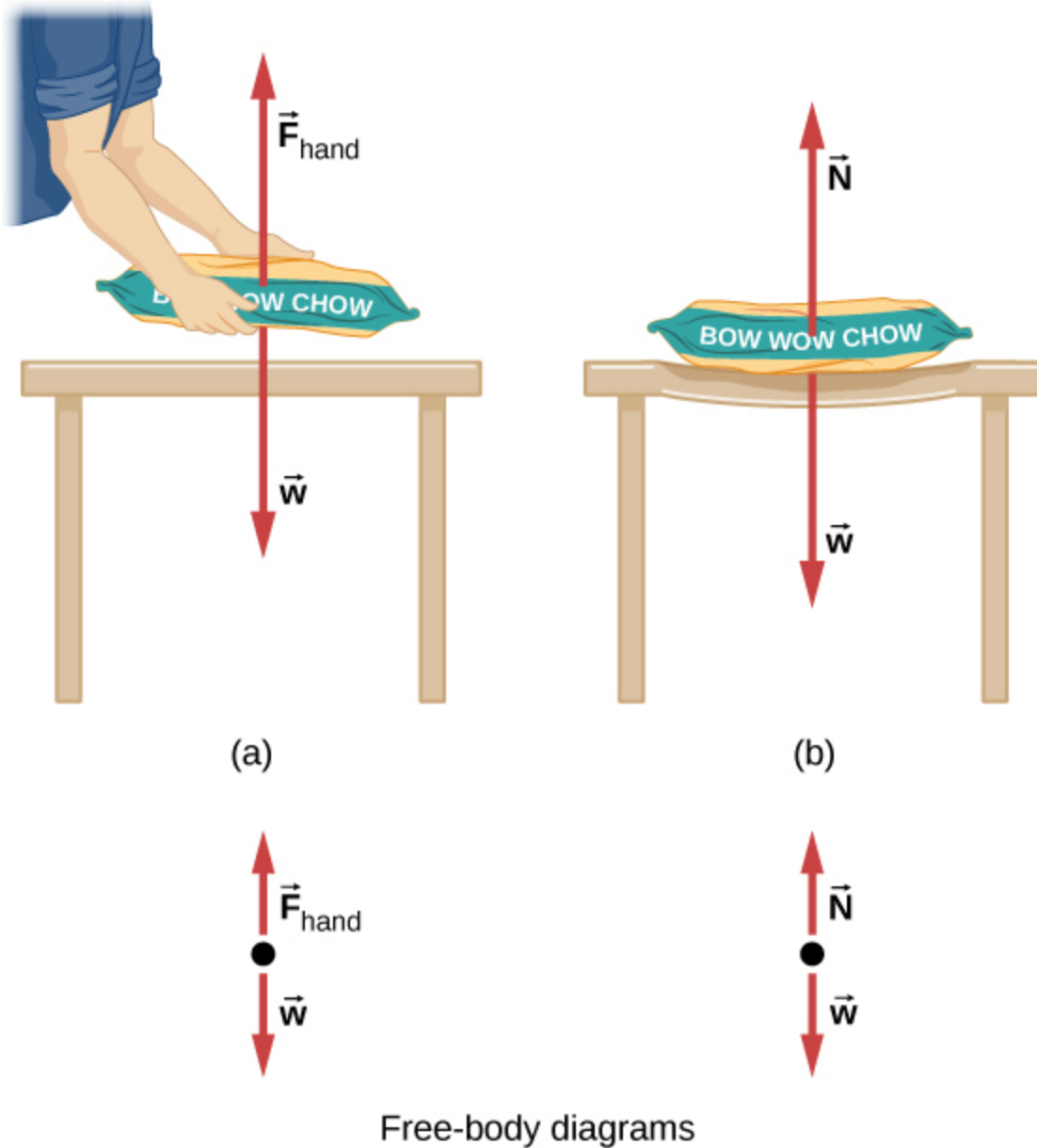
A Catalog of Forces: Normal, Tension, and Other Examples of Forces

A catalog of forces will be useful for reference as we solve various problems involving force and motion. These forces include normal force, tension, friction, and spring force.

Normal force

Weight (also called the force of gravity) is a pervasive force that acts at all times and must be counteracted to keep an object from falling. You must support the weight of a heavy object by pushing up on it when you hold it stationary, as illustrated in [\[link\]](#)(a). But how do inanimate objects like a table support the weight of a mass placed on them, such as shown in [\[link\]](#) (b)? When the bag of dog food is placed on the table, the table sags slightly under the load. This would be noticeable if the load were placed on a card table, but even a sturdy oak table deforms when a force is applied to it. Unless an object is deformed beyond its limit, it will exert a restoring force much like a deformed spring (or a trampoline or diving board). The greater the deformation, the greater the restoring force. Thus, when the load is

placed on the table, the table sags until the restoring force becomes as large as the weight of the load. At this point, the net external force on the load is zero. That is the situation when the load is stationary on the table. The table sags quickly and the sag is slight, so we do not notice it. But it is similar to the sagging of a trampoline when you climb onto it.



(a) The person holding the bag of dog food must supply an

upward force \vec{F}_{hand} equal in magnitude and opposite in direction to the weight of the food \vec{w} so that it doesn't drop to the ground. (b) The card table sags when the dog food is placed on it, much like a stiff trampoline. Elastic restoring forces in the table grow as it sags until they supply a force \vec{N} equal in magnitude and opposite in direction to the weight of the load.

We must conclude that whatever supports a load, be it animate or not, must supply an upward force equal to the weight of the load, as we assumed in a few of the previous examples. If the force supporting the weight of an object, or a load, is perpendicular to the surface of contact between the load and its support, this force is defined as a **normal force** and here is given by the symbol \vec{N} . (This is not the newton unit for force, or N.) The word *normal* means perpendicular to a surface. This means that the normal force experienced by an object resting on a horizontal surface can be expressed in vector form as follows:

Note:

Equation:

$$\vec{N} = -m\vec{g}.$$

In scalar form, this becomes

Note:

Equation:

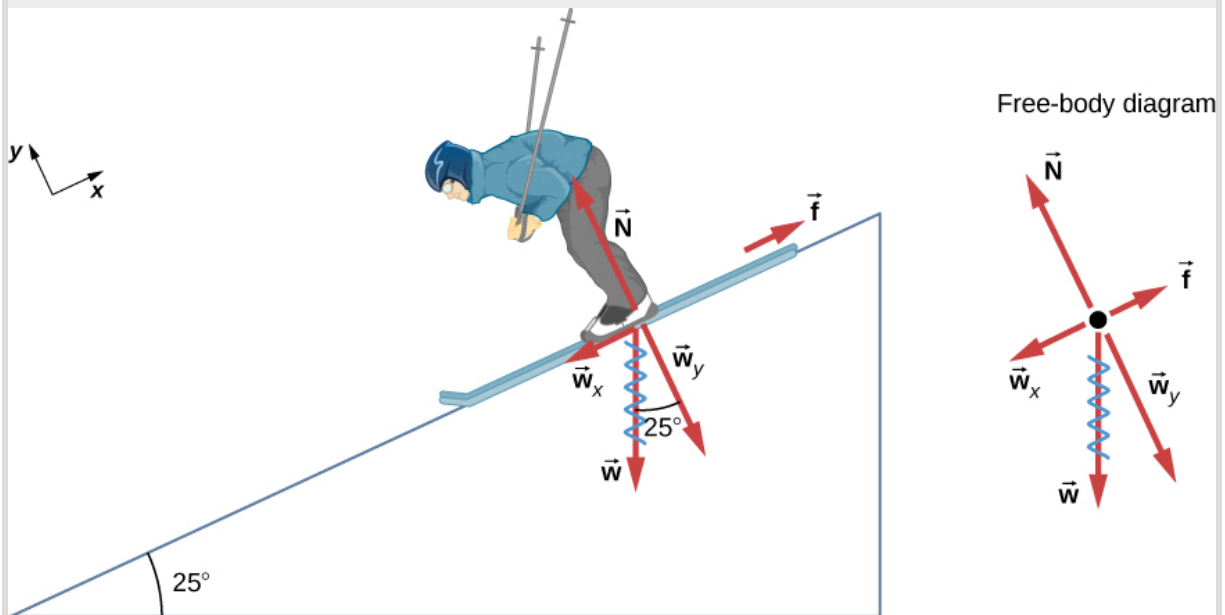
$$N = mg.$$

The normal force can be less than the object's weight if the object is on an incline.

Example:

Weight on an Incline

Consider the skier on the slope in [\[link\]](#). Her mass including equipment is 60.0 kg. (a) What is her acceleration if friction is negligible? (b) What is her acceleration if friction is 45.0 N?



Since the acceleration is parallel to the slope and acting down the slope, it is most convenient to project all forces onto a coordinate system where one axis is parallel to the slope and the other is perpendicular to it (axes shown to the left of the skier). \vec{N} is perpendicular to the slope and \vec{f} is parallel to the slope, but \vec{w} has components along both axes, namely, w_y and w_x . Here, \vec{w} has a squiggly line to show that it has been replaced by these components. The force \vec{N} is equal in magnitude to w_y , so there is no acceleration

perpendicular to the slope, but f is less than w_x , so there is a downslope acceleration (along the axis parallel to the slope).

Strategy

This is a two-dimensional problem, since not all forces on the skier (the system of interest) are parallel. The approach we have used in two-dimensional kinematics also works well here. Choose a convenient coordinate system and project the vectors onto its axes, creating two one-dimensional problems to solve. The most convenient coordinate system for motion on an incline is one that has one coordinate parallel to the slope and one perpendicular to the slope. (Motions along mutually perpendicular axes are independent.) We use x and y for the parallel and perpendicular directions, respectively. This choice of axes simplifies this type of problem, because there is no motion perpendicular to the slope and the acceleration is downslope. Regarding the forces, friction is drawn in opposition to motion (friction always opposes forward motion) and is always parallel to the slope, w_x is drawn parallel to the slope and downslope (it causes the motion of the skier down the slope), and w_y is drawn as the component of weight perpendicular to the slope. Then, we can consider the separate problems of forces parallel to the slope and forces perpendicular to the slope.

Solution

The magnitude of the component of weight parallel to the slope is

Equation:

$$w_x = w \sin 25^\circ = mg \sin 25^\circ ,$$

and the magnitude of the component of the weight perpendicular to the slope is

Equation:

$$w_y = w \cos 25^\circ = mg \cos 25^\circ .$$

a. Neglect friction. Since the acceleration is parallel to the slope, we need only consider forces parallel to the slope. (Forces perpendicular to the slope add to zero, since there is no acceleration in that direction.) The forces parallel to the slope are the component of the skier's weight parallel

to slope w_x and friction f . Using Newton's second law, with subscripts to denote quantities parallel to the slope,

Equation:

$$a_x = \frac{F_{\text{net } x}}{m}$$

where $F_{\text{net } x} = w_x - mg \sin 25^\circ$, assuming no friction for this part. Therefore,

Equation:

$$a_x = \frac{F_{\text{net } x}}{m} = \frac{mg \sin 25^\circ}{m} = g \sin 25^\circ \\ (9.80 \text{ m/s}^2) (0.4226) = 4.14 \text{ m/s}^2$$

is the acceleration.

b. Include friction. We have a given value for friction, and we know its direction is parallel to the slope and it opposes motion between surfaces in contact. So the net external force is

Equation:

$$F_{\text{net } x} = w_x - f.$$

Substituting this into Newton's second law, $a_x = F_{\text{net } x}/m$, gives

Equation:

$$a_x = \frac{F_{\text{net } x}}{m} = \frac{w_x - f}{m} = \frac{mg \sin 25^\circ - f}{m}.$$

We substitute known values to obtain

Equation:

$$a_x = \frac{(60.0 \text{ kg}) (9.80 \text{ m/s}^2) (0.4226) - 45.0 \text{ N}}{60.0 \text{ kg}}.$$

This gives us

Equation:

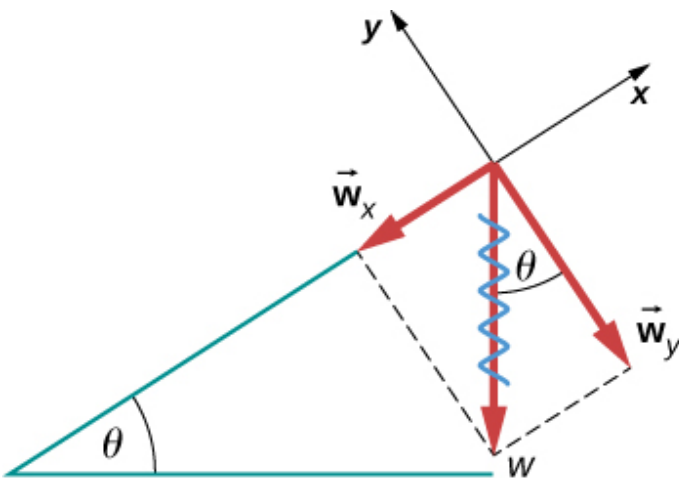
$$a_x = 3.39 \text{ m/s}^2,$$

which is the acceleration parallel to the incline when there is 45.0 N of opposing friction.

Significance

Since friction always opposes motion between surfaces, the acceleration is smaller when there is friction than when there is none. It is a general result that if friction on an incline is negligible, then the acceleration down the incline is $a = g \sin \theta$, regardless of mass. As discussed previously, all objects fall with the same acceleration in the absence of air resistance. Similarly, all objects, regardless of mass, slide down a frictionless incline with the same acceleration (if the angle is the same).

When an object rests on an incline that makes an angle θ with the horizontal, the force of gravity acting on the object is divided into two components: a force acting perpendicular to the plane, w_y , and a force acting parallel to the plane, w_x ([\[link\]](#)). The normal force \vec{N} is typically equal in magnitude and opposite in direction to the perpendicular component of the weight w_y . The force acting parallel to the plane, w_x , causes the object to accelerate down the incline.



$$w_x = w \sin(\theta) = mg \sin(\theta)$$
$$w_y = w \cos(\theta) = mg \cos(\theta)$$

An object rests on an incline that makes an angle θ with the horizontal.

Be careful when resolving the weight of the object into components. If the incline is at an angle θ to the horizontal, then the magnitudes of the weight components are

Equation:

$$w_x = w \sin \theta = mg \sin \theta$$

and

Equation:

$$w_y = w \cos \theta = mg \cos \theta.$$

We use the second equation to write the normal force experienced by an object resting on an inclined plane:

Note:

Equation:

$$N = mg \cos \theta.$$

Instead of memorizing these equations, it is helpful to be able to determine them from reason. To do this, we draw the right angle formed by the three weight vectors. The angle θ of the incline is the same as the angle formed between w and w_y . Knowing this property, we can use trigonometry to determine the magnitude of the weight components:

Equation:

$$\cos \theta = \frac{w_y}{w}, \quad w_y = w \cos \theta = mg \cos \theta$$

$$\sin \theta = \frac{w_x}{w}, \quad w_x = w \sin \theta = mg \sin \theta.$$

Note:

Exercise:

Problem:

Check Your Understanding A force of 1150 N acts parallel to a ramp to push a 250-kg gun safe into a moving van. The ramp is frictionless and inclined at 17° . (a) What is the acceleration of the safe up the ramp? (b) If we consider friction in this problem, with a friction force of 120 N, what is the acceleration of the safe?

Solution:

a. 1.7 m/s^2 ; b. 1.3 m/s^2

Tension

A **tension** is a force along the length of a medium; in particular, it is a pulling force that acts along a stretched flexible connector, such as a rope or cable. The word “tension” comes from a Latin word meaning “to stretch.” Not coincidentally, the flexible cords that carry muscle forces to other parts of the body are called *tendons*.

Any flexible connector, such as a string, rope, chain, wire, or cable, can only exert a pull parallel to its length; thus, a force carried by a flexible connector is a tension with a direction parallel to the connector. Tension is a pull in a connector. Consider the phrase: “You can’t push a rope.” Instead, tension force pulls outward along the two ends of a rope.

Consider a person holding a mass on a rope, as shown in [\[link\]](#). If the 5.00-kg mass in the figure is stationary, then its acceleration is zero and the net force is zero. The only external forces acting on the mass are its weight and the tension supplied by the rope. Thus,

Equation:

$$F_{\text{net}} = T - w = 0,$$

where T and w are the magnitudes of the tension and weight, respectively, and their signs indicate direction, with up being positive. As we proved using Newton's second law, the tension equals the weight of the supported mass:

Note:

Equation:

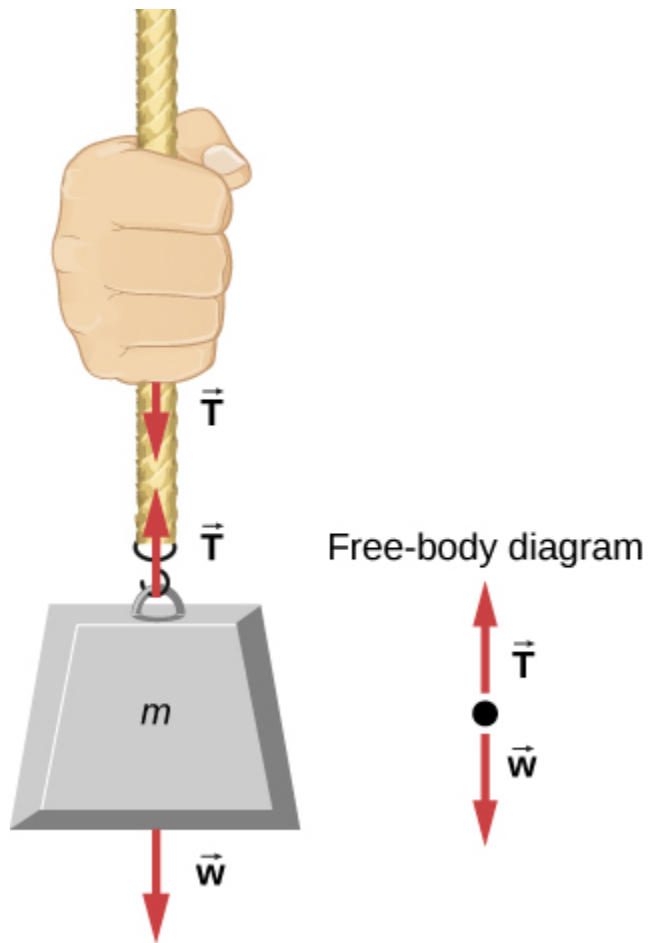
$$T = w = mg.$$

Thus, for a 5.00-kg mass (neglecting the mass of the rope), we see that

Equation:

$$T = mg = (5.00 \text{ kg}) (9.80 \text{ m/s}^2) = 49.0 \text{ N}.$$

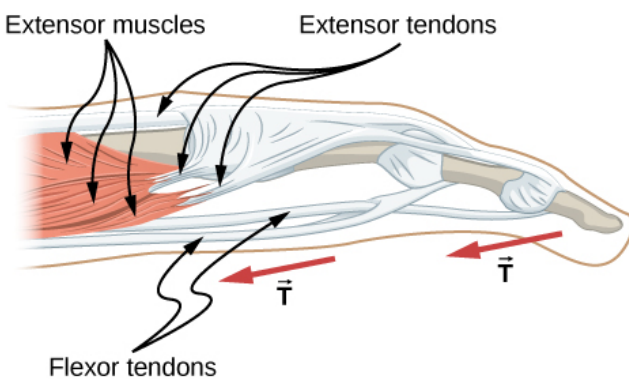
If we cut the rope and insert a spring, the spring would extend a length corresponding to a force of 49.0 N, providing a direct observation and measure of the tension force in the rope.



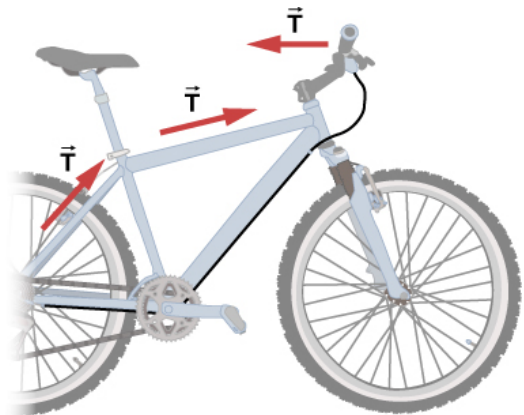
When a perfectly flexible connector (one requiring no force to bend it) such as this rope transmits a force \vec{T} , that force must be parallel to the length of the rope, as shown. By Newton's third law, the rope pulls with equal force but in opposite directions on the hand and the supported mass (neglecting the weight of the rope). The rope is the medium that carries the equal and opposite forces between the two objects. The tension anywhere in the rope between the

hand and the mass is equal. Once you have determined the tension in one location, you have determined the tension at all locations along the rope.

Flexible connectors are often used to transmit forces around corners, such as in a hospital traction system, a tendon, or a bicycle brake cable. If there is no friction, the tension transmission is undiminished; only its direction changes, and it is always parallel to the flexible connector, as shown in [\[link\]](#).



(a)

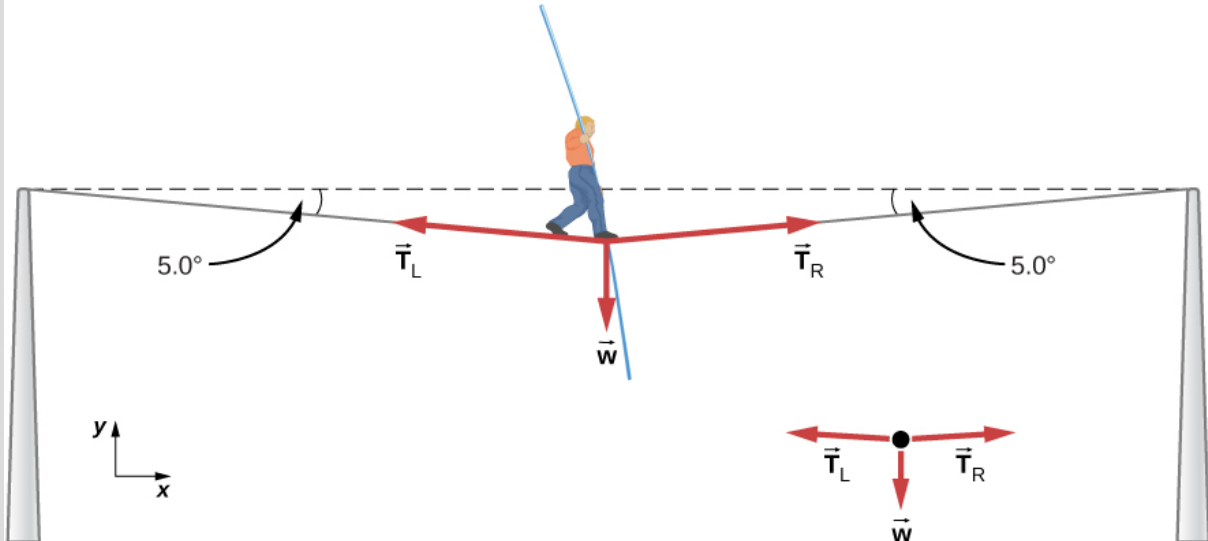


(b)

(a) Tendons in the finger carry force T from the muscles to other parts of the finger, usually changing the force's direction but not its magnitude (the tendons are relatively friction free). (b) The brake cable on a bicycle carries the tension T from the brake lever on the handlebars to the brake mechanism. Again, the direction but not the magnitude of T is changed.

Example:**What Is the Tension in a Tightrope?**

Calculate the tension in the wire supporting the 70.0-kg tightrope walker shown in [\[link\]](#).



The weight of a tightrope walker causes a wire to sag by 5.0° . The system of interest is the point in the wire at which the tightrope walker is standing.

Strategy

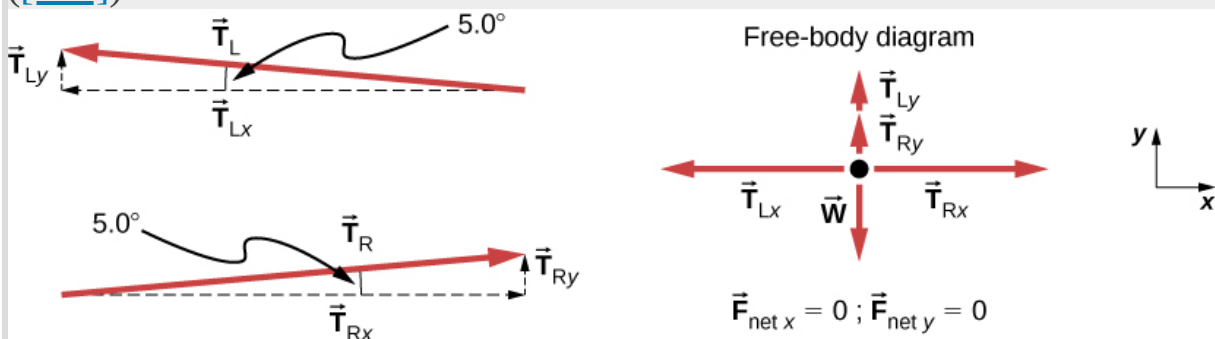
As you can see in [\[link\]](#), the wire is bent under the person's weight. Thus, the tension on either side of the person has an upward component that can support his weight. As usual, forces are vectors represented pictorially by arrows that have the same direction as the forces and lengths proportional to their magnitudes. The system is the tightrope walker, and the only external forces acting on him are his weight \vec{w} and the two tensions \vec{T}_L (left tension) and \vec{T}_R (right tension). It is reasonable to neglect the weight of the wire. The net external force is zero, because the system is static. We can use trigonometry to find the tensions. One conclusion is possible at the outset—we can see from [\[link\]](#)(b) that the magnitudes of the tensions T_L and T_R must be equal. We know this because there is no horizontal acceleration in the rope and the only forces acting to the left and right are

T_L and T_R . Thus, the magnitude of those horizontal components of the forces must be equal so that they cancel each other out.

Whenever we have two-dimensional vector problems in which no two vectors are parallel, the easiest method of solution is to pick a convenient coordinate system and project the vectors onto its axes. In this case, the best coordinate system has one horizontal axis (x) and one vertical axis (y).

Solution

First, we need to resolve the tension vectors into their horizontal and vertical components. It helps to look at a new free-body diagram showing all horizontal and vertical components of each force acting on the system ([\[link\]](#)).



When the vectors are projected onto vertical and horizontal axes, their components along these axes must add to zero, since the tightrope walker is stationary. The small angle results in T being much greater than w .

Consider the horizontal components of the forces (denoted with a subscript x):

Equation:

$$F_{\text{net } x} = T_{Rx} - T_{Lx}.$$

The net external horizontal force $F_{\text{net } x} = 0$, since the person is stationary. Thus,

Equation:

$$\begin{aligned} F_{\text{net } x} &= 0 = T_{Rx} - T_{Lx} \\ T_{Lx} &= T_{Rx}. \end{aligned}$$

Now observe [\[link\]](#). You can use trigonometry to determine the magnitude of T_L and T_R :

Equation:

$$\begin{aligned}\cos 5.0^\circ &= \frac{T_{Lx}}{T_L}, \quad T_{Lx} = T_L \cos 5.0^\circ \\ \cos 5.0^\circ &= \frac{T_{Rx}}{T_R}, \quad T_{Rx} = T_R \cos 5.0^\circ.\end{aligned}$$

Equating T_{Lx} and T_{Rx} :

Equation:

$$T_L \cos 5.0^\circ = T_R \cos 5.0^\circ.$$

Thus,

Equation:

$$T_L = T_R = T,$$

as predicted. Now, considering the vertical components (denoted by a subscript y), we can solve for T . Again, since the person is stationary, Newton's second law implies that $F_{\text{net } y} = 0$. Thus, as illustrated in the free-body diagram,

Equation:

$$F_{\text{net } y} = T_{Ly} + T_{Ry} - w = 0.$$

We can use trigonometry to determine the relationships among T_{Ly} , T_{Ry} , and T . As we determined from the analysis in the horizontal direction, $T_L = T_R = T$:

Equation:

$$\begin{aligned}\sin 5.0^\circ &= \frac{T_{Ly}}{T_L}, \quad T_{Ly} = T_L \sin 5.0^\circ = T \sin 5.0^\circ \\ \sin 5.0^\circ &= \frac{T_{Ry}}{T_R}, \quad T_{Ry} = T_R \sin 5.0^\circ = T \sin 5.0^\circ.\end{aligned}$$

Now we can substitute the values for T_{Ly} and T_{Ry} , into the net force equation in the vertical direction:

Equation:

$$\begin{aligned}F_{\text{net } y} &= T_{Ly} + T_{Ry} - w = 0 \\F_{\text{net } y} &= T \sin 5.0^\circ + T \sin 5.0^\circ - w = 0 \\2T \sin 5.0^\circ - w &= 0 \\2T \sin 5.0^\circ &= w\end{aligned}$$

and

Equation:

$$T = \frac{w}{2 \sin 5.0^\circ} = \frac{mg}{2 \sin 5.0^\circ},$$

so

Equation:

$$T = \frac{(70.0 \text{ kg}) (9.80 \text{ m/s}^2)}{2 (0.0872)},$$

and the tension is

Equation:

$$T = 3930 \text{ N}.$$

Significance

The vertical tension in the wire acts as a force that supports the weight of the tightrope walker. The tension is almost six times the 686-N weight of the tightrope walker. Since the wire is nearly horizontal, the vertical component of its tension is only a fraction of the tension in the wire. The large horizontal components are in opposite directions and cancel, so most of the tension in the wire is not used to support the weight of the tightrope walker.

If we wish to create a large tension, all we have to do is exert a force perpendicular to a taut flexible connector, as illustrated in [\[link\]](#). As we saw

in [\[link\]](#), the weight of the tightrope walker acts as a force perpendicular to the rope. We saw that the tension in the rope is related to the weight of the tightrope walker in the following way:

Equation:

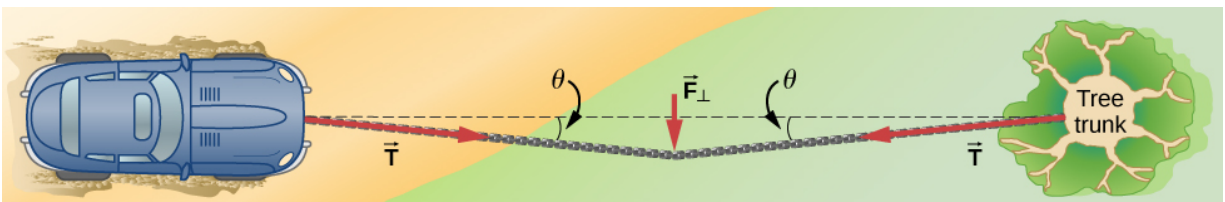
$$T = \frac{w}{2 \sin \theta}.$$

We can extend this expression to describe the tension T created when a perpendicular force (F_{\perp}) is exerted at the middle of a flexible connector:

Equation:

$$T = \frac{F_{\perp}}{2 \sin \theta}.$$

The angle between the horizontal and the bent connector is represented by θ . In this case, T becomes large as θ approaches zero. Even the relatively small weight of any flexible connector will cause it to sag, since an infinite tension would result if it were horizontal (i.e., $\theta = 0$ and $\sin \theta = 0$). For example, [\[link\]](#) shows a situation where we wish to pull a car out of the mud when no tow truck is available. Each time the car moves forward, the chain is tightened to keep it as straight as possible. The tension in the chain is given by $T = \frac{F_{\perp}}{2 \sin \theta}$, and since θ is small, T is large. This situation is analogous to the tightrope walker, except that the tensions shown here are those transmitted to the car and the tree rather than those acting at the point where F_{\perp} is applied.



We can create a large tension in the chain—and potentially a big mess—by pushing on it perpendicular to its length, as shown.

Note:

Exercise:

Problem:

Check Your Understanding One end of a 3.0-m rope is tied to a tree; the other end is tied to a car stuck in the mud. The motorist pulls sideways on the midpoint of the rope, displacing it a distance of 0.25 m. If he exerts a force of 200.0 N under these conditions, determine the force exerted on the car.

Solution:

$$6.0 \times 10^2 \text{ N}$$

In [Applications of Newton's Laws](#), we extend the discussion on tension in a cable to include cases in which the angles shown are not equal.

Friction

Friction is a resistive force opposing motion or its tendency. Imagine an object at rest on a horizontal surface. The net force acting on the object must be zero, leading to equality of the weight and the normal force, which act in opposite directions. If the surface is tilted, the normal force balances the component of the weight perpendicular to the surface. If the object does not slide downward, the component of the weight parallel to the inclined plane is balanced by friction. Friction is discussed in greater detail in the next chapter.

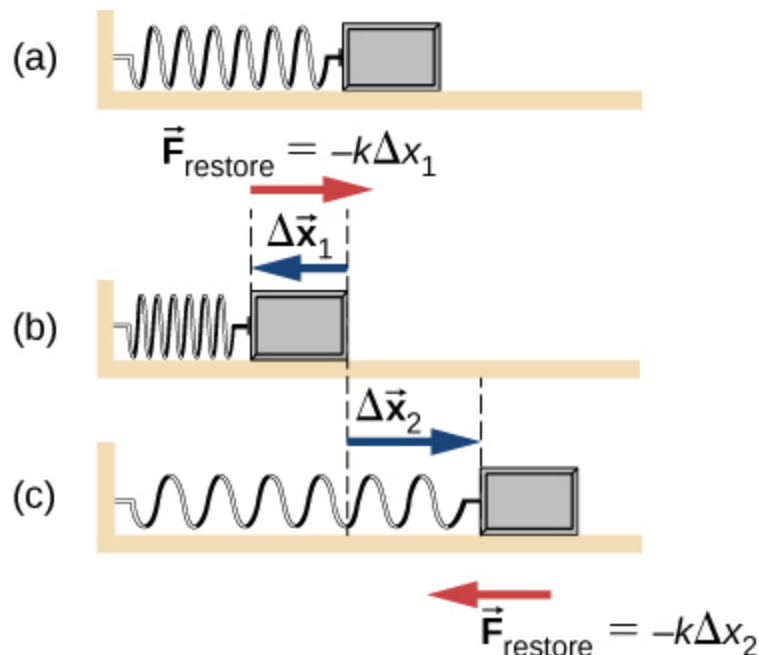
Spring force

A spring is a special medium with a specific atomic structure that has the ability to restore its shape, if deformed. To restore its shape, a spring exerts a restoring force that is proportional to and in the opposite direction in which it is stretched or compressed. This is the statement of a law known as Hooke's law, which has the mathematical form

Equation:

$$\vec{F} = -k\vec{x}.$$

The constant of proportionality k is a measure of the spring's stiffness. The line of action of this force is parallel to the spring axis, and the sense of the force is in the opposite direction of the displacement vector ([link](#)). The displacement must be measured from the relaxed position; $x = 0$ when the spring is relaxed.



A spring exerts its force proportional to a displacement, whether it is compressed

or stretched. (a) The spring is in a relaxed position and exerts no force on the block. (b) The spring is compressed by displacement $\Delta\vec{x}_1$ of the object and exerts restoring force $-k\Delta\vec{x}_1$. (c) The spring is stretched by displacement $\Delta\vec{x}_2$ of the object and exerts restoring force $-k\Delta\vec{x}_2$.

Real Forces and Inertial Frames

There is another distinction among forces: Some forces are real, whereas others are not. *Real forces* have some physical origin, such as a gravitational pull. In contrast, *fictitious forces* arise simply because an observer is in an accelerating or noninertial frame of reference, such as one that rotates (like a merry-go-round) or undergoes linear acceleration (like a car slowing down). For example, if a satellite is heading due north above Earth's Northern Hemisphere, then to an observer on Earth, it will appear to experience a force to the west that has no physical origin. Instead, Earth is rotating toward the east and moves east under the satellite. In Earth's frame, this looks like a westward force on the satellite, or it can be interpreted as a violation of Newton's first law (the law of inertia). We can identify a fictitious force by asking the question, "What is the reaction force?" If we cannot name the reaction force, then the force we are considering is fictitious. In the example of the satellite, the reaction force would have to be an eastward force on Earth. Recall that an inertial frame of reference is one in which all forces are real and, equivalently, one in which Newton's laws have the simple forms given in this chapter.

Earth's rotation is slow enough that Earth is nearly an inertial frame. You ordinarily must perform precise experiments to observe fictitious forces and the slight departures from Newton's laws, such as the effect just described. On a large scale, such as for the rotation of weather systems and ocean currents, the effects can be easily observed ([\[link\]](#)).



Hurricane Fran is shown heading toward the southeastern coast of the United States in September 1996. Notice the characteristic “eye” shape of the hurricane. This is a result of the Coriolis effect, which is the deflection of objects (in this case, air) when considered in a rotating frame of reference, like the spin of Earth. This hurricane shows a counter-clockwise rotation because it is a low pressure storm. (credit "runner": modification of work by "Greenwich Photography"/Flickr)

The crucial factor in determining whether a frame of reference is inertial is whether it accelerates or rotates relative to a known inertial frame. Unless stated otherwise, all phenomena discussed in this text are in inertial frames.

The forces discussed in this section are real forces, but they are not the only real forces. Lift and thrust, for example, are more specialized real forces. In

the long list of forces, are some more basic than others? Are some different manifestations of the same underlying force? The answer to both questions is yes, as you will see in the treatment of modern physics later in the text.

Note:

Explore forces and motion in this [interactive simulation](#) as you push household objects up and down a ramp. Lower and raise the ramp to see how the angle of inclination affects the parallel forces. Graphs show forces, energy, and work.

Note:

Stretch and compress springs in this [activity](#) to explore the relationships among force, spring constant, and displacement. Investigate what happens when two springs are connected in series and in parallel.

Summary

- When an object rests on a surface, the surface applies a force to the object that supports the weight of the object. This supporting force acts perpendicular to and away from the surface. It is called a normal force.
- When an object rests on a nonaccelerating horizontal surface, the magnitude of the normal force is equal to the weight of the object.
- When an object rests on an inclined plane that makes an angle θ with the horizontal surface, the weight of the object can be resolved into components that act perpendicular and parallel to the surface of the plane.
- The pulling force that acts along a stretched flexible connector, such as a rope or cable, is called tension. When a rope supports the weight of an object at rest, the tension in the rope is equal to the weight of the object. If the object is accelerating, tension is greater than weight, and if it is accelerating opposite to the motion, tension is less than weight.

- The force of friction is a force experienced by a moving object (or an object that has a tendency to move) parallel to the interface opposing the motion (or its tendency).
- The force developed in a spring obeys Hooke's law, according to which its magnitude is proportional to the displacement and has a sense in the opposite direction of the displacement.
- Real forces have a physical origin, whereas fictitious forces occur because the observer is in an accelerating or noninertial frame of reference.

Conceptual Questions

Exercise:

Problem:

A table is placed on a rug. Then a book is placed on the table. What does the floor exert a normal force on?

Exercise:

Problem:

A particle is moving to the right. (a) Can the force on it be acting to the left? If yes, what would happen? (b) Can that force be acting downward? If yes, why?

Solution:

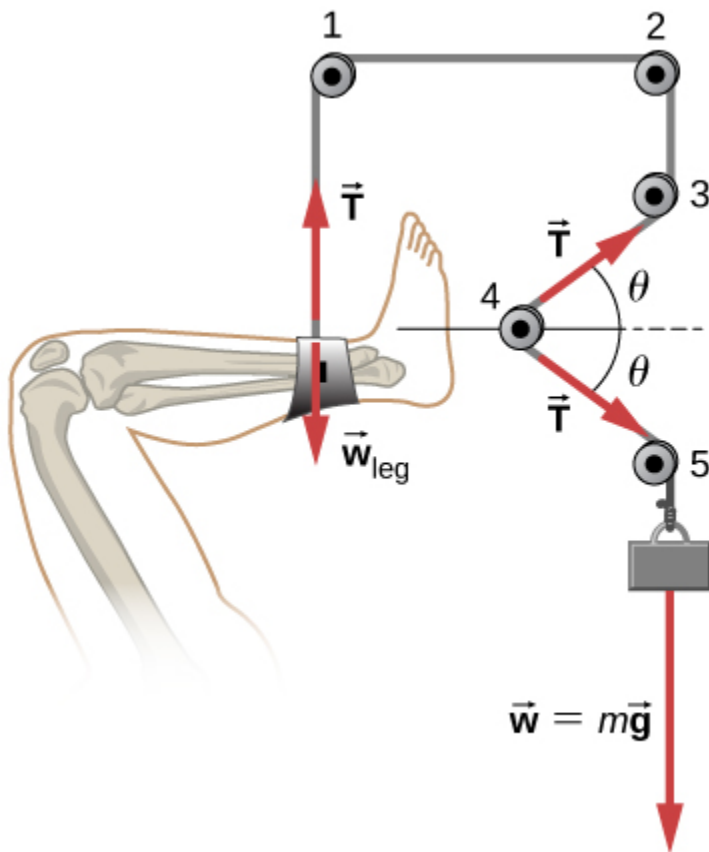
a. Yes, the force can be acting to the left; the particle would experience acceleration opposite to the motion and lose speed. B. Yes, the force can be acting downward because its weight acts downward even as it moves to the right.

Problems

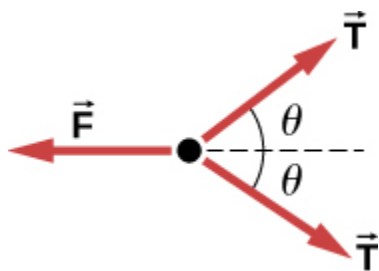
Exercise:

Problem:

A leg is suspended in a traction system, as shown below. (a) Which pulley in the figure is used to calculate the force exerted on the foot? (b) What is the tension in the rope? Here \vec{T} is the tension, \vec{w}_{leg} is the weight of the leg, and \vec{w} is the weight of the load that provides the tension.

**Solution:**

a. The free-body diagram of pulley 4:



b. $T = mg$, $F = 2T \cos \theta = 2mg \cos \theta$

Exercise:

Problem:

Suppose the shinbone in the preceding image was a femur in a traction setup for a broken bone, with pulleys and rope available. How might we be able to increase the force along the femur using the same weight?

Exercise:

Problem:

A team of nine members on a tall building tug on a string attached to a large boulder on an icy surface. The boulder has a mass of 200 kg and is tugged with a force of 2350 N. (a) What is magnitude of the acceleration? (b) What force would be required to produce a constant velocity?

Solution:

a. 1.95 m/s^2

b. 1960 N

Exercise:

Problem:

What force does a trampoline have to apply to Jennifer, a 45.0-kg gymnast, to accelerate her straight up at 7.50 m/s^2 ? The answer is independent of the velocity of the gymnast—she can be moving up or down or can be instantly stationary.

Exercise:**Problem:**

(a) Calculate the tension in a vertical strand of spider web if a spider of mass $2.00 \times 10^{-5} \text{ kg}$ hangs motionless on it. (b) Calculate the tension in a horizontal strand of spider web if the same spider sits motionless in the middle of it much like the tightrope walker in [\[link\]](#). The strand sags at an angle of 12° below the horizontal. Compare this with the tension in the vertical strand (find their ratio).

Solution:

a. $T = 1.96 \times 10^{-4} \text{ N};$

b. $T' = 4.71 \times 10^{-4} \text{ N}$

$$\frac{T'}{T} = 2.40 \text{ times the tension in the vertical strand}$$

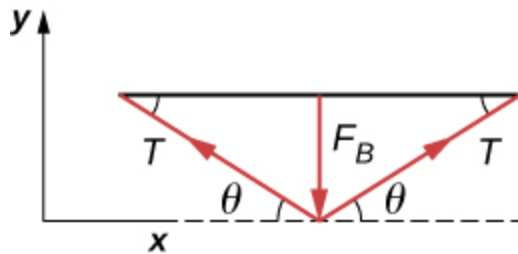
Exercise:**Problem:**

Suppose Kevin, a 60.0-kg gymnast, climbs a rope. (a) What is the tension in the rope if he climbs at a constant speed? (b) What is the tension in the rope if he accelerates upward at a rate of 1.50 m/s^2 ?

Exercise:

Problem:

Show that, as explained in the text, a force F_{\perp} exerted on a flexible medium at its center and perpendicular to its length (such as on the tightrope wire in [\[link\]](#)) gives rise to a tension of magnitude $T = F_{\perp}/2 \sin(\theta)$.

Solution:

$$F_{y \text{ net}} = F_{\perp} - 2T \sin \theta = 0$$

$$F_{\perp} = 2T \sin \theta$$

$$T = \frac{F_{\perp}}{2 \sin \theta}$$

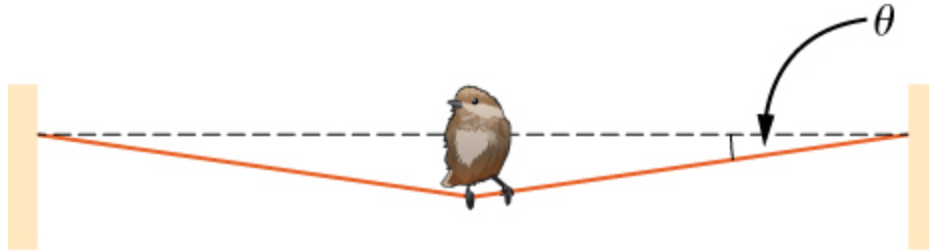
Exercise:**Problem:**

Consider [\[link\]](#). The driver attempts to get the car out of the mud by exerting a perpendicular force of 610.0 N, and the distance she pushes in the middle of the rope is 1.00 m while she stands 6.00 m away from the car on the left and 6.00 m away from the tree on the right. What is the tension T in the rope, and how do you find the answer?

Exercise:

Problem:

A bird has a mass of 26 g and perches in the middle of a stretched telephone line. (a) Show that the tension in the line can be calculated using the equation $T = \frac{mg}{2 \sin \theta}$. Determine the tension when (b) $\theta = 5^\circ$ and (c) $\theta = 0.5^\circ$. Assume that each half of the line is straight.

**Solution:**

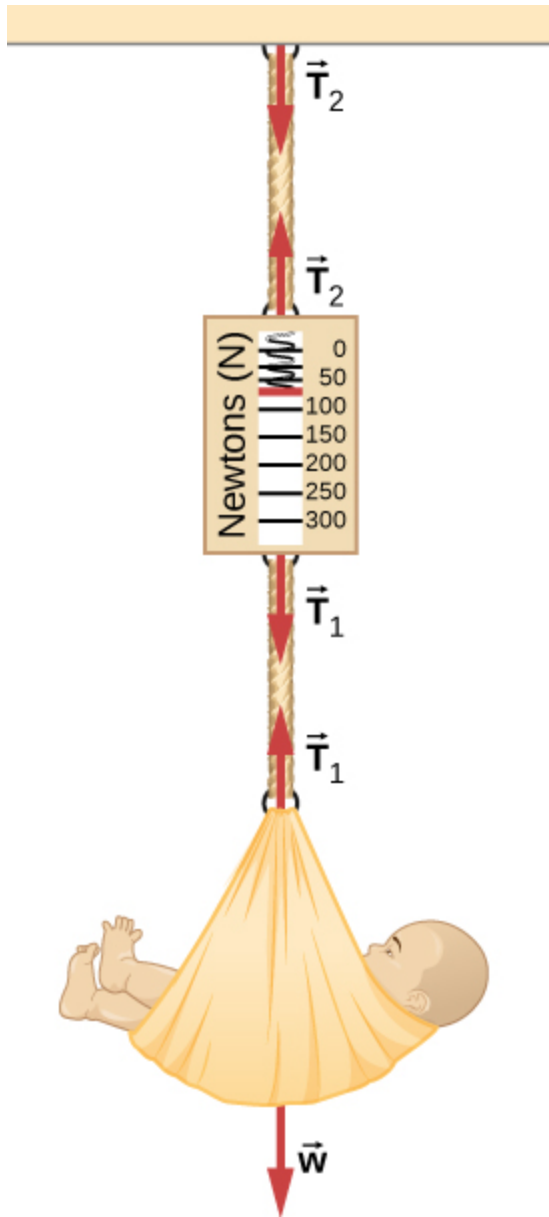
a. see [\[link\]](#); b. 1.5 N; c. 15 N

Exercise:**Problem:**

One end of a 30-m rope is tied to a tree; the other end is tied to a car stuck in the mud. The motorist pulls sideways on the midpoint of the rope, displacing it a distance of 2 m. If he exerts a force of 80 N under these conditions, determine the force exerted on the car.

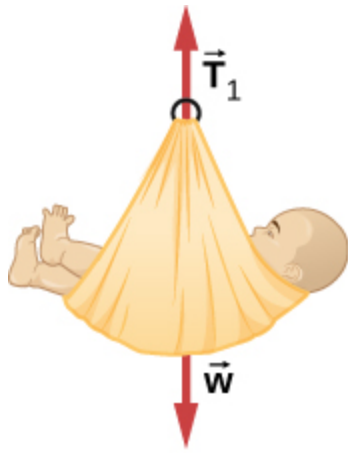
Exercise:**Problem:**

Consider the baby being weighed in the following figure. (a) What is the mass of the infant and basket if a scale reading of 55 N is observed? (b) What is tension T_1 in the cord attaching the baby to the scale? (c) What is tension T_2 in the cord attaching the scale to the ceiling, if the scale has a mass of 0.500 kg? (d) Sketch the situation, indicating the system of interest used to solve each part. The masses of the cords are negligible.

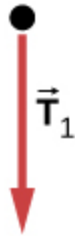


Solution:

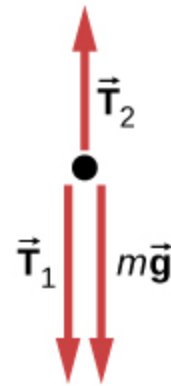
- a. 5.6 kg; b. 55 N; c. $T_2 = 60 \text{ N}$;
d.



(a)



(b)

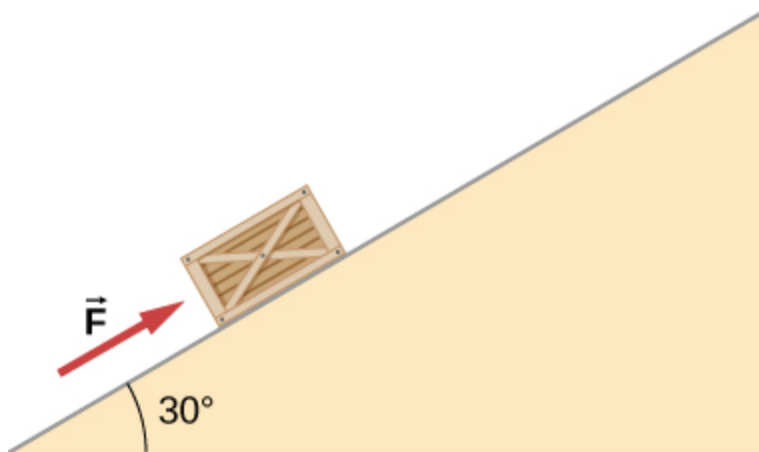


(c)

Exercise:

Problem:

What force must be applied to a 100.0-kg crate on a frictionless plane inclined at 30° to cause an acceleration of 2.0 m/s^2 up the plane?



Exercise:

Problem:

A 2.0-kg block is on a perfectly smooth ramp that makes an angle of 30° with the horizontal. (a) What is the block's acceleration down the ramp and the force of the ramp on the block? (b) What force applied upward along and parallel to the ramp would allow the block to move with constant velocity?

Solution:

a. 4.9 m/s^2 , 17 N; b. 9.8 N

Glossary

Hooke's law

in a spring, a restoring force proportional to and in the opposite direction of the imposed displacement

normal force

force supporting the weight of an object, or a load, that is perpendicular to the surface of contact between the load and its support; the surface applies this force to an object to support the weight of the object

tension

pulling force that acts along a stretched flexible connector, such as a rope or cable

Drawing Free-Body Diagrams

By the end of the section, you will be able to:

- Explain the rules for drawing a free-body diagram
- Construct free-body diagrams for different situations

The first step in describing and analyzing most phenomena in physics involves the careful drawing of a free-body diagram. Free-body diagrams have been used in examples throughout this chapter. Remember that a free-body diagram must only include the external forces acting on the body of interest. Once we have drawn an accurate free-body diagram, we can apply Newton's first law if the body is in equilibrium (balanced forces; that is, $F_{\text{net}} = 0$) or Newton's second law if the body is accelerating (unbalanced force; that is, $F_{\text{net}} \neq 0$).

In [Forces](#), we gave a brief problem-solving strategy to help you understand free-body diagrams. Here, we add some details to the strategy that will help you in constructing these diagrams.

Note:

Constructing Free-Body Diagrams

Observe the following rules when constructing a free-body diagram:

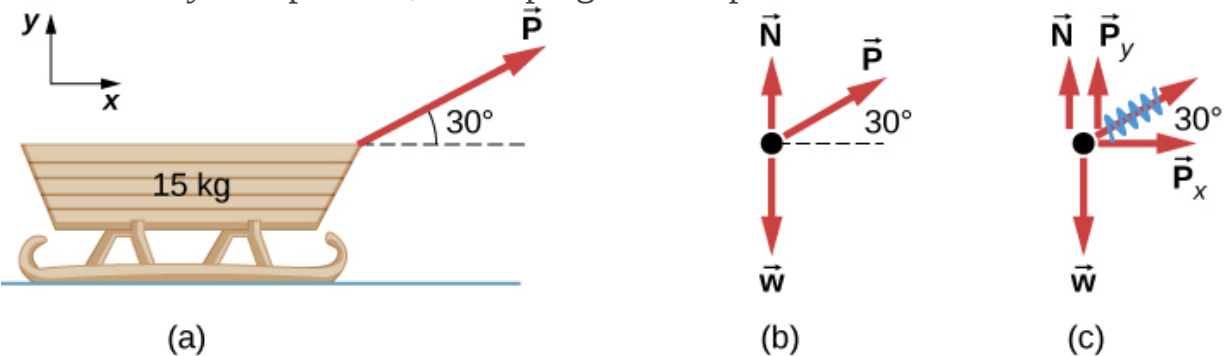
1. Draw the object under consideration; it does not have to be artistic. At first, you may want to draw a circle around the object of interest to be sure you focus on labeling the forces acting on the object. If you are treating the object as a particle (no size or shape and no rotation), represent the object as a point. We often place this point at the origin of an xy -coordinate system.
2. Include all forces that act on the object, representing these forces as vectors. Consider the types of forces described in [Common Forces](#)—normal force, friction, tension, and spring force—as well as weight and applied force. Do not include the net force on the object. With the exception of gravity, all of the forces we have discussed require direct contact with the object. However, forces that the object exerts on its

environment must not be included. We never include both forces of an action-reaction pair.

3. Convert the free-body diagram into a more detailed diagram showing the x - and y -components of a given force (this is often helpful when solving a problem using Newton's first or second law). In this case, place a squiggly line through the original vector to show that it is no longer in play—it has been replaced by its x - and y -components.
4. If there are two or more objects, or bodies, in the problem, draw a separate free-body diagram for each object.

Note: If there is acceleration, we do not directly include it in the free-body diagram; however, it may help to indicate acceleration outside the free-body diagram. You can label it in a different color to indicate that it is separate from the free-body diagram.

Let's apply the problem-solving strategy in drawing a free-body diagram for a sled. In [\[link\]](#)(a), a sled is pulled by force \vec{P} at an angle of 30° . In part (b), we show a free-body diagram for this situation, as described by steps 1 and 2 of the problem-solving strategy. In part (c), we show all forces in terms of their x - and y -components, in keeping with step 3.



(a) A moving sled is shown as (b) a free-body diagram and (c) a free-body diagram with force components.

Example:

Two Blocks on an Inclined Plane

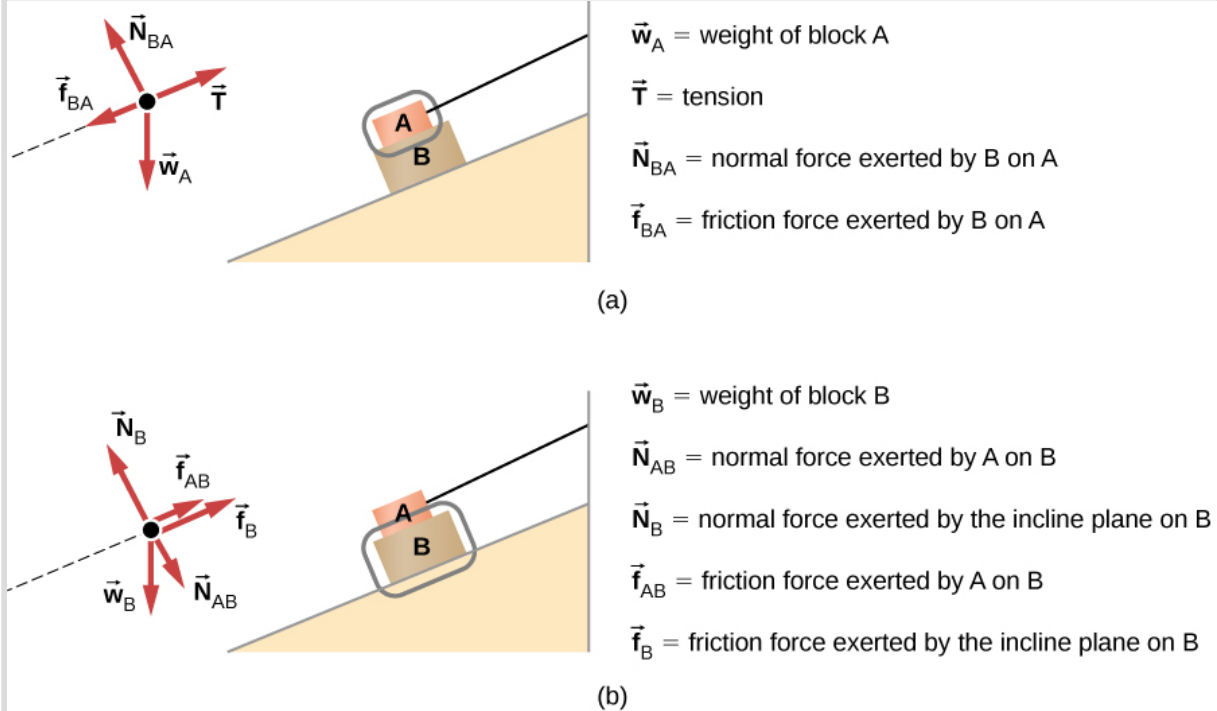
Construct the free-body diagram for object A and object B in [\[link\]](#).

Strategy

We follow the four steps listed in the problem-solving strategy.

Solution

We start by creating a diagram for the first object of interest. In [\[link\]](#)(a), object A is isolated (circled) and represented by a dot.



(a) The free-body diagram for isolated object A. (b) The free-body diagram for isolated object B. Comparing the two drawings, we see that friction acts in the opposite direction in the two figures. Because object A experiences a force that tends to pull it to the right, friction must act to the left. Because object B experiences a component of its weight that pulls it to the left, down the incline, the friction force must oppose it and act up the ramp. Friction always acts opposite the intended direction of motion.

We now include any force that acts on the body. Here, no applied force is present. The weight of the object acts as a force pointing vertically downward, and the presence of the cord indicates a force of tension pointing

away from the object. Object A has one interface and hence experiences a normal force, directed away from the interface. The source of this force is object B, and this normal force is labeled accordingly. Since object B has a tendency to slide down, object A has a tendency to slide up with respect to the interface, so the friction f_{BA} is directed downward parallel to the inclined plane.

As noted in step 4 of the problem-solving strategy, we then construct the free-body diagram in [\[link\]](#)(b) using the same approach. Object B experiences two normal forces and two friction forces due to the presence of two contact surfaces. The interface with the inclined plane exerts external forces of N_B and f_B , and the interface with object A exerts the normal force N_{AB} and friction f_{AB} ; N_{AB} is directed away from object B, and f_{AB} is opposing the tendency of the relative motion of object B with respect to object A.

Significance

The object under consideration in each part of this problem was circled in gray. When you are first learning how to draw free-body diagrams, you will find it helpful to circle the object before deciding what forces are acting on that particular object. This focuses your attention, preventing you from considering forces that are not acting on the body.

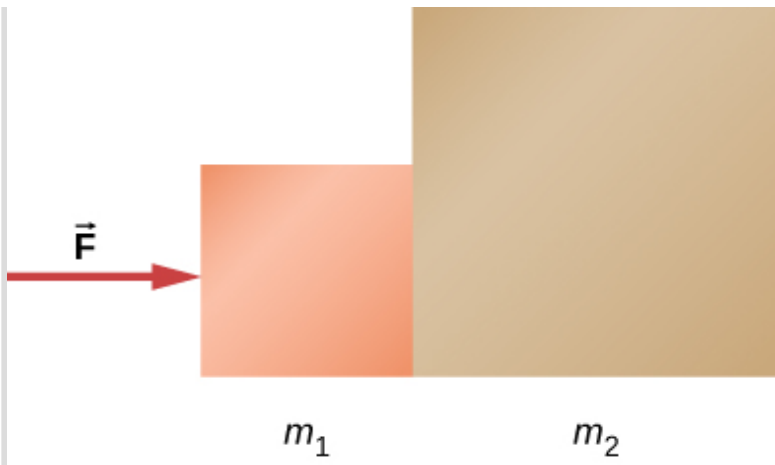
Example:

Two Blocks in Contact

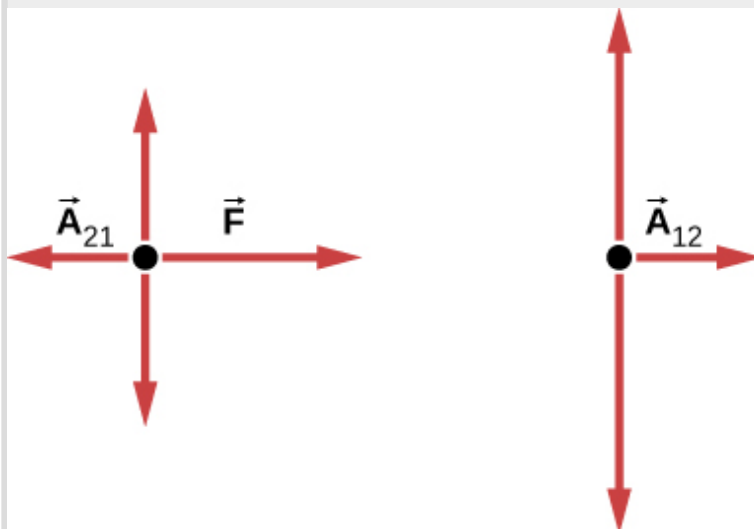
A force is applied to two blocks in contact, as shown.

Strategy

Draw a free-body diagram for each block. Be sure to consider Newton's third law at the interface where the two blocks touch.



Solution



Significance

\vec{A}_{21} is the action force of block 2 on block 1. \vec{A}_{12} is the reaction force of block 1 on block 2. We use these free-body diagrams in [Applications of Newton's Laws](#).

Example:

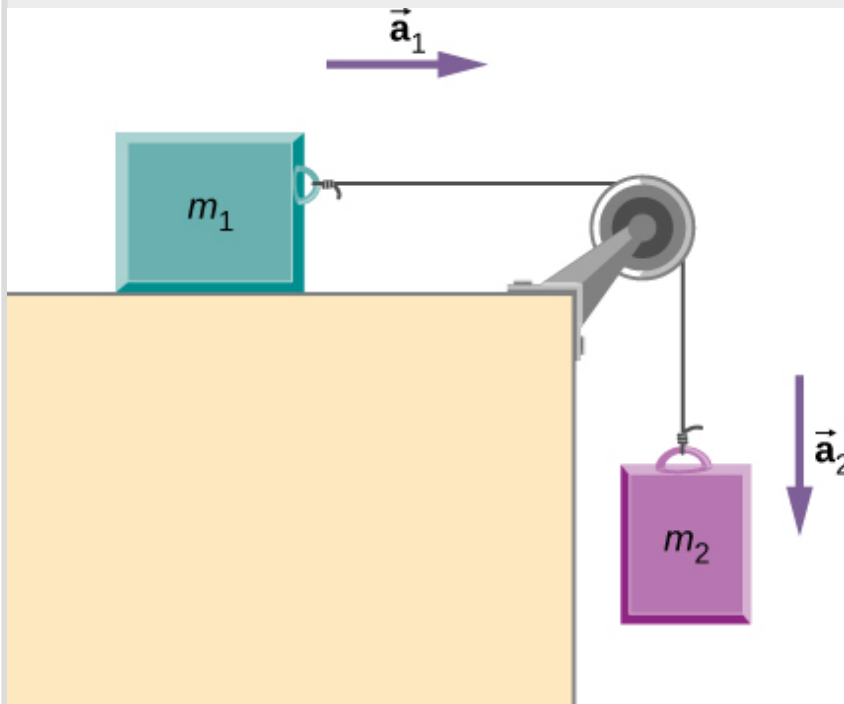
Block on the Table (Coupled Blocks)

A block rests on the table, as shown. A light rope is attached to it and runs over a pulley. The other end of the rope is attached to a second block. The

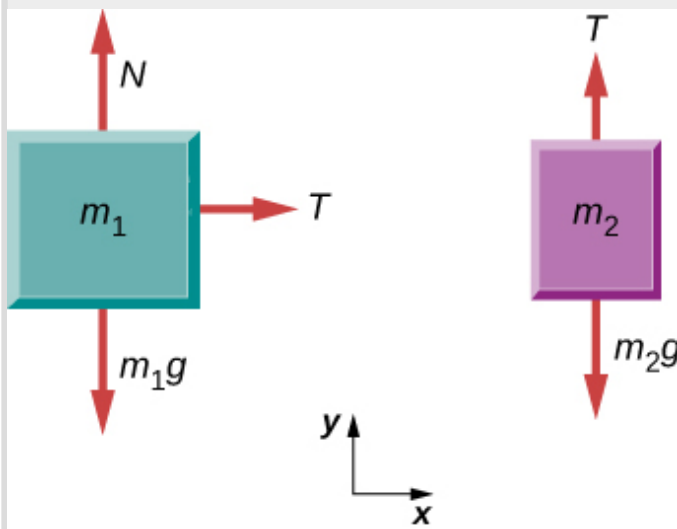
two blocks are said to be coupled. Block m_2 exerts a force due to its weight, which causes the system (two blocks and a string) to accelerate.

Strategy

We assume that the string has no mass so that we do not have to consider it as a separate object. Draw a free-body diagram for each block.



Solution



Significance

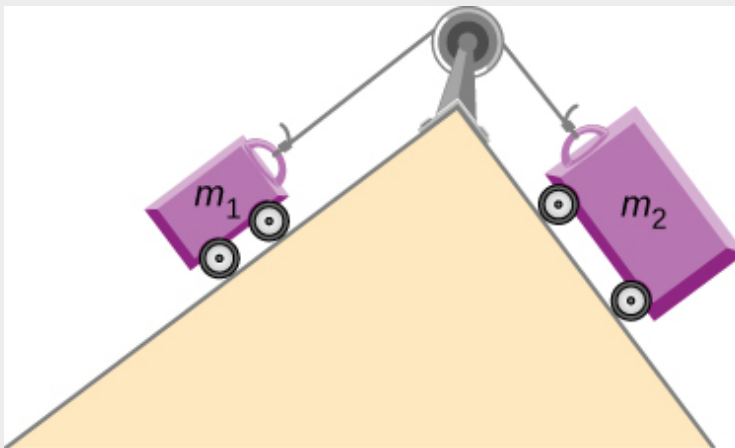
Each block accelerates (notice the labels shown for \vec{a}_1 and \vec{a}_2); however, assuming the string remains taut, the magnitudes of acceleration are equal. Thus, we have $|\vec{a}_1| = |\vec{a}_2|$. If we were to continue solving the problem, we could simply call the acceleration \vec{a} . Also, we use two free-body diagrams because we are usually finding tension T , which may require us to use a system of two equations in this type of problem. The tension is the same on both m_1 and m_2 .

Note:

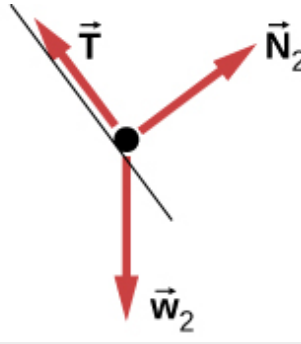
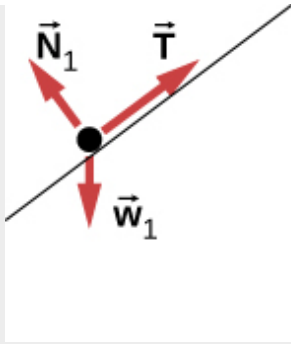
Exercise:

Problem:

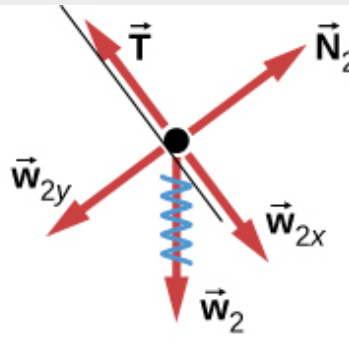
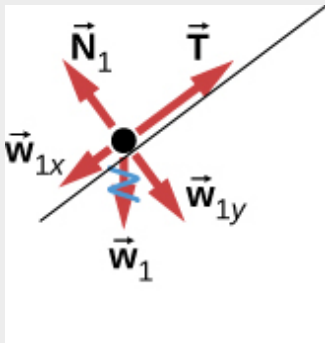
Check Your Understanding (a) Draw the free-body diagram for the situation shown. (b) Redraw it showing components; use x-axes parallel to the two ramps.



Solution:



;



Note:

View this [simulation](#) to predict, qualitatively, how an external force will affect the speed and direction of an object's motion. Explain the effects with the help of a free-body diagram. Use free-body diagrams to draw position, velocity, acceleration, and force graphs, and vice versa. Explain how the graphs relate to one another. Given a scenario or a graph, sketch all four graphs.

Summary

- To draw a free-body diagram, we draw the object of interest, draw all forces acting on that object, and resolve all force vectors into x- and y-components. We must draw a separate free-body diagram for each object in the problem.

- A free-body diagram is a useful means of describing and analyzing all the forces that act on a body to determine equilibrium according to Newton's first law or acceleration according to Newton's second law.

Key Equations

Net external force	$\vec{\mathbf{F}}_{\text{net}} = \sum \vec{\mathbf{F}} = \vec{\mathbf{F}}_1 + \vec{\mathbf{F}}_2 + \cdots$
Newton's first law	$\vec{\mathbf{v}} = \text{constant when } \vec{\mathbf{F}}_{\text{net}} = \vec{\mathbf{0}} \text{ N}$
Newton's second law, vector form	$\vec{\mathbf{F}}_{\text{net}} = \sum \vec{\mathbf{F}} = m\vec{\mathbf{a}}$
Newton's second law, scalar form	$F_{\text{net}} = ma$
Newton's second law, component form	$\sum \vec{\mathbf{F}}_x = m\vec{\mathbf{a}}_x, \sum \vec{\mathbf{F}}_y = m\vec{\mathbf{a}}_y, \text{ and } \sum \vec{\mathbf{F}}_z = m\vec{\mathbf{a}}_z.$
Newton's second law,	$\vec{\mathbf{F}}_{\text{net}} = \frac{d\vec{\mathbf{p}}}{dt}$

momentum form	
Definition of weight, vector form	$\vec{\mathbf{w}} = m\vec{\mathbf{g}}$
Definition of weight, scalar form	$w = mg$
Newton's third law	$\vec{\mathbf{F}}_{\text{AB}} = -\vec{\mathbf{F}}_{\text{BA}}$
Normal force on an object resting on a horizontal surface, vector form	$\vec{\mathbf{N}} = -m\vec{\mathbf{g}}$
Normal force on an object resting on a horizontal surface, scalar form	$N = mg$
Normal force on an object resting on	$N = mg\cos\theta$

an inclined plane, scalar form	
Tension in a cable supporting an object of mass m at rest, scalar form	$T = w = mg$

Conceptual Questions

Exercise:

Problem:

In completing the solution for a problem involving forces, what do we do after constructing the free-body diagram? That is, what do we apply?

Exercise:

Problem:

If a book is located on a table, how many forces should be shown in a free-body diagram of the book? Describe them.

Solution:

two forces of different types: weight acting downward and normal force acting upward

Exercise:

Problem:

If the book in the previous question is in free fall, how many forces should be shown in a free-body diagram of the book? Describe them.

Problems

Exercise:

Problem:

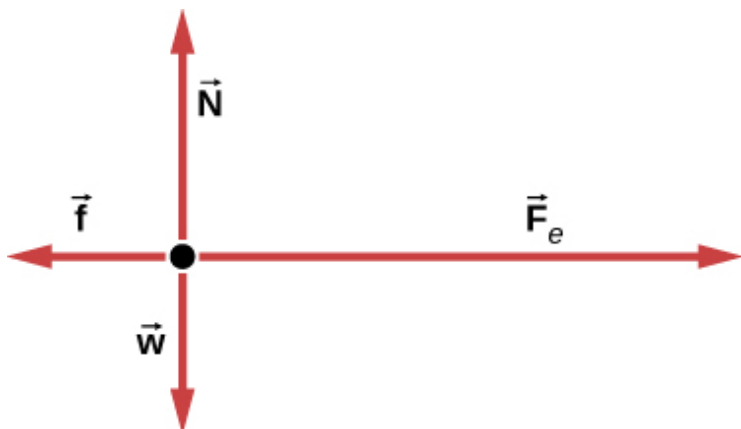
A ball of mass m hangs at rest, suspended by a string. (a) Sketch all forces. (b) Draw the free-body diagram for the ball.

Exercise:

Problem:

A car moves along a horizontal road. Draw a free-body diagram; be sure to include the friction of the road that opposes the forward motion of the car.

Solution:



Exercise:

Problem:

A runner pushes against the track, as shown. (a) Provide a free-body diagram showing all the forces on the runner. (*Hint:* Place all forces at the center of his body, and include his weight.) (b) Give a revised diagram showing the xy -component form.

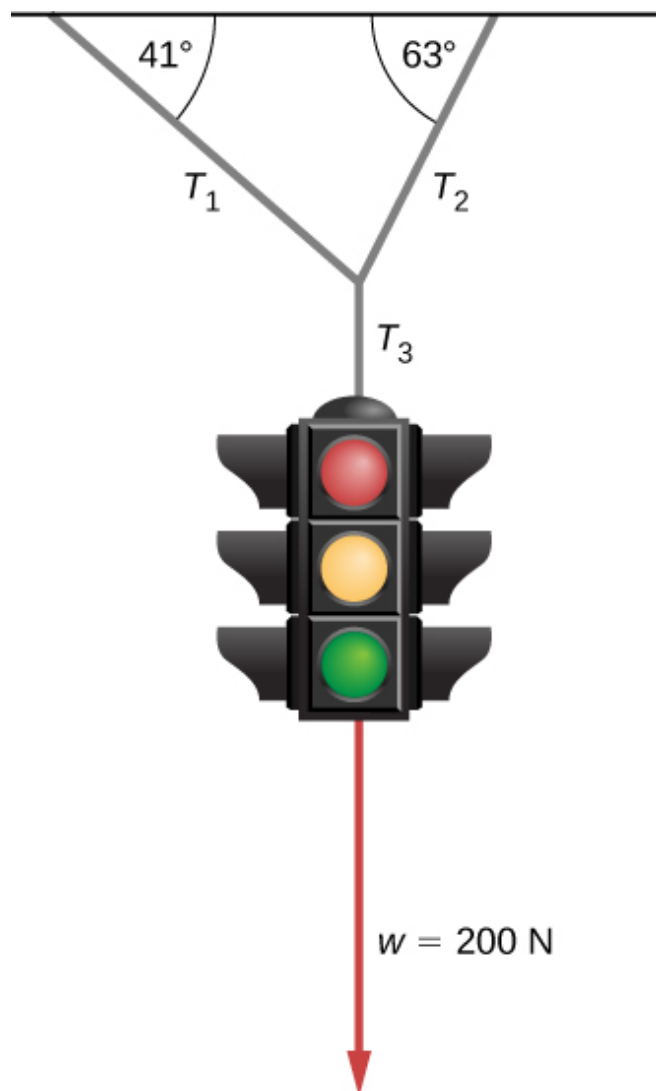


(credit: modification of work
by "Greenwich
Photography"/Flickr)

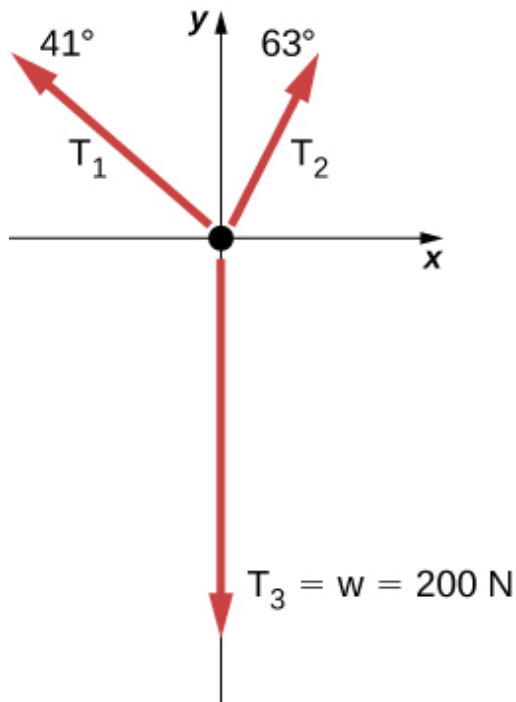
Exercise:

Problem:

The traffic light hangs from the cables as shown. Draw a free-body diagram on a coordinate plane for this situation.



Solution:



Additional Problems

Exercise:

Problem:

Two small forces, $\vec{F}_1 = -2.40\hat{i} - 6.10\hat{j}$ N and $\vec{F}_2 = 8.50\hat{i} - 9.70\hat{j}$ N, are exerted on a rogue asteroid by a pair of space tractors. (a) Find the net force. (b) What are the magnitude and direction of the net force? (c) If the mass of the asteroid is 125 kg, what acceleration does it experience (in vector form)? (d) What are the magnitude and direction of the acceleration?

Exercise:

Problem:

Two forces of 25 and 45 N act on an object. Their directions differ by 70° . The resulting acceleration has magnitude of 10.0 m/s^2 . What is the mass of the body?

Solution:

5.90 kg

Exercise:**Problem:**

A force of 1600 N acts parallel to a ramp to push a 300-kg piano into a moving van. The ramp is inclined at 20° . (a) What is the acceleration of the piano up the ramp? (b) What is the velocity of the piano when it reaches the top if the ramp is 4.0 m long and the piano starts from rest?

Exercise:**Problem:**

Draw a free-body diagram of a diver who has entered the water, moved downward, and is acted on by an upward force due to the water which balances the weight (that is, the diver is suspended).

Solution:**Exercise:**

Problem:

For a swimmer who has just jumped off a diving board, assume air resistance is negligible. The swimmer has a mass of 80.0 kg and jumps off a board 10.0 m above the water. Three seconds after entering the water, her downward motion is stopped. What average upward force did the water exert on her?

Exercise:**Problem:**

(a) Find an equation to determine the magnitude of the net force required to stop a car of mass m , given that the initial speed of the car is v_0 and the stopping distance is x . (b) Find the magnitude of the net force if the mass of the car is 1050 kg, the initial speed is 40.0 km/h, and the stopping distance is 25.0 m.

Solution:

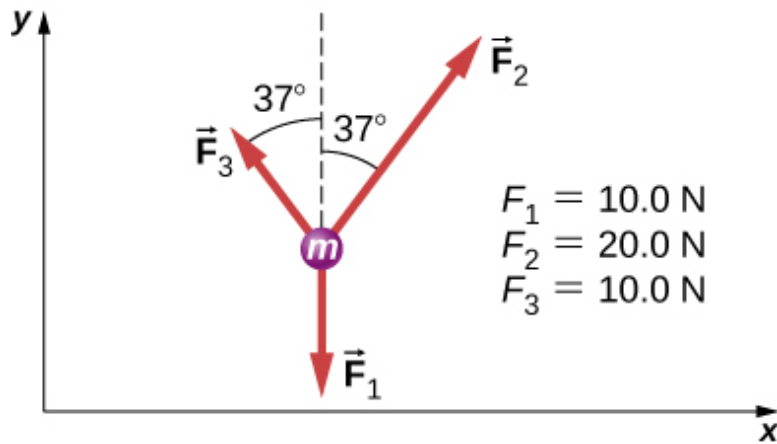
a. $F_{\text{net}} = \frac{m(v^2 - v_0^2)}{2x}$; b. 2590 N

Exercise:**Problem:**

A sailboat has a mass of 1.50×10^3 kg and is acted on by a force of 2.00×10^3 N toward the east, while the wind acts behind the sails with a force of 3.00×10^3 N in a direction 45° north of east. Find the magnitude and direction of the resulting acceleration.

Exercise:**Problem:**

Find the acceleration of the body of mass 10.0 kg shown below.



Solution:

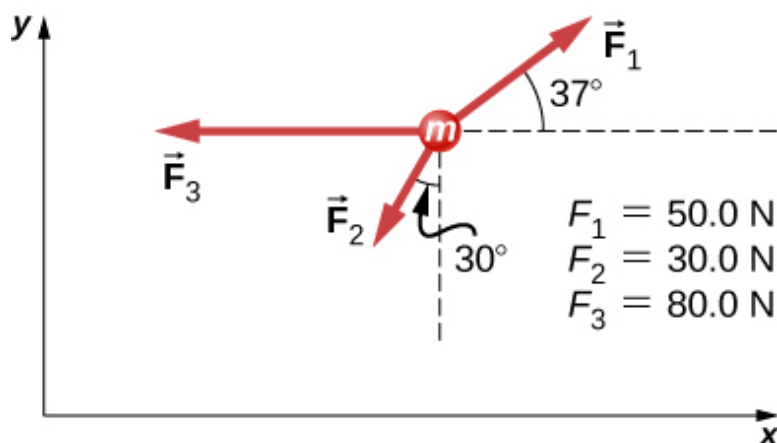
$$\vec{F}_{\text{net}} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 = (6.02\hat{i} + 14.0\hat{j})\text{N}$$

$$\vec{F}_{\text{net}} = m\vec{a} \Rightarrow \vec{a} = \frac{\vec{F}_{\text{net}}}{m} = \frac{6.02\hat{i} + 14.0\hat{j}\text{N}}{10.0\text{kg}} = (0.602\hat{i} + 1.40\hat{j})\text{m/s}^2$$

Exercise:

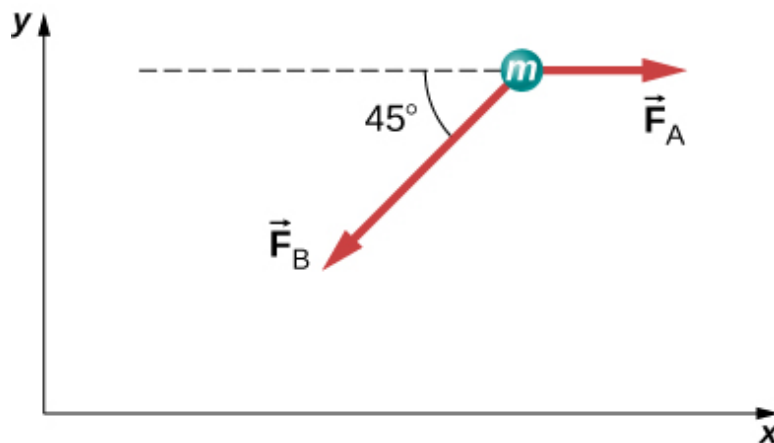
Problem:

A body of mass 2.0 kg is moving along the x -axis with a speed of 3.0 m/s at the instant represented below. (a) What is the acceleration of the body? (b) What is the body's velocity 10.0 s later? (c) What is its displacement after 10.0 s ?



Exercise:**Problem:**

Force \vec{F}_B has twice the magnitude of force \vec{F}_A . Find the direction in which the particle accelerates in this figure.

**Solution:**

$$\vec{F}_{\text{net}} = \vec{F}_A + \vec{F}_B$$

$$\vec{F}_{\text{net}} = A\hat{i} + (-1.41A\hat{i} - 1.41A\hat{j})$$

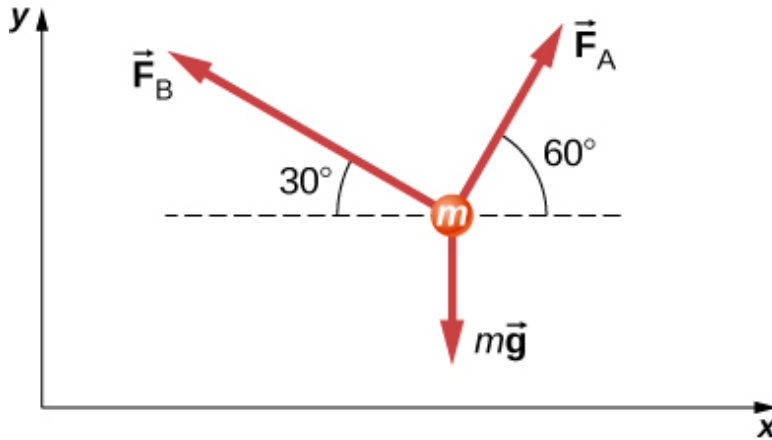
$$\vec{F}_{\text{net}} = A(-0.41\hat{i} - 1.41\hat{j})$$

$$\theta = 254^\circ$$

(We add 180° , because the angle is in quadrant IV.)

Exercise:**Problem:**

Shown below is a body of mass 1.0 kg under the influence of the forces \vec{F}_A , \vec{F}_B , and $m\vec{g}$. If the body accelerates to the left at 20 m/s^2 , what are \vec{F}_A and \vec{F}_B ?



Exercise:

Problem:

A force acts on a car of mass m so that the speed v of the car increases with position x as $v = kx^2$, where k is constant and all quantities are in SI units. Find the force acting on the car as a function of position.

Solution:

$F = 2mk^2x^2$; First, take the derivative of the velocity function to obtain $a = 2kxv = 2kx(kx^2) = 2k^2x^3$. Then apply Newton's second law $F = ma = 2mk^2x^2$.

Exercise:

Problem:

A 7.0-N force parallel to an incline is applied to a 1.0-kg crate. The ramp is tilted at 20° and is frictionless. (a) What is the acceleration of the crate? (b) If all other conditions are the same but the ramp has a friction force of 1.9 N, what is the acceleration?

Exercise:

Problem:

Two boxes, A and B, are at rest. Box A is on level ground, while box B rests on an inclined plane tilted at angle θ with the horizontal. (a) Write expressions for the normal force acting on each block. (b) Compare the two forces; that is, tell which one is larger or whether they are equal in magnitude. (c) If the angle of incline is 10° , which force is greater?

Solution:

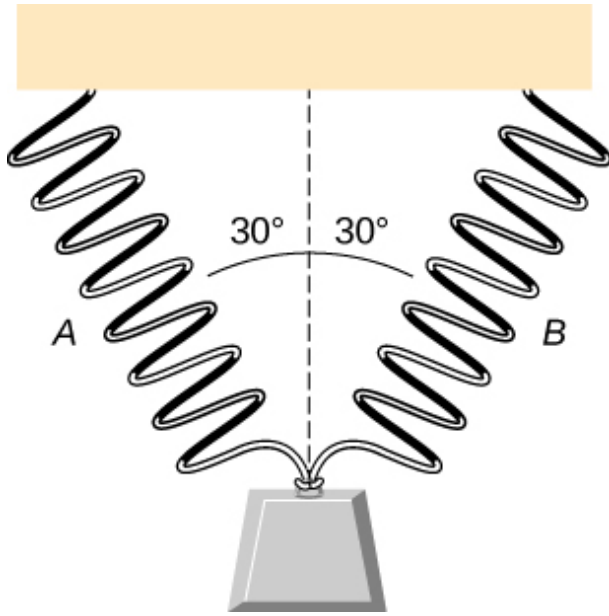
a. For box A, $N_A = mg$ and $N_B = mg \cos \theta$; b. $N_A > N_B$ because for $\theta < 90^\circ$, $\cos \theta < 1$; c. $N_A > N_B$ when $\theta = 10^\circ$

Exercise:**Problem:**

A mass of 250.0 g is suspended from a spring hanging vertically. The spring stretches 6.00 cm. How much will the spring stretch if the suspended mass is 530.0 g?

Exercise:**Problem:**

As shown below, two identical springs, each with the spring constant 20 N/m, support a 15.0-N weight. (a) What is the tension in spring A? (b) What is the amount of stretch of spring A from the rest position?



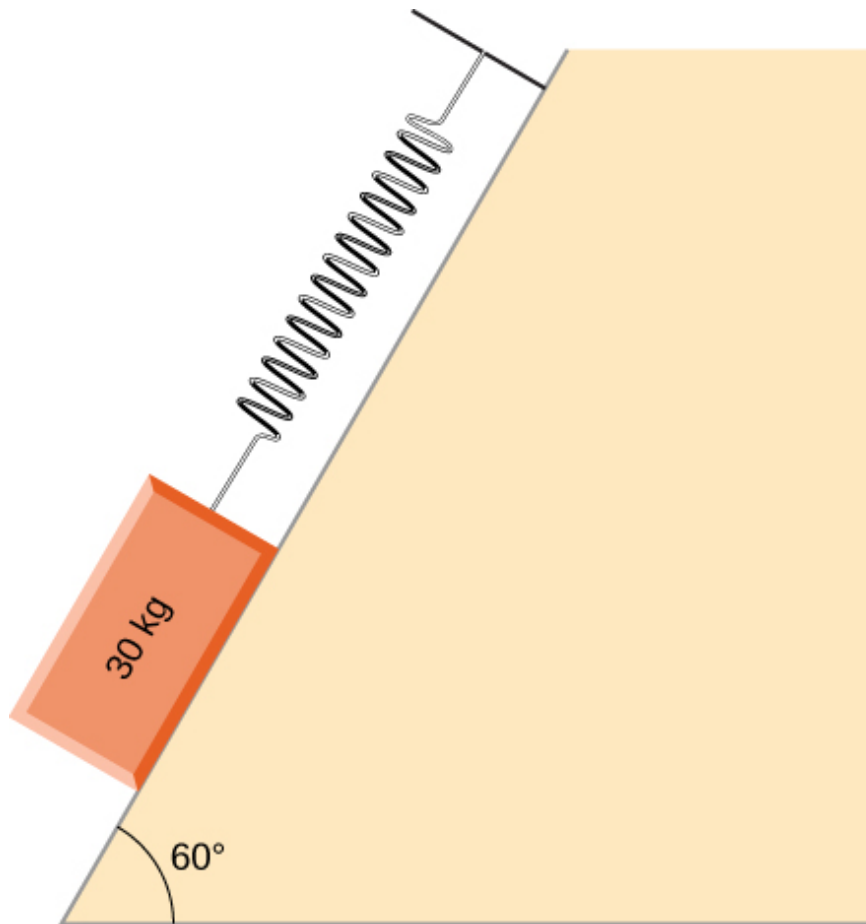
Solution:

a. 8.66 N; b. 0.433 m

Exercise:

Problem:

Shown below is a 30.0-kg block resting on a frictionless ramp inclined at 60° to the horizontal. The block is held by a spring that is stretched 5.0 cm. What is the force constant of the spring?



Exercise:

Problem:

In building a house, carpenters use nails from a large box. The box is suspended from a spring twice during the day to measure the usage of nails. At the beginning of the day, the spring stretches 50 cm. At the end of the day, the spring stretches 30 cm. What fraction or percentage of the nails have been used?

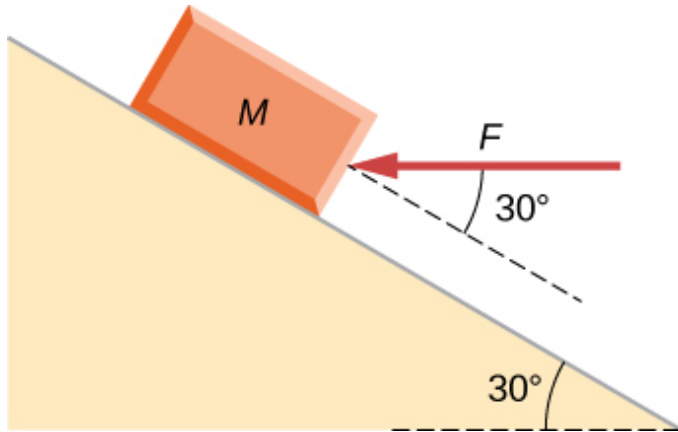
Solution:

0.40 or 40%

Exercise:

Problem:

A force is applied to a block to move it up a 30° incline. The incline is frictionless. If $F = 65.0\text{ N}$ and $M = 5.00\text{ kg}$, what is the magnitude of the acceleration of the block?

**Exercise:****Problem:**

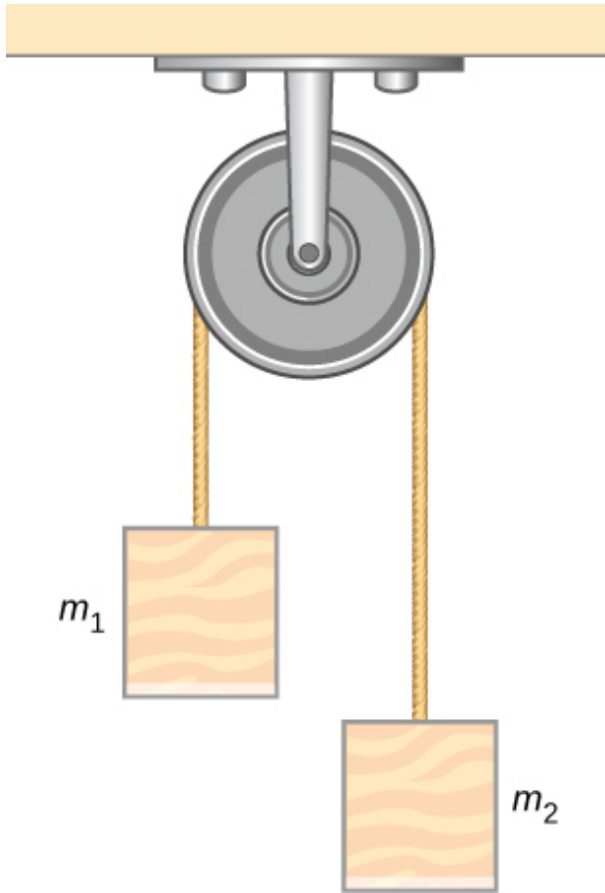
Two forces are applied to a 5.0-kg object, and it accelerates at a rate of 2.0 m/s^2 in the positive y -direction. If one of the forces acts in the positive x -direction with magnitude 12.0 N , find the magnitude of the other force.

Solution:

16 N

Exercise:**Problem:**

The block on the right shown below has more mass than the block on the left ($m_2 > m_1$). Draw free-body diagrams for each block.

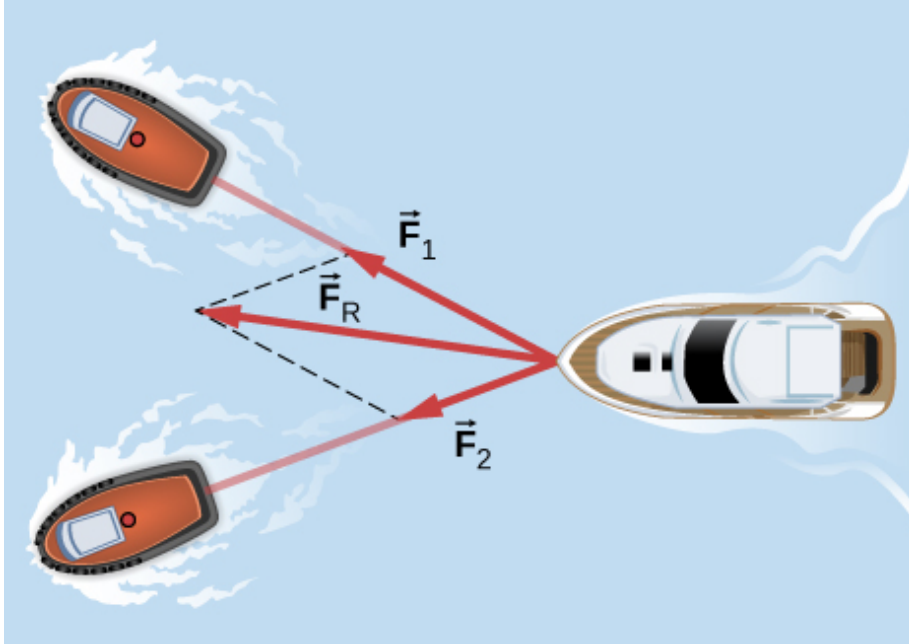


Challenge Problems

Exercise:

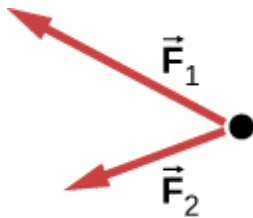
Problem:

If two tugboats pull on a disabled vessel, as shown here in an overhead view, the disabled vessel will be pulled along the direction indicated by the result of the exerted forces. (a) Draw a free-body diagram for the vessel. Assume no friction or drag forces affect the vessel. (b) Did you include all forces in the overhead view in your free-body diagram? Why or why not?



Solution:

a.



; b. No; \vec{F}_R is not shown, because it would replace \vec{F}_1 and \vec{F}_2 . (If we want to show it, we could draw it and then place squiggly lines on \vec{F}_1 and \vec{F}_2 to show that they are no longer considered.)

Exercise:

Problem:

A 10.0-kg object is initially moving east at 15.0 m/s. Then a force acts on it for 2.00 s, after which it moves northwest, also at 15.0 m/s. What are the magnitude and direction of the average force that acted on the object over the 2.00-s interval?

Exercise:**Problem:**

On June 25, 1983, shot-putter Udo Beyer of East Germany threw the 7.26-kg shot 22.22 m, which at that time was a world record. (a) If the shot was released at a height of 2.20 m with a projection angle of 45.0° , what was its initial velocity? (b) If while in Beyer's hand the shot was accelerated uniformly over a distance of 1.20 m, what was the net force on it?

Solution:

a. 14.1 m/s; b. 601 N

Exercise:**Problem:**

A body of mass m moves in a horizontal direction such that at time t its position is given by $x(t) = at^4 + bt^3 + ct$, where a , b , and c are constants. (a) What is the acceleration of the body? (b) What is the time-dependent force acting on the body?

Exercise:**Problem:**

A body of mass m has initial velocity v_0 in the positive x -direction. It is acted on by a constant force F for time t until the velocity becomes zero; the force continues to act on the body until its velocity becomes $-v_0$ in the same amount of time. Write an expression for the total distance the body travels in terms of the variables indicated.

Solution:

$$\frac{F}{m} t^2$$

Exercise:

Problem:

The velocities of a 3.0-kg object at $t = 6.0$ s and $t = 8.0$ s are $(3.0\hat{\mathbf{i}} - 6.0\hat{\mathbf{j}} + 4.0\hat{\mathbf{k}})$ m/s and $(-2.0\hat{\mathbf{i}} + 4.0\hat{\mathbf{k}})$ m/s, respectively. If the object is moving at constant acceleration, what is the force acting on it?

Exercise:**Problem:**

A 120-kg astronaut is riding in a rocket sled that is sliding along an inclined plane. The sled has a horizontal component of acceleration of 5.0 m/s^2 and a downward component of 3.8 m/s^2 . Calculate the magnitude of the force on the rider by the sled. (*Hint: Remember that gravitational acceleration must be considered.*)

Solution:

936 N

Exercise:**Problem:**

Two forces are acting on a 5.0-kg object that moves with acceleration 2.0 m/s^2 in the positive y -direction. If one of the forces acts in the positive x -direction and has magnitude of 12 N, what is the magnitude of the other force?

Exercise:**Problem:**

Suppose that you are viewing a soccer game from a helicopter above the playing field. Two soccer players simultaneously kick a stationary soccer ball on the flat field; the soccer ball has mass 0.420 kg. The first player kicks with force 162 N at 9.0° north of west. At the same instant, the second player kicks with force 215 N at 15° east of south. Find the acceleration of the ball in $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ form.

Solution:

$$\vec{a} = -248\hat{i} - 433\hat{j}\text{m/s}^2$$

Exercise:**Problem:**

A 10.0-kg mass hangs from a spring that has the spring constant 535 N/m. Find the position of the end of the spring away from its rest position. (Use $g = 9.80 \text{ m/s}^2$.)

Exercise:**Problem:**

A 0.0502-kg pair of fuzzy dice is attached to the rearview mirror of a car by a short string. The car accelerates at constant rate, and the dice hang at an angle of 3.20° from the vertical because of the car's acceleration. What is the magnitude of the acceleration of the car?

Solution:

$$0.548 \text{ m/s}^2$$

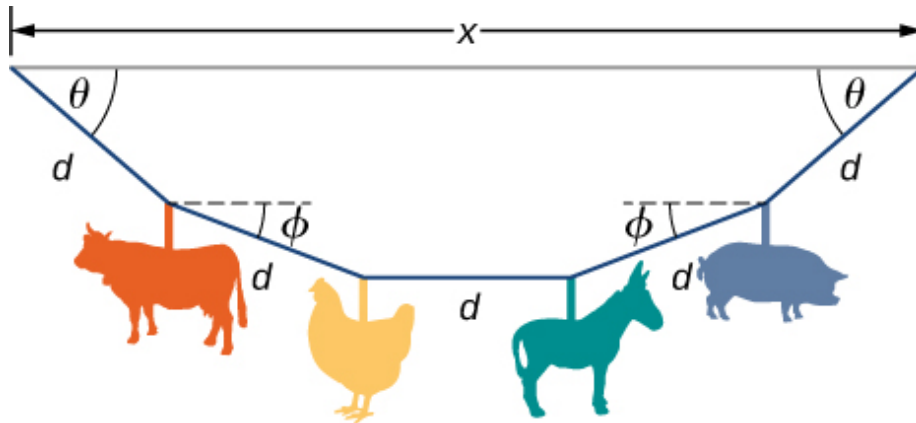
Exercise:**Problem:**

At a circus, a donkey pulls on a sled carrying a small clown with a force given by $2.48\hat{i} + 4.33\hat{j}$ N. A horse pulls on the same sled, aiding the hapless donkey, with a force of $6.56\hat{i} + 5.33\hat{j}$ N. The mass of the sled is 575 kg. Using \hat{i} and \hat{j} form for the answer to each problem, find (a) the net force on the sled when the two animals act together, (b) the acceleration of the sled, and (c) the velocity after 6.50 s.

Exercise:

Problem:

Hanging from the ceiling over a baby bed, well out of baby's reach, is a string with plastic shapes, as shown here. The string is taut (there is no slack), as shown by the straight segments. Each plastic shape has the same mass m , and they are equally spaced by a distance d , as shown. The angles labeled θ describe the angle formed by the end of the string and the ceiling at each end. The center length of string is horizontal. The remaining two segments each form an angle with the horizontal, labeled ϕ . Let T_1 be the tension in the leftmost section of the string, T_2 be the tension in the section adjacent to it, and T_3 be the tension in the horizontal segment. (a) Find an equation for the tension in each section of the string in terms of the variables m , g , and θ . (b) Find the angle ϕ in terms of the angle θ . (c) If $\theta = 5.10^\circ$, what is the value of ϕ ? (d) Find the distance x between the endpoints in terms of d and θ .

**Solution:**

a. $T_1 = \frac{2mg}{\sin \theta}$, $T_2 = \frac{mg}{\sin(\arctan(\frac{1}{2}\tan \theta))}$, $T_3 = \frac{2mg}{\tan \theta}$; b.

$\phi = \arctan(\frac{1}{2}\tan \theta)$; c. 2.56° ; (d)

$x = d(2 \cos \theta + 2 \cos(\arctan(\frac{1}{2}\tan \theta)) + 1)$

Exercise:

Problem:

A bullet shot from a rifle has mass of 10.0 g and travels to the right at 350 m/s. It strikes a target, a large bag of sand, penetrating it a distance of 34.0 cm. Find the magnitude and direction of the retarding force that slows and stops the bullet.

Exercise:**Problem:**

An object is acted on by three simultaneous forces:

$$\vec{F}_1 = (-3.00\hat{i} + 2.00\hat{j}) \text{ N}, \vec{F}_2 = (6.00\hat{i} - 4.00\hat{j}) \text{ N}, \text{ and}$$

$$\vec{F}_3 = (2.00\hat{i} + 5.00\hat{j}) \text{ N. The object experiences acceleration of}$$

4.23 m/s^2 . (a) Find the acceleration vector in terms of m . (b) Find the mass of the object. (c) If the object begins from rest, find its speed after 5.00 s. (d) Find the components of the velocity of the object after 5.00 s.

Solution:

$$\text{a. } \vec{a} = \left(\frac{5.00}{m}\hat{i} + \frac{3.00}{m}\hat{j} \right) \text{ m/s}^2; \text{ b. } 1.38 \text{ kg; c. } 21.2 \text{ m/s; d.}$$

$$\vec{v} = (18.1\hat{i} + 10.9\hat{j}) \text{ m/s}^2$$

Exercise:**Problem:**

In a particle accelerator, a proton has mass $1.67 \times 10^{-27} \text{ kg}$ and an initial speed of $2.00 \times 10^5 \text{ m/s}$. It moves in a straight line, and its speed increases to $9.00 \times 10^5 \text{ m/s}$ in a distance of 10.0 cm. Assume that the acceleration is constant. Find the magnitude of the force exerted on the proton.

Exercise:

Problem:

A drone is being directed across a frictionless ice-covered lake. The mass of the drone is 1.50 kg, and its velocity is $3.00\hat{i}\text{ m/s}$. After 10.0 s, the velocity is $9.00\hat{i} + 4.00\hat{j}\text{ m/s}$. If a constant force in the horizontal direction is causing this change in motion, find (a) the components of the force and (b) the magnitude of the force.

Solution:

a. $0.900\hat{i} + 0.600\hat{j}\text{ N}$; b. 1.08 N

Introduction

class="introduction"

Stock cars
racing in the
Grand
National
Divisional
race at Iowa
Speedway in
May, 2015.

Cars often
reach speeds
of 200 mph
(320 km/h).

(credit:
modification
of work by
Erik
Schneider/U.S
. Navy)



Car racing has grown in popularity in recent years. As each car moves in a curved path around the turn, its wheels also spin rapidly. The wheels complete many revolutions while the car makes only part of one (a circular arc). How can we describe the velocities, accelerations, and forces involved? What force keeps a racecar from spinning out, hitting the wall bordering the track? What provides this force? Why is the track banked? We answer all of these questions in this chapter as we expand our consideration of Newton's laws of motion.

Solving Problems with Newton's Laws

By the end of the section, you will be able to:

- Apply problem-solving techniques to solve for quantities in more complex systems of forces
- Use concepts from kinematics to solve problems using Newton's laws of motion
- Solve more complex equilibrium problems
- Solve more complex acceleration problems
- Apply calculus to more advanced dynamics problems

Success in problem solving is necessary to understand and apply physical principles. We developed a pattern of analyzing and setting up the solutions to problems involving Newton's laws in [Newton's Laws of Motion](#); in this chapter, we continue to discuss these strategies and apply a step-by-step process.

Problem-Solving Strategies

We follow here the basics of problem solving presented earlier in this text, but we emphasize specific strategies that are useful in applying Newton's laws of motion. Once you identify the physical principles involved in the problem and determine that they include Newton's laws of motion, you can apply these steps to find a solution. These techniques also reinforce concepts that are useful in many other areas of physics. Many problem-solving strategies are stated outright in the worked examples, so the following techniques should reinforce skills you have already begun to develop.

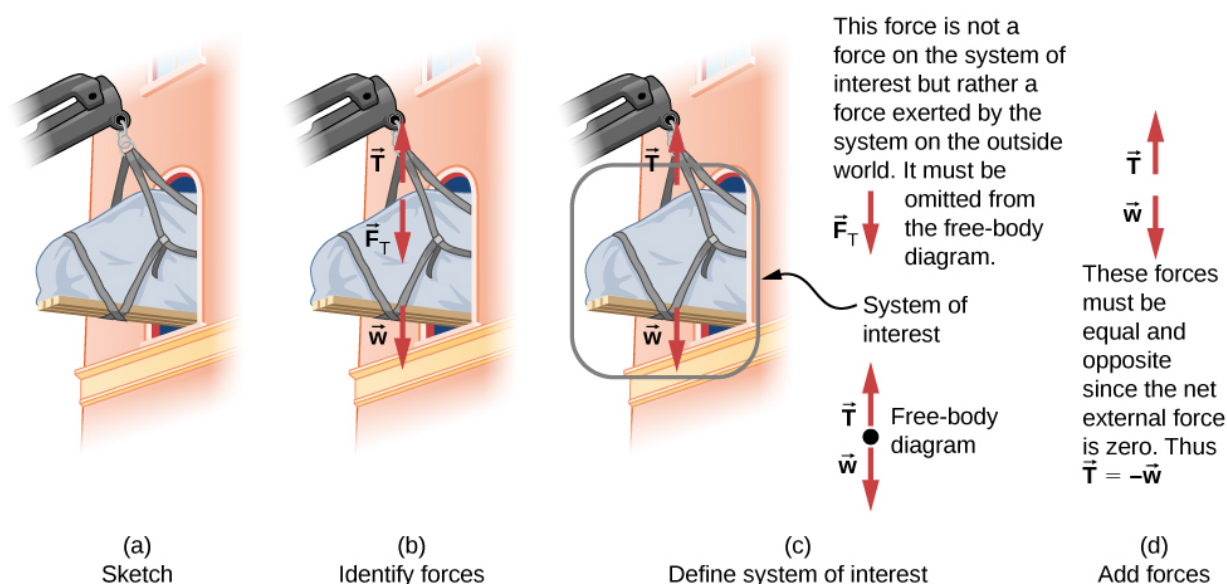
Note:

Applying Newton's Laws of Motion

1. Identify the physical principles involved by listing the givens and the quantities to be calculated.
2. Sketch the situation, using arrows to represent all forces.
3. Determine the system of interest. The result is a free-body diagram that is essential to solving the problem.
4. Apply Newton's second law to solve the problem. If necessary, apply appropriate kinematic equations from the chapter on motion along a straight line.
5. Check the solution to see whether it is reasonable.

Let's apply this problem-solving strategy to the challenge of lifting a grand piano into a second-story apartment. Once we have determined that Newton's laws of motion are involved (if the problem involves forces), it is particularly important to draw a careful sketch of the situation. Such a sketch is shown in [\[link\]](#)(a). Then, as in [\[link\]](#)(b), we can

represent all forces with arrows. Whenever sufficient information exists, it is best to label these arrows carefully and make the length and direction of each correspond to the represented force.



(a) A grand piano is being lifted to a second-story apartment. (b) Arrows are used to represent all forces: \vec{T} is the tension in the rope above the piano, \vec{F}_T is the force that the piano exerts on the rope, and \vec{w} is the weight of the piano. All other forces, such as the nudge of a breeze, are assumed to be negligible. (c) Suppose we are given the piano's mass and asked to find the tension in the rope. We then define the system of interest as shown and draw a free-body diagram. Now \vec{F}_T is no longer shown, because it is not a force acting on the system of interest; rather, \vec{F}_T acts on the outside world. (d) Showing only the arrows, the head-to-tail method of addition is used. It is apparent that if the piano is stationary, $\vec{T} = -\vec{w}$.

As with most problems, we next need to identify what needs to be determined and what is known or can be inferred from the problem as stated, that is, make a list of knowns and unknowns. It is particularly crucial to identify the system of interest, since Newton's second law involves only external forces. We can then determine which forces are external and which are internal, a necessary step to employ Newton's second law. (See [\[link\]](#)(c).) Newton's third law may be used to identify whether forces are exerted between components of a system (internal) or between the system and something outside (external). As illustrated in [Newton's Laws of Motion](#), the system of interest depends on the question we need to answer. Only forces are shown in free-body diagrams, not acceleration or velocity. We have drawn several free-body diagrams in previous worked examples. [\[link\]](#)(c) shows a free-

body diagram for the system of interest. Note that no internal forces are shown in a free-body diagram.

Once a free-body diagram is drawn, we apply Newton's second law. This is done in [\[link\]](#) (d) for a particular situation. In general, once external forces are clearly identified in free-body diagrams, it should be a straightforward task to put them into equation form and solve for the unknown, as done in all previous examples. If the problem is one-dimensional—that is, if all forces are parallel—then the forces can be handled algebraically. If the problem is two-dimensional, then it must be broken down into a pair of one-dimensional problems. We do this by projecting the force vectors onto a set of axes chosen for convenience. As seen in previous examples, the choice of axes can simplify the problem. For example, when an incline is involved, a set of axes with one axis parallel to the incline and one perpendicular to it is most convenient. It is almost always convenient to make one axis parallel to the direction of motion, if this is known. Generally, just write Newton's second law in components along the different directions. Then, you have the following equations:

Equation:

$$\sum F_x = ma_x, \quad \sum F_y = ma_y.$$

(If, for example, the system is accelerating horizontally, then you can then set $a_y = 0$.) We need this information to determine unknown forces acting on a system.

As always, we must check the solution. In some cases, it is easy to tell whether the solution is reasonable. For example, it is reasonable to find that friction causes an object to slide down an incline more slowly than when no friction exists. In practice, intuition develops gradually through problem solving; with experience, it becomes progressively easier to judge whether an answer is reasonable. Another way to check a solution is to check the units. If we are solving for force and end up with units of millimeters per second, then we have made a mistake.

There are many interesting applications of Newton's laws of motion, a few more of which are presented in this section. These serve also to illustrate some further subtleties of physics and to help build problem-solving skills. We look first at problems involving particle equilibrium, which make use of Newton's first law, and then consider particle acceleration, which involves Newton's second law.

Particle Equilibrium

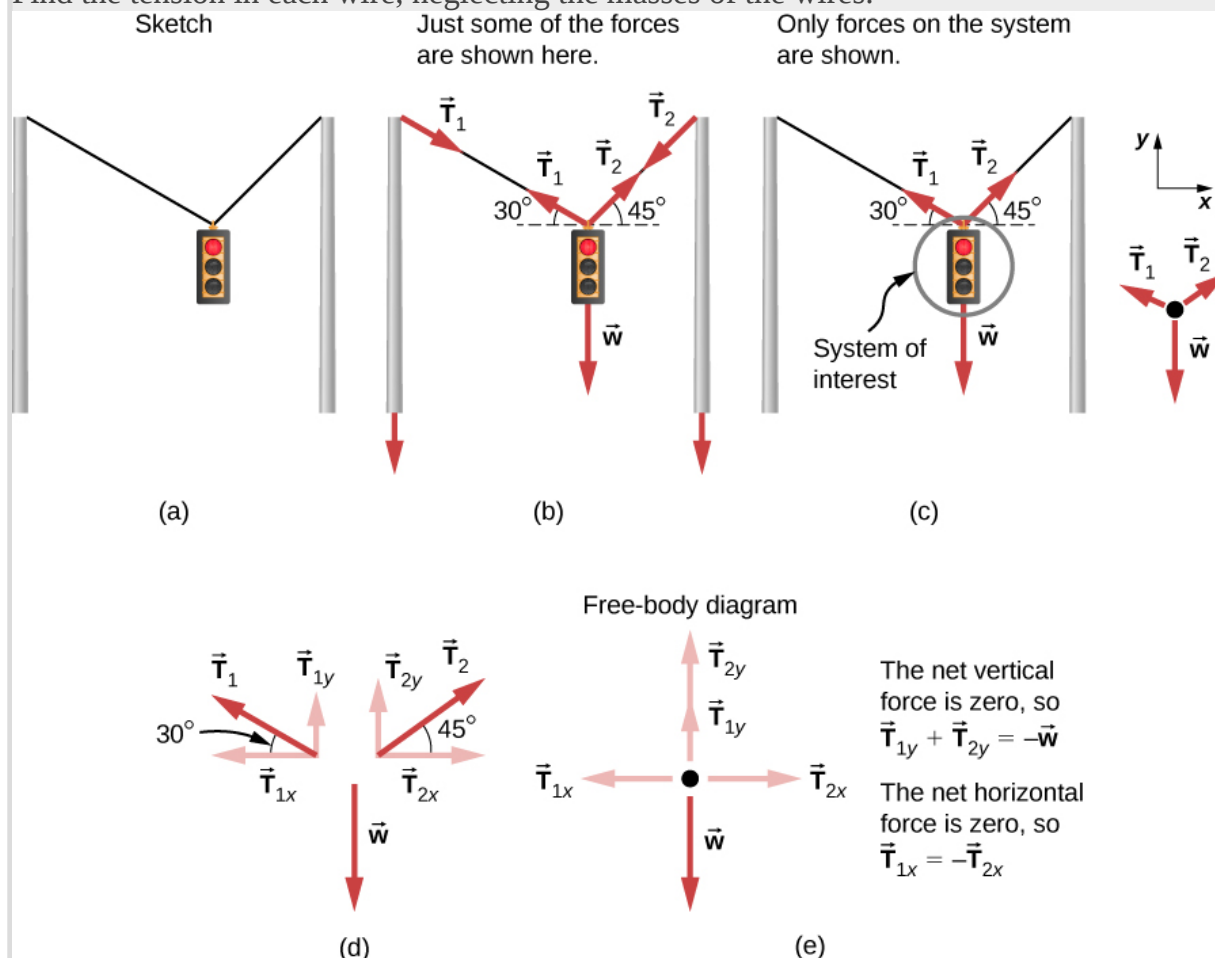
Recall that a particle in equilibrium is one for which the external forces are balanced. Static equilibrium involves objects at rest, and dynamic equilibrium involves objects in motion without acceleration, but it is important to remember that these conditions are relative. For example, an object may be at rest when viewed from our frame of reference, but the same object would appear to be in motion when viewed by someone moving at a constant velocity. We now make use of the knowledge attained in [Newton's Laws of Motion](#),

regarding the different types of forces and the use of free-body diagrams, to solve additional problems in particle equilibrium.

Example:

Different Tensions at Different Angles

Consider the traffic light (mass of 15.0 kg) suspended from two wires as shown in [\[link\]](#). Find the tension in each wire, neglecting the masses of the wires.



A traffic light is suspended from two wires. (b) Some of the forces involved. (c) Only forces acting on the system are shown here. The free-body diagram for the traffic light is also shown. (d) The forces projected onto vertical (y) and horizontal (x) axes. The horizontal components of the tensions must cancel, and the sum of the vertical components of the tensions must equal the weight of the traffic light. (e) The free-body diagram shows the vertical and horizontal forces acting on the traffic light.

Strategy

The system of interest is the traffic light, and its free-body diagram is shown in [\[link\]](#)(c). The three forces involved are not parallel, and so they must be projected onto a coordinate system. The most convenient coordinate system has one axis vertical and one horizontal, and the vector projections on it are shown in [\[link\]](#)(d). There are two unknowns in this problem (T_1 and T_2), so two equations are needed to find them. These two equations come from applying Newton's second law along the vertical and horizontal axes, noting that the net external force is zero along each axis because acceleration is zero.

Solution

First consider the horizontal or x -axis:

Equation:

$$F_{\text{net } x} = T_{2x} + T_{1x} = 0.$$

Thus, as you might expect,

Equation:

$$|T_{1x}| = |T_{2x}|.$$

This gives us the following relationship:

Equation:

$$T_1 \cos 30^\circ = T_2 \cos 45^\circ.$$

Thus,

Equation:

$$T_2 = 1.225T_1.$$

Note that T_1 and T_2 are not equal in this case because the angles on either side are not equal. It is reasonable that T_2 ends up being greater than T_1 because it is exerted more vertically than T_1 .

Now consider the force components along the vertical or y -axis:

Equation:

$$F_{\text{net } y} = T_{1y} + T_{2y} - w = 0.$$

This implies

Equation:

$$T_{1y} + T_{2y} = w.$$

Substituting the expressions for the vertical components gives

Equation:

$$T_1 \sin 30^\circ + T_2 \sin 45^\circ = w.$$

There are two unknowns in this equation, but substituting the expression for T_2 in terms of T_1 reduces this to one equation with one unknown:

Equation:

$$T_1(0.500) + (1.225T_1)(0.707) = w = mg,$$

which yields

Equation:

$$1.366T_1 = (15.0 \text{ kg})(9.80 \text{ m/s}^2).$$

Solving this last equation gives the magnitude of T_1 to be

Equation:

$$T_1 = 108 \text{ N}.$$

Finally, we find the magnitude of T_2 by using the relationship between them,

$T_2 = 1.225T_1$, found above. Thus we obtain

Equation:

$$T_2 = 132 \text{ N}.$$

Significance

Both tensions would be larger if both wires were more horizontal, and they will be equal if and only if the angles on either side are the same (as they were in the earlier example of a tightrope walker in [Newton's Laws of Motion](#)).

Particle Acceleration

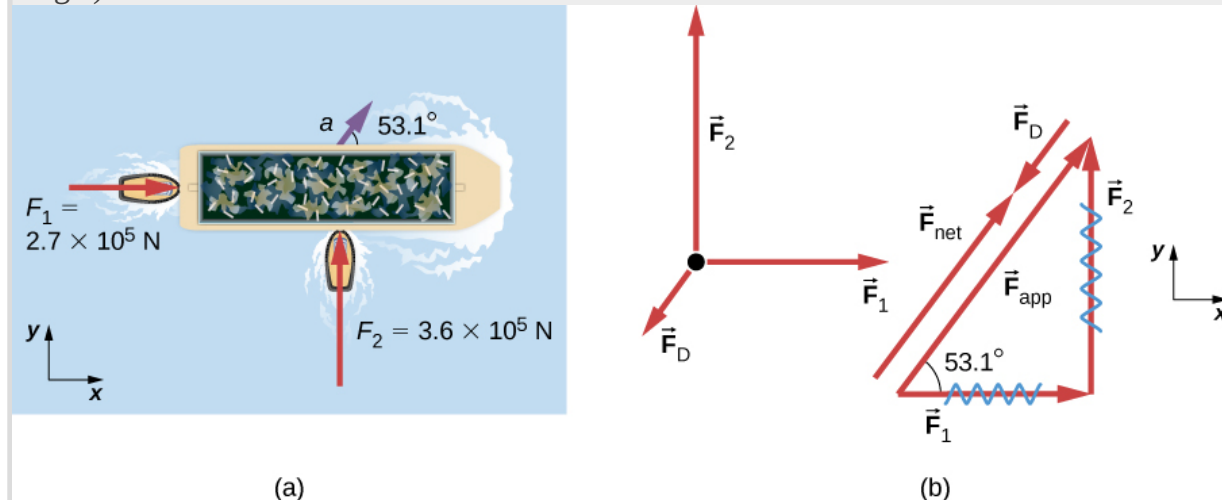
We have given a variety of examples of particles in equilibrium. We now turn our attention to particle acceleration problems, which are the result of a nonzero net force. Refer again to the steps given at the beginning of this section, and notice how they are applied to the following examples.

Example:

Drag Force on a Barge

Two tugboats push on a barge at different angles ([link](#)). The first tugboat exerts a force of $2.7 \times 10^5 \text{ N}$ in the x-direction, and the second tugboat exerts a force of $3.6 \times 10^5 \text{ N}$ in the y-direction. The mass of the barge is $5.0 \times 10^6 \text{ kg}$ and its acceleration is observed to be $7.5 \times 10^{-2} \text{ m/s}^2$ in the direction shown. What is the drag force of the water on the barge resisting the motion? (*Note:* Drag force is a frictional force exerted by fluids, such as air or water. The drag force opposes the motion of the object. Since the barge is flat

bottomed, we can assume that the drag force is in the direction opposite of motion of the barge.)



(a) A view from above of two tugboats pushing on a barge. (b) The free-body diagram for the ship contains only forces acting in the plane of the water. It omits the two vertical forces—the weight of the barge and the buoyant force of the water supporting it cancel and are not shown. Note that \vec{F}_{app} is the total applied force of the tugboats.

Strategy

The directions and magnitudes of acceleration and the applied forces are given in [\[link\]\(a\)](#). We define the total force of the tugboats on the barge as \vec{F}_{app} so that

Equation:

$$\vec{F}_{\text{app}} = \vec{F}_1 + \vec{F}_2.$$

The drag of the water \vec{F}_D is in the direction opposite to the direction of motion of the boat; this force thus works against \vec{F}_{app} , as shown in the free-body diagram in [\[link\]\(b\)](#). The system of interest here is the barge, since the forces on it are given as well as its acceleration. Because the applied forces are perpendicular, the x - and y -axes are in the same direction as \vec{F}_1 and \vec{F}_2 . The problem quickly becomes a one-dimensional problem along the direction of \vec{F}_{app} , since friction is in the direction opposite to \vec{F}_{app} . Our strategy is to find the magnitude and direction of the net applied force \vec{F}_{app} and then apply Newton's second law to solve for the drag force \vec{F}_D .

Solution

Since F_x and F_y are perpendicular, we can find the magnitude and direction of \vec{F}_{app} directly. First, the resultant magnitude is given by the Pythagorean theorem:

Equation:

$$F_{\text{app}} = \sqrt{F_1^2 + F_2^2} = \sqrt{(2.7 \times 10^5 \text{ N})^2 + (3.6 \times 10^5 \text{ N})^2} = 4.5 \times 10^5 \text{ N}.$$

The angle is given by

Equation:

$$\theta = \tan^{-1} \left(\frac{F_2}{F_1} \right) = \tan^{-1} \left(\frac{3.6 \times 10^5 \text{ N}}{2.7 \times 10^5 \text{ N}} \right) = 53.1^\circ.$$

From Newton's first law, we know this is the same direction as the acceleration. We also know that \vec{F}_D is in the opposite direction of \vec{F}_{app} , since it acts to slow down the acceleration. Therefore, the net external force is in the same direction as \vec{F}_{app} , but its magnitude is slightly less than \vec{F}_{app} . The problem is now one-dimensional. From the free-body diagram, we can see that

Equation:

$$F_{\text{net}} = F_{\text{app}} - F_D.$$

However, Newton's second law states that

Equation:

$$F_{\text{net}} = ma.$$

Thus,

Equation:

$$F_{\text{app}} - F_D = ma.$$

This can be solved for the magnitude of the drag force of the water F_D in terms of known quantities:

Equation:

$$F_D = F_{\text{app}} - ma.$$

Substituting known values gives

Equation:

$$F_D = (4.5 \times 10^5 \text{ N}) - (5.0 \times 10^6 \text{ kg}) (7.5 \times 10^{-2} \text{ m/s}^2) = 7.5 \times 10^4 \text{ N}.$$

The direction of \vec{F}_D has already been determined to be in the direction opposite to \vec{F}_{app} , or at an angle of 53° south of west.

Significance

The numbers used in this example are reasonable for a moderately large barge. It is certainly difficult to obtain larger accelerations with tugboats, and small speeds are desirable to avoid running the barge into the docks. Drag is relatively small for a well-

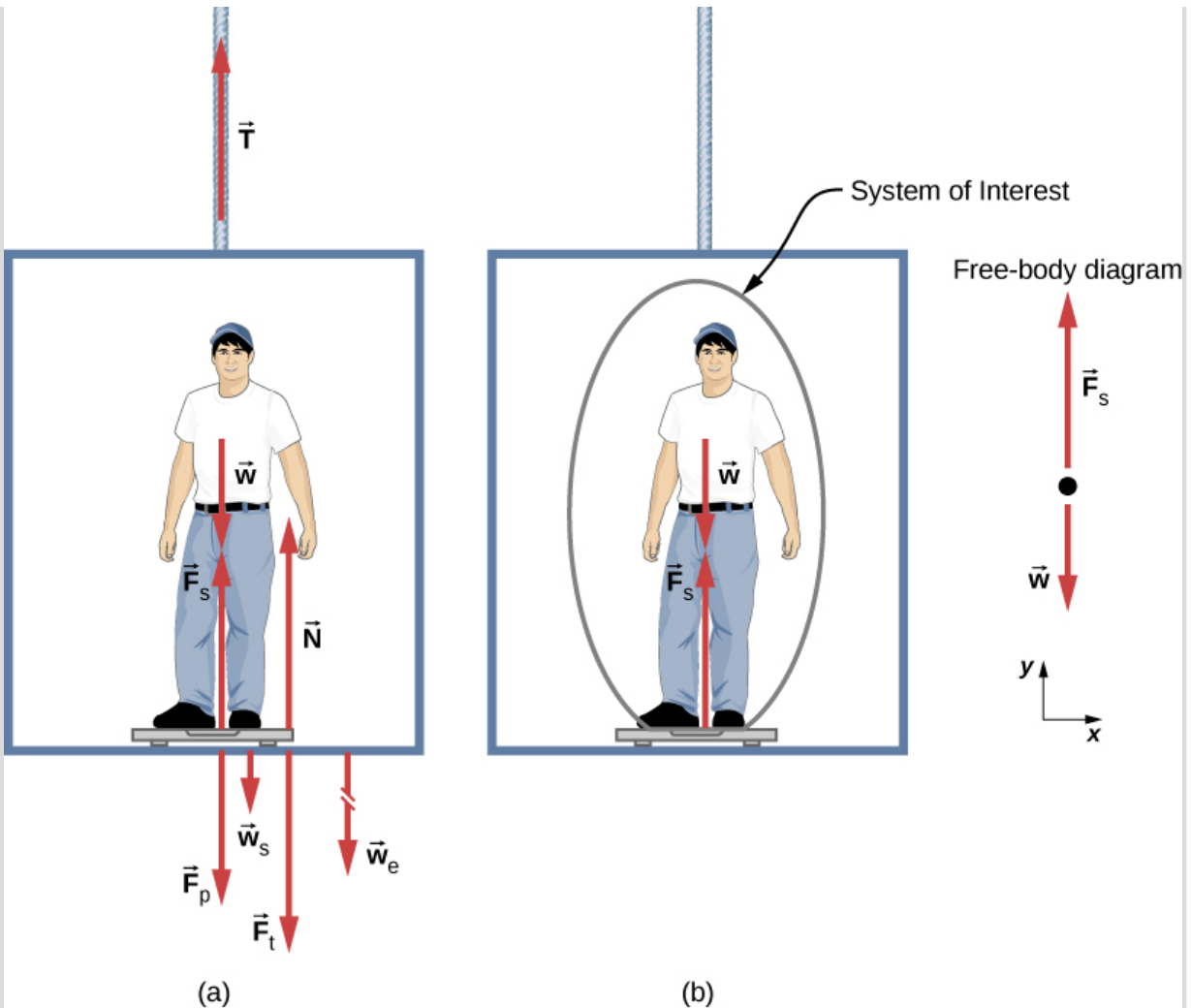
designed hull at low speeds, consistent with the answer to this example, where F_D is less than 1/600th of the weight of the ship.

In [Newton's Laws of Motion](#), we discussed the normal force, which is a contact force that acts normal to the surface so that an object does not have an acceleration perpendicular to the surface. The bathroom scale is an excellent example of a normal force acting on a body. It provides a quantitative reading of how much it must push upward to support the weight of an object. But can you predict what you would see on the dial of a bathroom scale if you stood on it during an elevator ride? Will you see a value greater than your weight when the elevator starts up? What about when the elevator moves upward at a constant speed? Take a guess before reading the next example.

Example:

What Does the Bathroom Scale Read in an Elevator?

[\[link\]](#) shows a 75.0-kg man (weight of about 165 lb.) standing on a bathroom scale in an elevator. Calculate the scale reading: (a) if the elevator accelerates upward at a rate of 1.20 m/s^2 , and (b) if the elevator moves upward at a constant speed of 1 m/s.



(a) The various forces acting when a person stands on a bathroom scale in an elevator. The arrows are approximately correct for when the elevator is accelerating upward—broken arrows represent forces too large to be drawn to scale. \vec{T} is the tension in the supporting cable, \vec{W} is the weight of the person, \vec{W}_s is the weight of the scale, \vec{W}_e is the weight of the elevator, \vec{F}_s is the force of the scale on the person, \vec{F}_p is the force of the person on the scale, \vec{F}_t is the force of the scale on the floor of the elevator, and \vec{N} is the force of the floor upward on the scale. (b) The free-body diagram shows only the external forces acting *on* the designated system of interest—the person—and is the diagram we use for the solution of the problem.

Strategy

If the scale at rest is accurate, its reading equals \vec{F}_p , the magnitude of the force the person exerts downward on it. [\[link\]](#)(a) shows the numerous forces acting on the elevator, scale, and person. It makes this one-dimensional problem look much more formidable than if the

person is chosen to be the system of interest and a free-body diagram is drawn, as in [\[link\]](#) (b). Analysis of the free-body diagram using Newton's laws can produce answers to both [\[link\]](#) (a) and (b) of this example, as well as some other questions that might arise. The only forces acting on the person are his weight \vec{w} and the upward force of the scale \vec{F}_s .

According to Newton's third law, \vec{F}_p and \vec{F}_s are equal in magnitude and opposite in direction, so that we need to find F_s in order to find what the scale reads. We can do this, as usual, by applying Newton's second law,

Equation:

$$\vec{F}_{\text{net}} = m\vec{a}.$$

From the free-body diagram, we see that $\vec{F}_{\text{net}} = \vec{F}_s - \vec{w}$, so we have

Equation:

$$F_s - w = ma.$$

Solving for F_s gives us an equation with only one unknown:

Equation:

$$F_s = ma + w,$$

or, because $w = mg$, simply

Equation:

$$F_s = ma + mg.$$

No assumptions were made about the acceleration, so this solution should be valid for a variety of accelerations in addition to those in this situation. (*Note:* We are considering the case when the elevator is accelerating upward. If the elevator is accelerating downward, Newton's second law becomes $F_s - w = -ma$.)

Solution

- a. We have $a = 1.20 \text{ m/s}^2$, so that

Equation:

$$F_s = (75.0 \text{ kg})(9.80 \text{ m/s}^2) + (75.0 \text{ kg})(1.20 \text{ m/s}^2)$$

yielding

Equation:

$$F_s = 825 \text{ N}.$$

- b. Now, what happens when the elevator reaches a constant upward velocity? Will the scale still read more than his weight? For any constant velocity—up, down, or stationary—acceleration is zero because $a = \frac{\Delta v}{\Delta t}$ and $\Delta v = 0$. Thus,

Equation:

$$F_s = ma + mg = 0 + mg$$

or

Equation:

$$F_s = (75.0 \text{ kg})(9.80 \text{ m/s}^2),$$

which gives

Equation:

$$F_s = 735 \text{ N}.$$

Significance

The scale reading in [\[link\]](#)(a) is about 185 lb. What would the scale have read if he were stationary? Since his acceleration would be zero, the force of the scale would be equal to his weight:

Equation:

$$F_{\text{net}} = ma = 0 = F_s - w$$

Equation:

$$F_s = w = mg$$

Equation:

$$F_s = (75.0 \text{ kg})(9.80 \text{ m/s}^2) = 735 \text{ N}.$$

Thus, the scale reading in the elevator is greater than his 735-N (165-lb.) weight. This means that the scale is pushing up on the person with a force greater than his weight, as it must in order to accelerate him upward. Clearly, the greater the acceleration of the elevator, the greater the scale reading, consistent with what you feel in rapidly accelerating versus slowly accelerating elevators. In [\[link\]](#)(b), the scale reading is 735 N, which equals the person's weight. This is the case whenever the elevator has a constant velocity—moving up, moving down, or stationary.

Note:

Exercise:

Problem:

Check Your Understanding Now calculate the scale reading when the elevator accelerates downward at a rate of 1.20 m/s^2 .

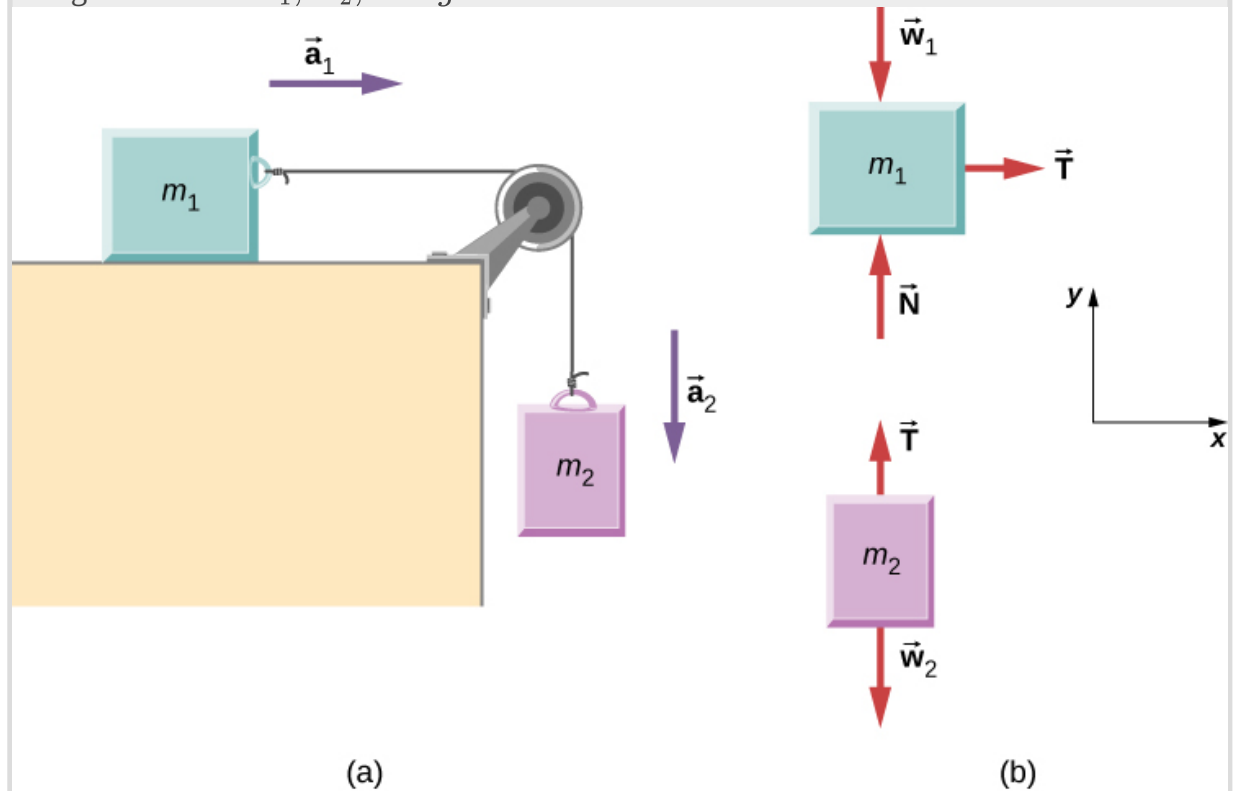
Solution:

$$F_s = 645 \text{ N}$$

The solution to the previous example also applies to an elevator accelerating downward, as mentioned. When an elevator accelerates downward, a is negative, and the scale reading is *less* than the weight of the person. If a constant downward velocity is reached, the scale reading again becomes equal to the person's weight. If the elevator is in free fall and accelerating downward at g , then the scale reading is zero and the person appears to be weightless.

Example:**Two Attached Blocks**

[\[link\]](#) shows a block of mass m_1 on a frictionless, horizontal surface. It is pulled by a light string that passes over a frictionless and massless pulley. The other end of the string is connected to a block of mass m_2 . Find the acceleration of the blocks and the tension in the string in terms of m_1 , m_2 , and g .



(a) Block 1 is connected by a light string to block 2. (b) The free-body diagrams of the blocks.

Strategy

We draw a free-body diagram for each mass separately, as shown in [\[link\]](#). Then we analyze each one to find the required unknowns. The forces on block 1 are the gravitational force, the contact force of the surface, and the tension in the string. Block 2 is subjected to the gravitational force and the string tension. Newton's second law applies to each, so we write two vector equations:

For block 1: $\vec{T} + \vec{w}_1 + \vec{N} = m_1 \vec{a}_1$

For block 2: $\vec{T} + \vec{w}_2 = m_2 \vec{a}_2$.

Notice that \vec{T} is the same for both blocks. Since the string and the pulley have negligible mass, and since there is no friction in the pulley, the tension is the same throughout the string. We can now write component equations for each block. All forces are either horizontal or vertical, so we can use the same horizontal/vertical coordinate system for both objects

Solution

The component equations follow from the vector equations above. We see that block 1 has the vertical forces balanced, so we ignore them and write an equation relating the x -components. There are no horizontal forces on block 2, so only the y -equation is written. We obtain these results:

Equation:

Block 1	Block 2
$\sum F_x = ma_x$	$\sum F_y = ma_y$
$T_x = m_1 a_{1x}$	$T_y - m_2 g = m_2 a_{2y}$

When block 1 moves to the right, block 2 travels an equal distance downward; thus, $a_{1x} = -a_{2y}$. Writing the common acceleration of the blocks as $a = a_{1x} = -a_{2y}$, we now have

Equation:

$$T = m_1 a$$

and

Equation:

$$T - m_2 g = -m_2 a.$$

From these two equations, we can express a and T in terms of the masses m_1 and m_2 , and g :

Equation:

$$a = \frac{m_2}{m_1 + m_2} g$$

and

Equation:

$$T = \frac{m_1 m_2}{m_1 + m_2} g.$$

Significance

Notice that the tension in the string is *less* than the weight of the block hanging from the end of it. A common error in problems like this is to set $T = m_2 g$. You can see from the free-body diagram of block 2 that cannot be correct if the block is accelerating.

Note:

Exercise:

Problem:

Check Your Understanding Calculate the acceleration of the system, and the tension in the string, when the masses are $m_1 = 5.00$ kg and $m_2 = 3.00$ kg.

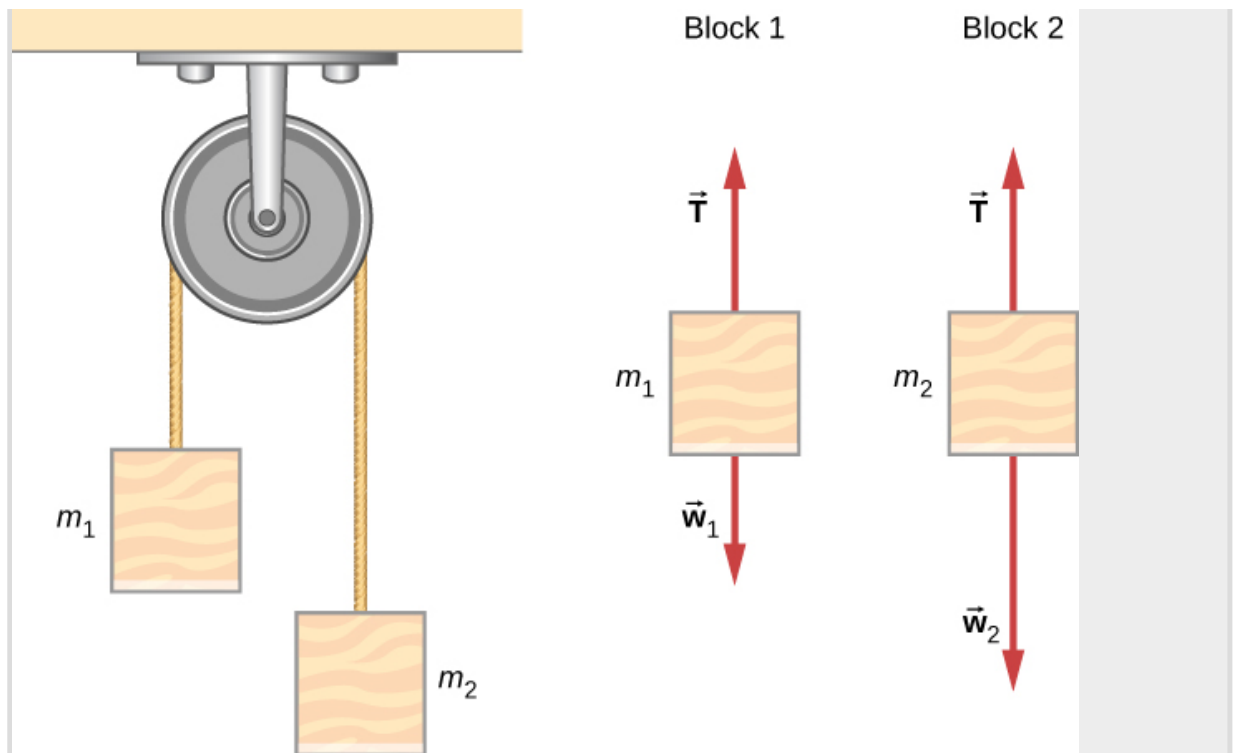
Solution:

$$a = 3.68 \text{ m/s}^2, T = 18.4 \text{ N}$$

Example:

Atwood Machine

A classic problem in physics, similar to the one we just solved, is that of the Atwood machine, which consists of a rope running over a pulley, with two objects of different mass attached. It is particularly useful in understanding the connection between force and motion. In [\[link\]](#), $m_1 = 2.00$ kg and $m_2 = 4.00$ kg. Consider the pulley to be frictionless. (a) If m_2 is released, what will its acceleration be? (b) What is the tension in the string?



An Atwood machine and free-body diagrams for each of the two blocks.

Strategy

We draw a free-body diagram for each mass separately, as shown in the figure. Then we analyze each diagram to find the required unknowns. This may involve the solution of simultaneous equations. It is also important to note the similarity with the previous example. As block 2 accelerates with acceleration a_2 in the downward direction, block 1 accelerates upward with acceleration a_1 . Thus, $a = a_1 = -a_2$.

Solution

a. We have

Equation:

$$\text{For } m_1, \sum F_y = T - m_1g = m_1a. \quad \text{For } m_2, \sum F_y = T - m_2g = -m_2a.$$

(The negative sign in front of m_2a indicates that m_2 accelerates downward; both blocks accelerate at the same rate, but in opposite directions.) Solve the two equations simultaneously (subtract them) and the result is

Equation:

$$(m_2 - m_1)g = (m_1 + m_2)a.$$

Solving for a :

Equation:

$$a = \frac{m_2 - m_1}{m_1 + m_2}g = \frac{4 \text{ kg} - 2 \text{ kg}}{4 \text{ kg} + 2 \text{ kg}}(9.8 \text{ m/s}^2) = 3.27 \text{ m/s}^2.$$

b. Observing the first block, we see that

Equation:

$$T - m_1g = m_1a$$

$$T = m_1(g + a) = (2 \text{ kg})(9.8 \text{ m/s}^2 + 3.27 \text{ m/s}^2) = 26.1 \text{ N}.$$

Significance

The result for the acceleration given in the solution can be interpreted as the ratio of the unbalanced force on the system, $(m_2 - m_1)g$, to the total mass of the system, $m_1 + m_2$. We can also use the Atwood machine to measure local gravitational field strength.

Note:**Exercise:****Problem:**

Check Your Understanding Determine a general formula in terms of m_1 , m_2 and g for calculating the tension in the string for the Atwood machine shown above.

Solution:

$T = \frac{2m_1m_2}{m_1+m_2}g$ (This is found by substituting the equation for acceleration in [\[link\]\(a\)](#), into the equation for tension in [\[link\]\(b\)](#).)

Newton's Laws of Motion and Kinematics

Physics is most interesting and most powerful when applied to general situations that involve more than a narrow set of physical principles. Newton's laws of motion can also be integrated with other concepts that have been discussed previously in this text to solve problems of motion. For example, forces produce accelerations, a topic of kinematics, and hence the relevance of earlier chapters.

When approaching problems that involve various types of forces, acceleration, velocity, and/or position, listing the givens and the quantities to be calculated will allow you to identify the principles involved. Then, you can refer to the chapters that deal with a

particular topic and solve the problem using strategies outlined in the text. The following worked example illustrates how the problem-solving strategy given earlier in this chapter, as well as strategies presented in other chapters, is applied to an integrated concept problem.

Example:

What Force Must a Soccer Player Exert to Reach Top Speed?

A soccer player starts at rest and accelerates forward, reaching a velocity of 8.00 m/s in 2.50 s. (a) What is her average acceleration? (b) What average force does the ground exert forward on the runner so that she achieves this acceleration? The player's mass is 70.0 kg, and air resistance is negligible.

Strategy

To find the answers to this problem, we use the problem-solving strategy given earlier in this chapter. The solutions to each part of the example illustrate how to apply specific problem-solving steps. In this case, we do not need to use all of the steps. We simply identify the physical principles, and thus the knowns and unknowns; apply Newton's second law; and check to see whether the answer is reasonable.

Solution

- a. We are given the initial and final velocities (zero and 8.00 m/s forward); thus, the change in velocity is $\Delta v = 8.00 \text{ m/s}$. We are given the elapsed time, so $\Delta t = 2.50 \text{ s}$. The unknown is acceleration, which can be found from its definition:

Equation:

$$a = \frac{\Delta v}{\Delta t}.$$

Substituting the known values yields

Equation:

$$a = \frac{8.00 \text{ m/s}}{2.50 \text{ s}} = 3.20 \text{ m/s}^2.$$

- b. Here we are asked to find the average force the ground exerts on the runner to produce this acceleration. (Remember that we are dealing with the force or forces acting on the object of interest.) This is the reaction force to that exerted by the player backward against the ground, by Newton's third law. Neglecting air resistance, this would be equal in magnitude to the net external force on the player, since this force causes her acceleration. Since we now know the player's acceleration and are given her mass, we can use Newton's second law to find the force exerted. That is,

Equation:

$$F_{\text{net}} = ma.$$

Substituting the known values of m and a gives

Equation:

$$F_{\text{net}} = (70.0 \text{ kg})(3.20 \text{ m/s}^2) = 224 \text{ N}.$$

This is a reasonable result: The acceleration is attainable for an athlete in good condition. The force is about 50 pounds, a reasonable average force.

Significance

This example illustrates how to apply problem-solving strategies to situations that include topics from different chapters. The first step is to identify the physical principles, the knowns, and the unknowns involved in the problem. The second step is to solve for the unknown, in this case using Newton's second law. Finally, we check our answer to ensure it is reasonable. These techniques for integrated concept problems will be useful in applications of physics outside of a physics course, such as in your profession, in other science disciplines, and in everyday life.

Note:**Exercise:****Problem:**

Check Your Understanding The soccer player stops after completing the play described above, but now notices that the ball is in position to be stolen. If she now experiences a force of 126 N to attempt to steal the ball, which is 2.00 m away from her, how long will it take her to get to the ball?

Solution:

1.49 s

Example:**What Force Acts on a Model Helicopter?**

A 1.50-kg model helicopter has a velocity of $5.00\hat{\mathbf{j}} \text{ m/s}$ at $t = 0$. It is accelerated at a constant rate for two seconds (2.00 s) after which it has a velocity of

$(6.00\hat{\mathbf{i}} + 12.00\hat{\mathbf{j}}) \text{ m/s}$. What is the magnitude of the resultant force acting on the helicopter during this time interval?

Strategy

We can easily set up a coordinate system in which the x-axis ($\hat{\mathbf{i}}$ direction) is horizontal, and the y-axis ($\hat{\mathbf{j}}$ direction) is vertical. We know that $\Delta t = 2.00 \text{ s}$ and

$\Delta v = (6.00\hat{\mathbf{i}} + 12.00\hat{\mathbf{j}} \text{ m/s}) - (5.00\hat{\mathbf{j}} \text{ m/s})$. From this, we can calculate the acceleration by the definition; we can then apply Newton's second law.

Solution

We have

Equation:

$$a = \frac{\Delta v}{\Delta t} = \frac{(6.00\hat{\mathbf{i}} + 12.00\hat{\mathbf{j}} \text{ m/s}) - (5.00\hat{\mathbf{j}} \text{ m/s})}{2.00 \text{ s}} = 3.00\hat{\mathbf{i}} + 3.50\hat{\mathbf{j}} \text{ m/s}^2$$

Equation:

$$\sum \vec{\mathbf{F}} = m\vec{\mathbf{a}} = (1.50 \text{ kg})(3.00\hat{\mathbf{i}} + 3.50\hat{\mathbf{j}} \text{ m/s}^2) = 4.50\hat{\mathbf{i}} + 5.25\hat{\mathbf{j}} \text{ N}.$$

The magnitude of the force is now easily found:

Equation:

$$F = \sqrt{(4.50 \text{ N})^2 + (5.25 \text{ N})^2} = 6.91 \text{ N}.$$

Significance

The original problem was stated in terms of $\hat{\mathbf{i}} - \hat{\mathbf{j}}$ vector components, so we used vector methods. Compare this example with the previous example.

Note:

Exercise:

Problem:

Check Your Understanding Find the direction of the resultant for the 1.50-kg model helicopter.

Solution:

49.4 degrees

Example:

Baggage Tractor

[\[link\]](#) (a) shows a baggage tractor pulling luggage carts from an airplane. The tractor has mass 650.0 kg, while cart A has mass 250.0 kg and cart B has mass 150.0 kg. The driving force acting for a brief period of time accelerates the system from rest and acts for 3.00 s. (a) If this driving force is given by $F = (820.0t) \text{ N}$, find the speed after 3.00 seconds. (b)

Diagram (a) shows three objects: a tractor, Cart A, and Cart B. The tractor has a normal force \vec{N}_{tractor} pointing up, a weight \vec{w}_{tractor} pointing down, and a friction force \vec{F}_{tractor} pointing left. Cart A has a normal force \vec{N}_A pointing up and a weight \vec{w}_A pointing down. Cart B has a normal force \vec{N}_B pointing up and a weight \vec{w}_B pointing down. The tractor is connected to Cart A, which is connected to Cart B.

Diagram (b) shows the tractor with an additional tension force \vec{T} pointing right, representing the force exerted by Cart A on the tractor.

Strategy

This exposes the coupling force $\vec{\mathbf{T}}$, which is our objective.

a. $\sum F_x = m_{\text{system}} a_x$ and $\sum F_x = 820.0t$, so

$$\begin{aligned} 820.0t &= (650.0 + 250.0 + 150.0)a \\ a &= 0.7809t. \end{aligned}$$

Equation:

b. Refer to the free-body diagram in [\[link\]](#)(b).

Equation:

$$\begin{aligned}
 \sum F_x &= m_{\text{tractor}} a_x \\
 820.0t - T &= m_{\text{tractor}}(0.7805)t \\
 (820.0)(3.00) - T &= (650.0)(0.7805)(3.00) \\
 T &= 938 \text{ N.}
 \end{aligned}$$

Significance

Since the force varies with time, we must use calculus to solve this problem. Notice how the total mass of the system was important in solving [\[link\]\(a\)](#), whereas only the mass of the truck (since it supplied the force) was of use in [\[link\]\(b\)](#).

Recall that $v = \frac{ds}{dt}$ and $a = \frac{dv}{dt}$. If acceleration is a function of time, we can use the calculus forms developed in [Motion Along a Straight Line](#), as shown in this example. However, sometimes acceleration is a function of displacement. In this case, we can derive an important result from these calculus relations. Solving for dt in each, we have $dt = \frac{ds}{v}$ and $dt = \frac{dv}{a}$. Now, equating these expressions, we have $\frac{ds}{v} = \frac{dv}{a}$. We can rearrange this to obtain $a ds = v dv$.

Example:

Motion of a Projectile Fired Vertically

A 10.0-kg mortar shell is fired vertically upward from the ground, with an initial velocity of 50.0 m/s (see [\[link\]](#)). Determine the maximum height it will travel if atmospheric resistance is measured as $F_D = (0.0100v^2)$ N, where v is the speed at any instant.



(a)



(b)

(a) The mortar fires a shell straight up; we consider the friction force provided by the air. (b) A free-body diagram is shown which indicates all the forces on the mortar shell. (credit a: modification of work by OS541/DoD; The appearance of U.S. Department of Defense (DoD) visual information does not imply or constitute DoD endorsement.)

Strategy

The known force on the mortar shell can be related to its acceleration using the equations of motion. Kinematics can then be used to relate the mortar shell's acceleration to its position.

Solution

Initially, $y_0 = 0$ and $v_0 = 50.0 \text{ m/s}$. At the maximum height $y = h$, $v = 0$. The free-body diagram shows F_D to act downward, because it slows the upward motion of the mortar shell. Thus, we can write

Equation:

$$\sum F_y = ma_y$$

Equation:

$$\begin{aligned} -F_D - w &= ma_y \\ -0.0100v^2 - 98.0 &= 10.0a \\ a &= -0.00100v^2 - 9.80. \end{aligned}$$

The acceleration depends on v and is therefore variable. Since $a = f(v)$, we can relate a to v using the rearrangement described above,

Equation:

$$a ds = v dv.$$

We replace ds with dy because we are dealing with the vertical direction,

Equation:

$$a dy = v dv, \quad (-0.00100v^2 - 9.80) dy = v dv.$$

We now separate the variables (v 's and dv 's on one side; dy on the other):

Equation:

$$\begin{aligned} \int_0^h dy &= \int_{50.0}^0 \frac{v dv}{(-0.00100v^2 - 9.80)} \\ \int_0^h dy &= - \int_{50.0}^0 \frac{v dv}{(0.00100v^2 + 9.80)} = (-5 \times 10^3) \ln(0.00100v^2 + 9.80) \Big|_{50.0}^0. \end{aligned}$$

Thus, $h = 114$ m.

Significance

Notice the need to apply calculus since the force is not constant, which also means that acceleration is not constant. To make matters worse, the force depends on v (not t), and so we must use the trick explained prior to the example. The answer for the height indicates a lower elevation if there were air resistance. We will deal with the effects of air resistance and other drag forces in greater detail in [Drag Force and Terminal Speed](#).

Note:

Exercise:

Problem:

Check Your Understanding If atmospheric resistance is neglected, find the maximum height for the mortar shell. Is calculus required for this solution?

Solution:

128 m; no

Note:

Explore the forces at work in this [simulation](#) when you try to push a filing cabinet. Create an applied force and see the resulting frictional force and total force acting on the cabinet. Charts show the forces, position, velocity, and acceleration vs. time. View a free-body diagram of all the forces (including gravitational and normal forces).

Summary

- Newton's laws of motion can be applied in numerous situations to solve motion problems.
- Some problems contain multiple force vectors acting in different directions on an object. Be sure to draw diagrams, resolve all force vectors into horizontal and vertical components, and draw a free-body diagram. Always analyze the direction in which an object accelerates so that you can determine whether $F_{\text{net}} = ma$ or $F_{\text{net}} = 0$.
- The normal force on an object is not always equal in magnitude to the weight of the object. If an object is accelerating vertically, the normal force is less than or greater than the weight of the object. Also, if the object is on an inclined plane, the normal force is always less than the full weight of the object.
- Some problems contain several physical quantities, such as forces, acceleration, velocity, or position. You can apply concepts from kinematics and dynamics to solve these problems.

Conceptual Questions

Exercise:**Problem:**

To simulate the apparent weightlessness of space orbit, astronauts are trained in the hold of a cargo aircraft that is accelerating downward at g . Why do they appear to be weightless, as measured by standing on a bathroom scale, in this accelerated frame of reference? Is there any difference between their apparent weightlessness in orbit and in the aircraft?

Solution:

The scale is in free fall along with the astronauts, so the reading on the scale would be 0. There is no difference in the apparent weightlessness; in the aircraft and in orbit, free fall is occurring.

Problems

Exercise:

Problem:

A 30.0-kg girl in a swing is pushed to one side and held at rest by a horizontal force \vec{F} so that the swing ropes are 30.0° with respect to the vertical. (a) Calculate the tension in each of the two ropes supporting the swing under these conditions. (b) Calculate the magnitude of \vec{F} .

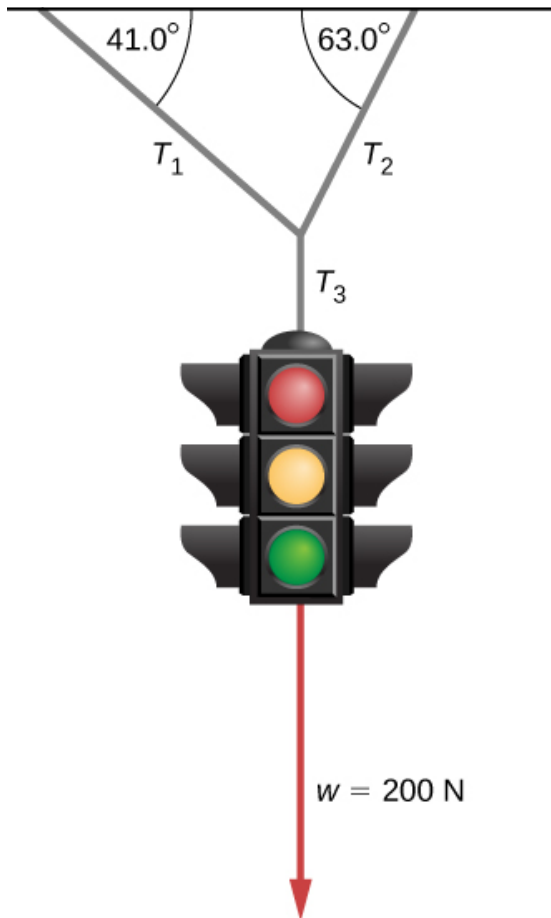
Solution:

a. 170 N; b. 170 N

Exercise:

Problem:

Find the tension in each of the three cables supporting the traffic light if it weighs 2.00×10^2 N.



Exercise:

Problem:

Three forces act on an object, considered to be a particle, which moves with constant velocity $v = (3\hat{i} - 2\hat{j}) \text{ m/s}$. Two of the forces are $\vec{F}_1 = (3\hat{i} + 5\hat{j} - 6\hat{k}) \text{ N}$ and $\vec{F}_2 = (4\hat{i} - 7\hat{j} + 2\hat{k}) \text{ N}$. Find the third force.

Solution:

$$\vec{F}_3 = (-7\hat{i} + 2\hat{j} + 4\hat{k}) \text{ N}$$

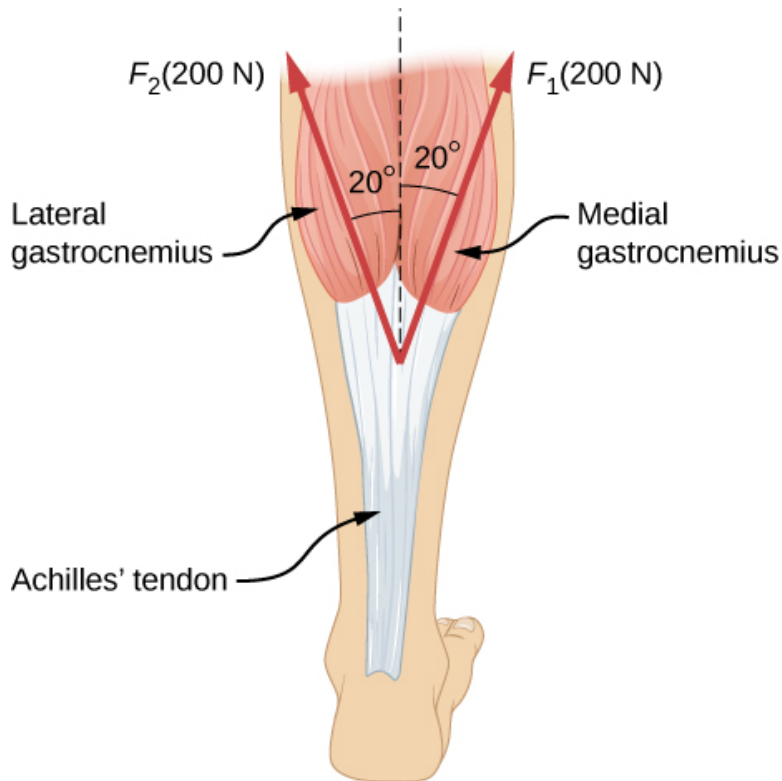
Exercise:

Problem:

A flea jumps by exerting a force of $1.20 \times 10^{-5} \text{ N}$ straight down on the ground. A breeze blowing on the flea parallel to the ground exerts a force of $0.500 \times 10^{-6} \text{ N}$ on the flea while the flea is still in contact with the ground. Find the direction and magnitude of the acceleration of the flea if its mass is $6.00 \times 10^{-7} \text{ kg}$. Do not neglect the gravitational force.

Exercise:**Problem:**

Two muscles in the back of the leg pull upward on the Achilles tendon, as shown below. (These muscles are called the medial and lateral heads of the gastrocnemius muscle.) Find the magnitude and direction of the total force on the Achilles tendon. What type of movement could be caused by this force?

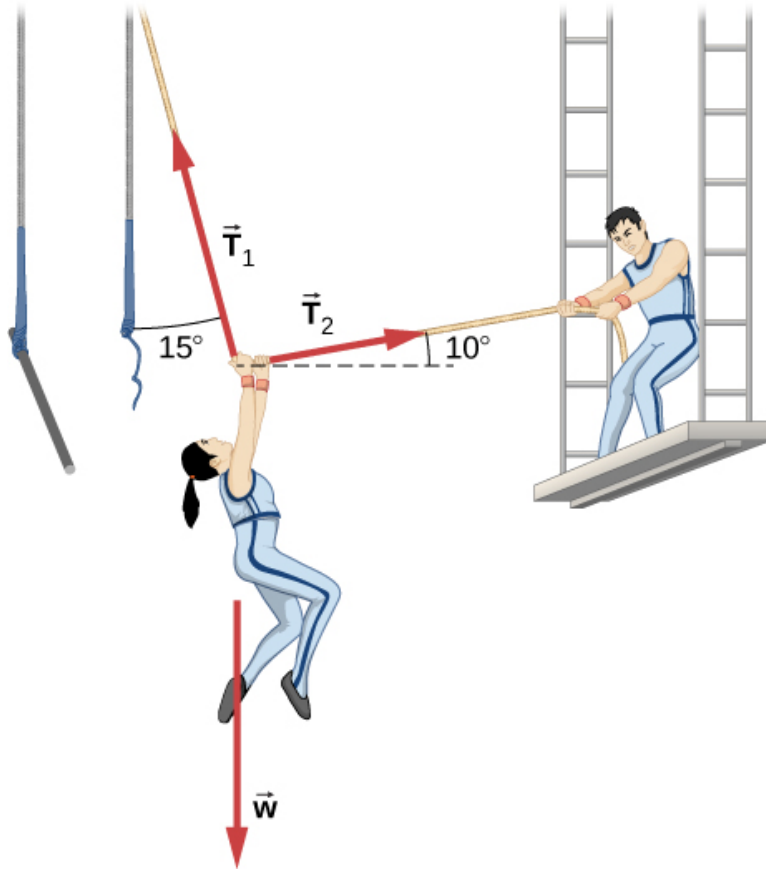


Solution:

376 N pointing up (along the dashed line in the figure); the force is used to raise the heel of the foot.

Exercise:**Problem:**

After a mishap, a 76.0-kg circus performer clings to a trapeze, which is being pulled to the side by another circus artist, as shown here. Calculate the tension in the two ropes if the person is momentarily motionless. Include a free-body diagram in your solution.



Exercise:

Problem:

A 35.0-kg dolphin accelerates opposite to the motion from 12.0 to 7.50 m/s in 2.30 s to join another dolphin in play. What average force was exerted to slow the first dolphin if it was moving horizontally? (The gravitational force is balanced by the buoyant force of the water.)

Solution:

-68.5 N

Exercise:

Problem:

When starting a foot race, a 70.0-kg sprinter exerts an average force of 650 N backward on the ground for 0.800 s. (a) What is his final speed? (b) How far does he travel?

Exercise:

Problem:

A large rocket has a mass of 2.00×10^6 kg at takeoff, and its engines produce a thrust of 3.50×10^7 N. (a) Find its initial acceleration if it takes off vertically. (b) How long does it take to reach a velocity of 120 km/h straight up, assuming constant mass and thrust?

Solution:

a. 7.70 m/s^2 ; b. 4.33 s

Exercise:**Problem:**

A basketball player jumps straight up for a ball. To do this, he lowers his body 0.300 m and then accelerates through this distance by forcefully straightening his legs. This player leaves the floor with a vertical velocity sufficient to carry him 0.900 m above the floor. (a) Calculate his velocity when he leaves the floor. (b) Calculate his acceleration while he is straightening his legs. He goes from zero to the velocity found in (a) in a distance of 0.300 m. (c) Calculate the force he exerts on the floor to do this, given that his mass is 110.0 kg.

Exercise:**Problem:**

A 2.50-kg fireworks shell is fired straight up from a mortar and reaches a height of 110.0 m. (a) Neglecting air resistance (a poor assumption, but we will make it for this example), calculate the shell's velocity when it leaves the mortar. (b) The mortar itself is a tube 0.450 m long. Calculate the average acceleration of the shell in the tube as it goes from zero to the velocity found in (a). (c) What is the average force on the shell in the mortar? Express your answer in newtons and as a ratio to the weight of the shell.

Solution:

a. 46.4 m/s ; b. $2.40 \times 10^3 \text{ m/s}^2$; c. $5.99 \times 10^3 \text{ N}$; ratio of 245

Exercise:**Problem:**

A 0.500-kg potato is fired at an angle of 80.0° above the horizontal from a PVC pipe used as a "potato gun" and reaches a height of 110.0 m. (a) Neglecting air resistance, calculate the potato's velocity when it leaves the gun. (b) The gun itself is a tube 0.450 m long. Calculate the average acceleration of the potato in the tube as it goes from zero to the velocity found in (a). (c) What is the average force on the potato in the gun? Express your answer in newtons and as a ratio to the weight of the potato.

Exercise:**Problem:**

An elevator filled with passengers has a mass of 1.70×10^3 kg. (a) The elevator accelerates upward from rest at a rate of 1.20 m/s^2 for 1.50 s. Calculate the tension in the cable supporting the elevator. (b) The elevator continues upward at constant velocity for 8.50 s. What is the tension in the cable during this time? (c) The elevator accelerates opposite to the motion at a rate of 0.600 m/s^2 for 3.00 s. What is the tension in the cable during acceleration opposite to the motion? (d) How high has the elevator moved above its original starting point, and what is its final velocity?

Solution:

a. 1.87×10^4 N; b. 1.67×10^4 N; c. 1.56×10^4 N; d. 19.4 m, 0 m/s

Exercise:**Problem:**

A 20.0-g ball hangs from the roof of a freight car by a string. When the freight car begins to move, the string makes an angle of 35.0° with the vertical. (a) What is the acceleration of the freight car? (b) What is the tension in the string?

Exercise:**Problem:**

A student's backpack, full of textbooks, is hung from a spring scale attached to the ceiling of an elevator. When the elevator is accelerating downward at 3.8 m/s^2 , the scale reads 60 N. (a) What is the mass of the backpack? (b) What does the scale read if the elevator moves upward while speeding up at a rate 3.8 m/s^2 ? (c) What does the scale read if the elevator moves upward at constant velocity? (d) If the elevator had no brakes and the cable supporting it were to break loose so that the elevator could fall freely, what would the spring scale read?

Solution:

a. 10 kg; b. 140 N; c. 98 N; d. 0

Exercise:**Problem:**

A service elevator takes a load of garbage, mass 10.0 kg, from a floor of a skyscraper under construction, down to ground level, accelerating downward at a rate of 1.2 m/s^2 . Find the magnitude of the force the garbage exerts on the floor of the service elevator?

Exercise:

Problem:

A roller coaster car starts from rest at the top of a track 30.0 m long and inclined at 20.0° to the horizontal. Assume that friction can be ignored. (a) What is the acceleration of the car? (b) How much time elapses before it reaches the bottom of the track?

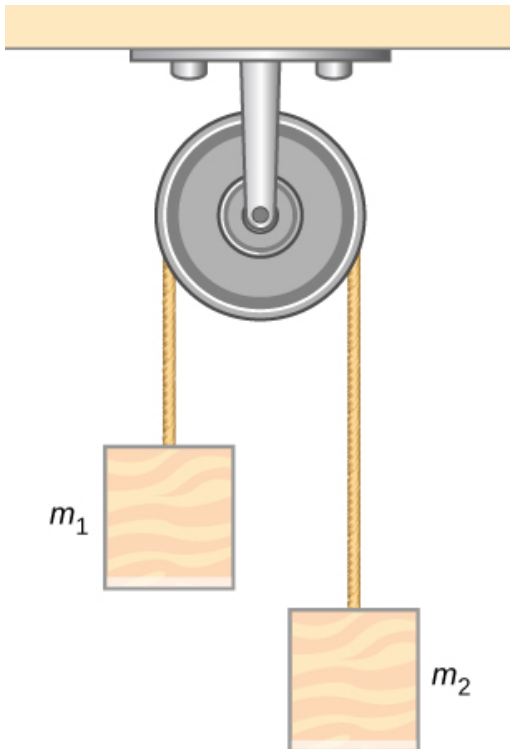
Solution:

a. 3.35 m/s^2 ; b. 4.2 s

Exercise:

Problem:

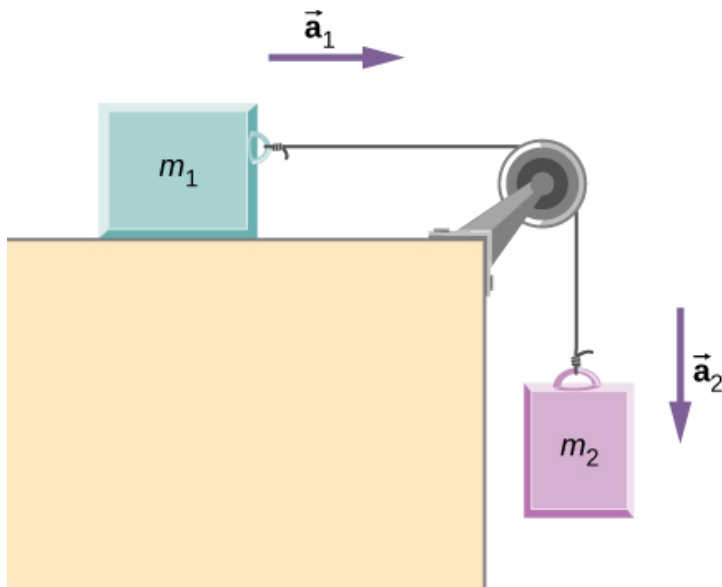
The device shown below is the Atwood's machine considered in [\[link\]](#). Assuming that the masses of the string and the frictionless pulley are negligible, (a) find an equation for the acceleration of the two blocks; (b) find an equation for the tension in the string; and (c) find both the acceleration and tension when block 1 has mass 2.00 kg and block 2 has mass 4.00 kg.



Exercise:

Problem:

Two blocks are connected by a massless rope as shown below. The mass of the block on the table is 4.0 kg and the hanging mass is 1.0 kg. The table and the pulley are frictionless. (a) Find the acceleration of the system. (b) Find the tension in the rope. (c) Find the speed with which the hanging mass hits the floor if it starts from rest and is initially located 1.0 m from the floor.

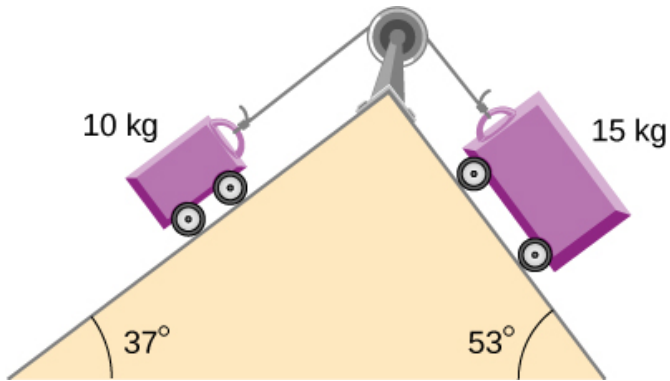


Solution:

a. 2.0 m/s^2 ; b. 7.8 N ; c. 2.0 m/s

Exercise:**Problem:**

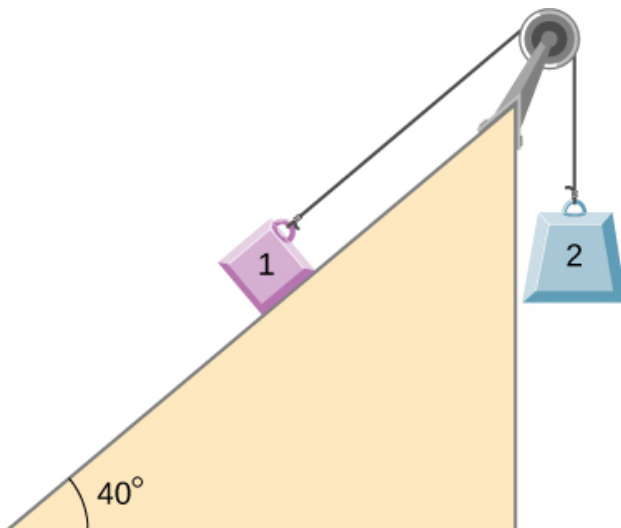
Shown below are two carts connected by a cord that passes over a small frictionless pulley. Each cart rolls freely with negligible friction. Calculate the acceleration of the carts and the tension in the cord.



Exercise:

Problem:

A 2.00 kg block (mass 1) and a 4.00 kg block (mass 2) are connected by a light string as shown; the inclination of the ramp is 40.0° . Friction is negligible. What is (a) the acceleration of each block and (b) the tension in the string?



Solution:

a. 4.43 m/s^2 (mass 1 accelerates up the ramp as mass 2 falls with the same acceleration); b. 21.5 N

Friction

By the end of the section, you will be able to:

- Describe the general characteristics of friction
- List the various types of friction
- Calculate the magnitude of static and kinetic friction, and use these in problems involving Newton's laws of motion

When a body is in motion, it has resistance because the body interacts with its surroundings. This resistance is a force of friction. Friction opposes relative motion between systems in contact but also allows us to move, a concept that becomes obvious if you try to walk on ice. Friction is a common yet complex force, and its behavior still not completely understood. Still, it is possible to understand the circumstances in which it behaves.

Static and Kinetic Friction

The basic definition of **friction** is relatively simple to state.

Note:

Friction

Friction is a force that opposes relative motion between systems in contact.

There are several forms of friction. One of the simpler characteristics of sliding friction is that it is parallel to the contact surfaces between systems and is always in a direction that opposes motion or attempted motion of the systems relative to each other. If two systems are in contact and moving relative to one another, then the friction between them is called kinetic friction. For example, friction slows a hockey puck sliding on ice. When objects are stationary, static friction can act between them; the static friction is usually greater than the kinetic friction between two objects.

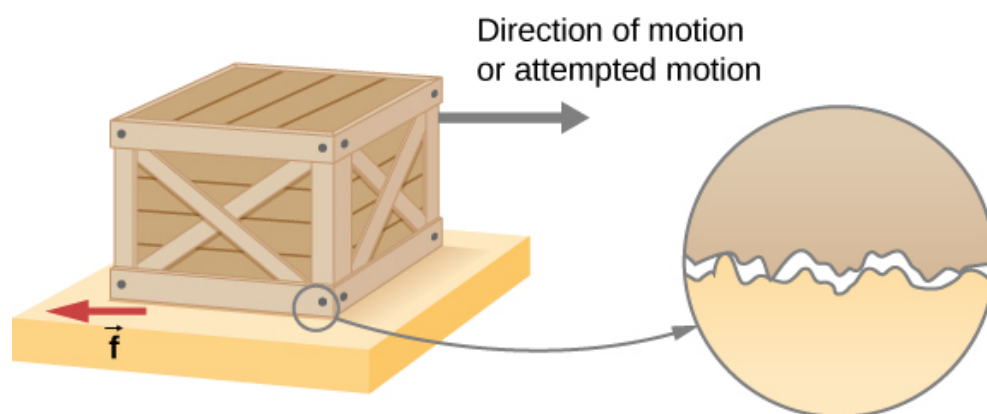
Note:

Static and Kinetic Friction

If two systems are in contact and stationary relative to one another, then the friction between them is called **static friction**. If two systems are in contact and moving relative to one another, then the friction between them is called **kinetic friction**.

Imagine, for example, trying to slide a heavy crate across a concrete floor—you might push very hard on the crate and not move it at all. This means that the static friction responds to what you do—it increases to be equal to and in the opposite direction of your push. If you finally push hard enough, the crate seems to slip suddenly and starts to move. Now static friction gives way to kinetic friction. Once in motion, it is easier to keep it in motion than it was to get it started, indicating that the kinetic frictional force is less than the static frictional force. If you add mass to the crate, say by placing a box on top of it, you need to push even harder to get it started and also to keep it moving. Furthermore, if you oiled the concrete you would find it easier to get the crate started and keep it going (as you might expect).

[\[link\]](#) is a crude pictorial representation of how friction occurs at the interface between two objects. Close-up inspection of these surfaces shows them to be rough. Thus, when you push to get an object moving (in this case, a crate), you must raise the object until it can skip along with just the tips of the surface hitting, breaking off the points, or both. A considerable force can be resisted by friction with no apparent motion. The harder the surfaces are pushed together (such as if another box is placed on the crate), the more force is needed to move them. Part of the friction is due to adhesive forces between the surface molecules of the two objects, which explains the dependence of friction on the nature of the substances. For example, rubber-soled shoes slip less than those with leather soles. Adhesion varies with substances in contact and is a complicated aspect of surface physics. Once an object is moving, there are fewer points of contact (fewer molecules adhering), so less force is required to keep the object moving. At small but nonzero speeds, friction is nearly independent of speed.



Frictional forces, such as \vec{f} , always oppose motion or attempted motion between objects in contact. Friction arises in part because of the roughness of the surfaces in contact, as seen in the expanded view. For the object to move, it must rise to where the peaks of the top surface can skip along the bottom surface. Thus, a

force is required just to set the object in motion. Some of the peaks will be broken off, also requiring a force to maintain motion. Much of the friction is actually due to attractive forces between molecules making up the two objects, so that even perfectly smooth surfaces are not friction-free. (In fact, perfectly smooth, clean surfaces of similar materials would adhere, forming a bond called a “cold weld.”)

The magnitude of the frictional force has two forms: one for static situations (static friction), the other for situations involving motion (kinetic friction). What follows is an approximate empirical (experimentally determined) model only. These equations for static and kinetic friction are not vector equations.

Note:

Magnitude of Static Friction

The magnitude of static friction f_s is

Equation:

$$f_s \leq \mu_s N,$$

where μ_s is the coefficient of static friction and N is the magnitude of the normal force.

The symbol \leq means *less than or equal to*, implying that static friction can have a maximum value of $\mu_s N$. Static friction is a responsive force that increases to be equal and opposite to whatever force is exerted, up to its maximum limit. Once the applied force exceeds

$f_s(\text{max})$, the object moves. Thus,

Equation:

$$f_s(\text{max}) = \mu_s N.$$

Note:

Magnitude of Kinetic Friction

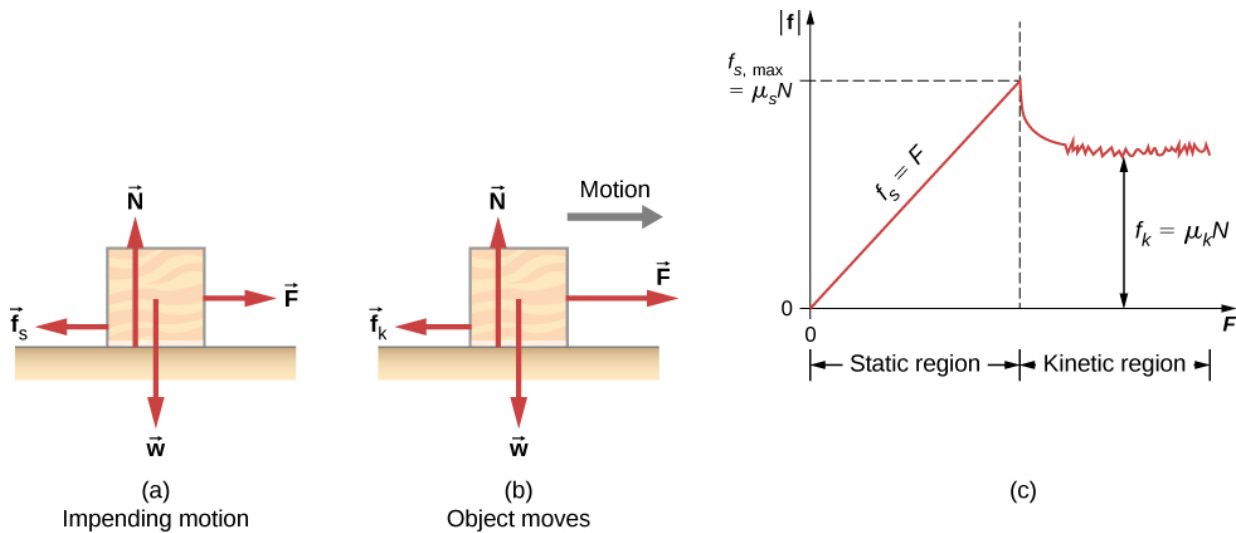
The magnitude of kinetic friction f_k is given by

Equation:

$$f_k = \mu_k N,$$

where μ_k is the coefficient of kinetic friction.

A system in which $f_k = \mu_k N$ is described as a system in which *friction behaves simply*. The transition from static friction to kinetic friction is illustrated in [\[link\]](#).



(a) The force of friction \vec{f} between the block and the rough surface opposes the direction of the applied force \vec{F} . The magnitude of the static friction balances that of the applied force. This is shown in the left side of the graph in (c). (b) At some point, the magnitude of the applied force is greater than the force of kinetic friction, and the block moves to the right. This is shown in the right side of the graph. (c) The graph of the frictional force versus the applied force; note that $f_s(\text{max}) > f_k$.

This means that $\mu_s > \mu_k$.

As you can see in [\[link\]](#), the coefficients of kinetic friction are less than their static counterparts. The approximate values of μ are stated to only one or two digits to indicate the approximate description of friction given by the preceding two equations.

System	Static Friction μ_s	Kinetic Friction μ_k
Rubber on dry concrete	1.0	0.7
Rubber on wet concrete	0.5-0.7	0.3-0.5
Wood on wood	0.5	0.3
Waxed wood on wet snow	0.14	0.1
Metal on wood	0.5	0.3
Steel on steel (dry)	0.6	0.3
Steel on steel (oiled)	0.05	0.03
Teflon on steel	0.04	0.04
Bone lubricated by synovial fluid	0.016	0.015
Shoes on wood	0.9	0.7
Shoes on ice	0.1	0.05
Ice on ice	0.1	0.03
Steel on ice	0.4	0.02

Approximate Coefficients of Static and Kinetic Friction

[\[link\]](#) and [\[link\]](#) include the dependence of friction on materials and the normal force. The direction of friction is always opposite that of motion, parallel to the surface between objects, and perpendicular to the normal force. For example, if the crate you try to push (with a force parallel to the floor) has a mass of 100 kg, then the normal force is equal to its weight,

Equation:

$$w = mg = (100 \text{ kg}) (9.80 \text{ m/s}^2) = 980 \text{ N},$$

perpendicular to the floor. If the coefficient of static friction is 0.45, you would have to exert a force parallel to the floor greater than

Equation:

$$f_s(\text{max}) = \mu_s N = (0.45)(980 \text{ N}) = 440 \text{ N}$$

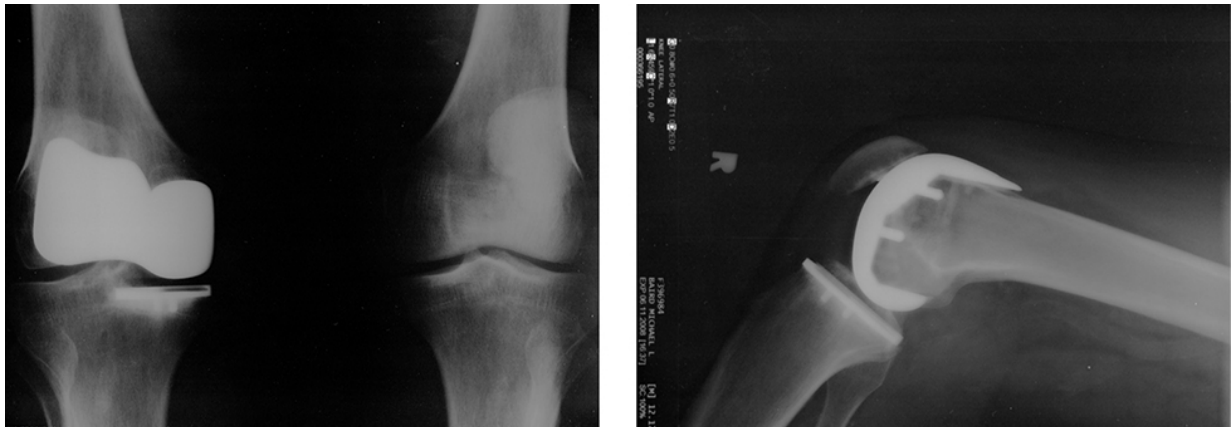
to move the crate. Once there is motion, friction is less and the coefficient of kinetic friction might be 0.30, so that a force of only

Equation:

$$f_k = \mu_k N = (0.30)(980 \text{ N}) = 290 \text{ N}$$

keeps it moving at a constant speed. If the floor is lubricated, both coefficients are considerably less than they would be without lubrication. Coefficient of friction is a unitless quantity with a magnitude usually between 0 and 1.0. The actual value depends on the two surfaces that are in contact.

Many people have experienced the slipperiness of walking on ice. However, many parts of the body, especially the joints, have much smaller coefficients of friction—often three or four times less than ice. A joint is formed by the ends of two bones, which are connected by thick tissues. The knee joint is formed by the lower leg bone (the tibia) and the thighbone (the femur). The hip is a ball (at the end of the femur) and socket (part of the pelvis) joint. The ends of the bones in the joint are covered by cartilage, which provides a smooth, almost-glassy surface. The joints also produce a fluid (synovial fluid) that reduces friction and wear. A damaged or arthritic joint can be replaced by an artificial joint ([link](#)). These replacements can be made of metals (stainless steel or titanium) or plastic (polyethylene), also with very small coefficients of friction.



Artificial knee replacement is a procedure that has been performed for more than 20 years. These post-operative X-rays show a right knee joint replacement. (credit: modification of work by Mike Baird)

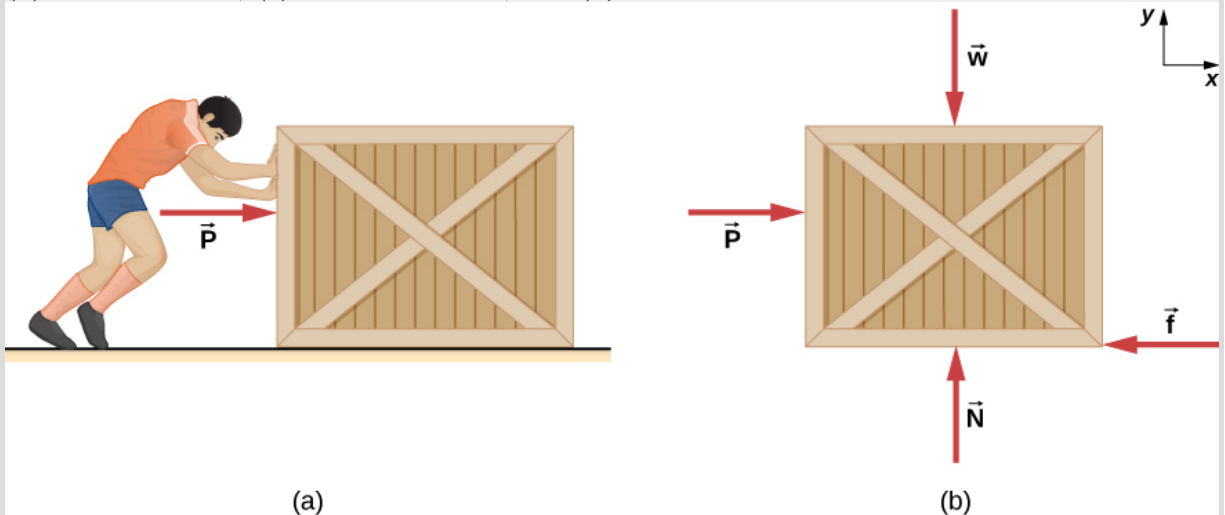
Natural lubricants include saliva produced in our mouths to aid in the swallowing process, and the slippery mucus found between organs in the body, allowing them to move freely past each other during heartbeats, during breathing, and when a person moves. Hospitals and doctor's clinics commonly use artificial lubricants, such as gels, to reduce friction.

The equations given for static and kinetic friction are empirical laws that describe the behavior of the forces of friction. While these formulas are very useful for practical purposes, they do not have the status of mathematical statements that represent general principles (e.g., Newton's second law). In fact, there are cases for which these equations are not even good approximations. For instance, neither formula is accurate for lubricated surfaces or for two surfaces sliding across each other at high speeds. Unless specified, we will not be concerned with these exceptions.

Example:

Static and Kinetic Friction

A 20.0-kg crate is at rest on a floor as shown in [\[link\]](#). The coefficient of static friction between the crate and floor is 0.700 and the coefficient of kinetic friction is 0.600. A horizontal force \vec{P} is applied to the crate. Find the force of friction if (a) $\vec{P} = 20.0 \text{ N}\hat{i}$, (b) $\vec{P} = 30.0 \text{ N}\hat{i}$, (c) $\vec{P} = 120.0 \text{ N}\hat{i}$, and (d) $\vec{P} = 180.0 \text{ N}\hat{i}$.



(a) A crate on a horizontal surface is pushed with a force \vec{P} . (b) The forces on the crate. Here, \vec{f} may represent either the static or the kinetic frictional force.

Strategy

The free-body diagram of the crate is shown in [\[link\]](#)(b). We apply Newton's second law in the horizontal and vertical directions, including the friction force in opposition to the direction of motion of the box.

Solution

Newton's second law GIVES

Equation:

$$\begin{aligned}\sum F_x &= ma_x & \sum F_y &= ma_y \\ P - f &= ma_x & N - w &= 0.\end{aligned}$$

Here we are using the symbol f to represent the frictional force since we have not yet determined whether the crate is subject to static friction or kinetic friction. We do this whenever we are unsure what type of friction is acting. Now the weight of the crate is

Equation:

$$w = (20.0 \text{ kg})(9.80 \text{ m/s}^2) = 196 \text{ N},$$

which is also equal to N . The maximum force of static friction is therefore $(0.700)(196 \text{ N}) = 137 \text{ N}$. As long as \vec{P} is less than 137 N, the force of static friction keeps the crate stationary and $f_s = \vec{P}$. Thus, (a) $f_s = 20.0 \text{ N}$, (b) $f_s = 30.0 \text{ N}$, and (c) $f_s = 120.0 \text{ N}$.

(d) If $\vec{P} = 180.0 \text{ N}$, the applied force is greater than the maximum force of static friction (137 N), so the crate can no longer remain at rest. Once the crate is in motion, kinetic friction acts. Then

Equation:

$$f_k = \mu_k N = (0.600)(196 \text{ N}) = 118 \text{ N},$$

and the acceleration is

Equation:

$$a_x = \frac{P - f_k}{m} = \frac{180.0 \text{ N} - 118 \text{ N}}{20.0 \text{ kg}} = 3.10 \text{ m/s}^2.$$

Significance

This example illustrates how we consider friction in a dynamics problem. Notice that static friction has a value that matches the applied force, until we reach the maximum value of static friction. Also, no motion can occur until the applied force equals the force of static friction, but the force of kinetic friction will then become smaller.

Note:

Exercise:**Problem:**

Check Your Understanding A block of mass 1.0 kg rests on a horizontal surface. The frictional coefficients for the block and surface are $\mu_s = 0.50$ and $\mu_k = 0.40$. (a) What is the minimum horizontal force required to move the block? (b) What is the block's acceleration when this force is applied?

Solution:

a. 4.9 N; b. 0.98 m/s^2

Friction and the Inclined Plane

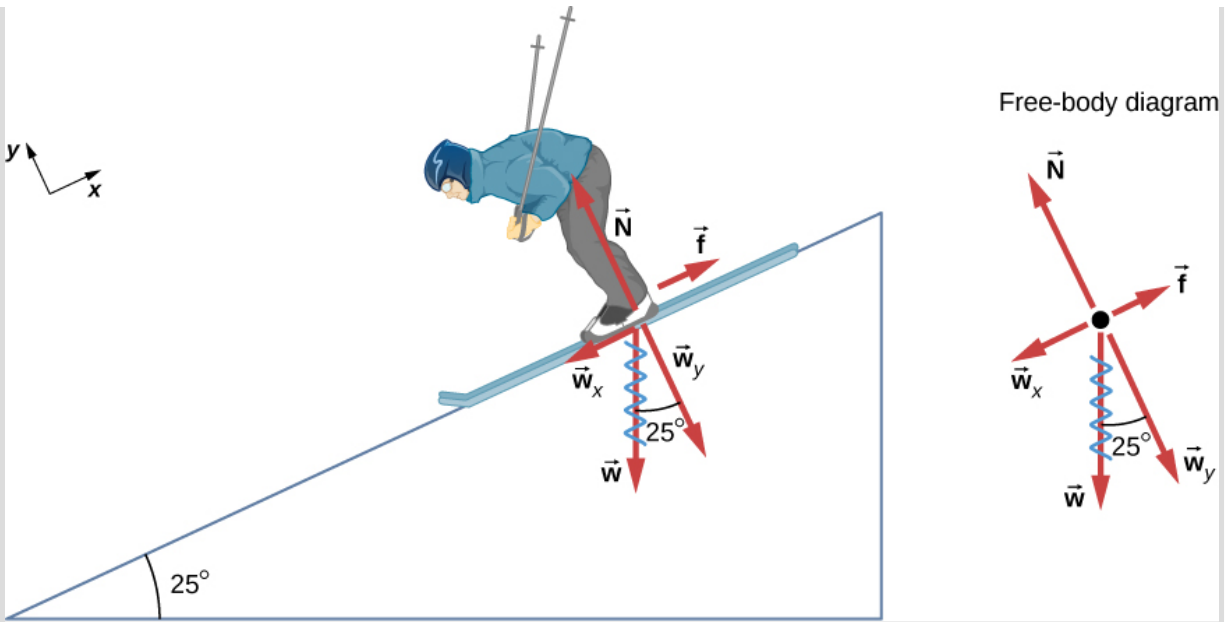
One situation where friction plays an obvious role is that of an object on a slope. It might be a crate being pushed up a ramp to a loading dock or a skateboarder coasting down a mountain, but the basic physics is the same. We usually generalize the sloping surface and call it an inclined plane but then pretend that the surface is flat. Let's look at an example of analyzing motion on an inclined plane with friction.

Example:**Downhill Skier**

A skier with a mass of 62 kg is sliding down a snowy slope at a constant acceleration. Find the coefficient of kinetic friction for the skier if friction is known to be 45.0 N.

Strategy

The magnitude of kinetic friction is given as 45.0 N. Kinetic friction is related to the normal force N by $f_k = \mu_k N$; thus, we can find the coefficient of kinetic friction if we can find the normal force on the skier. The normal force is always perpendicular to the surface, and since there is no motion perpendicular to the surface, the normal force should equal the component of the skier's weight perpendicular to the slope. (See [\[link\]](#), which repeats a figure from the chapter on Newton's laws of motion.)



The motion of the skier and friction are parallel to the slope, so it is most convenient to project all forces onto a coordinate system where one axis is parallel to the slope and the other is perpendicular (axes shown to left of skier). The normal force \vec{N} is perpendicular to the slope, and friction \vec{f} is parallel to the slope, but the skier's weight \vec{w} has components along both axes, namely \vec{w}_y and \vec{w}_x . The normal force \vec{N} is equal in magnitude to \vec{w}_y , so there is no motion perpendicular to the slope.

We have

Equation:

$$N = w_y = w \cos 25^\circ = mg \cos 25^\circ.$$

Substituting this into our expression for kinetic friction, we obtain

Equation:

$$f_k = \mu_k mg \cos 25^\circ,$$

which can now be solved for the coefficient of kinetic friction μ_k .

Solution

Solving for μ_k gives

Equation:

$$\mu_k = \frac{f_k}{N} = \frac{f_k}{w \cos 25^\circ} = \frac{f_k}{mg \cos 25^\circ}.$$

Substituting known values on the right-hand side of the equation,

Equation:

$$\mu_k = \frac{45.0 \text{ N}}{(62 \text{ kg})(9.80 \text{ m/s}^2)(0.906)} = 0.082.$$

Significance

This result is a little smaller than the coefficient listed in [\[link\]](#) for waxed wood on snow, but it is still reasonable since values of the coefficients of friction can vary greatly. In situations like this, where an object of mass m slides down a slope that makes an angle θ with the horizontal, friction is given by $f_k = \mu_k mg \cos \theta$. All objects slide down a slope with constant acceleration under these circumstances.

We have discussed that when an object rests on a horizontal surface, the normal force supporting it is equal in magnitude to its weight. Furthermore, simple friction is always proportional to the normal force. When an object is not on a horizontal surface, as with the inclined plane, we must find the force acting on the object that is directed perpendicular to the surface; it is a component of the weight.

We now derive a useful relationship for calculating coefficient of friction on an inclined plane. Notice that the result applies only for situations in which the object slides at constant speed down the ramp.

An object slides down an inclined plane at a constant velocity if the net force on the object is zero. We can use this fact to measure the coefficient of kinetic friction between two objects. As shown in [\[link\]](#), the kinetic friction on a slope is $f_k = \mu_k mg \cos \theta$. The component of the weight down the slope is equal to $mg \sin \theta$ (see the free-body diagram in [\[link\]](#)). These forces act in opposite directions, so when they have equal magnitude, the acceleration is zero. Writing these out,

Equation:

$$\mu_k mg \cos \theta = mg \sin \theta.$$

Solving for μ_k , we find that

Equation:

$$\mu_k = \frac{mg \sin \theta}{mg \cos \theta} = \tan \theta.$$

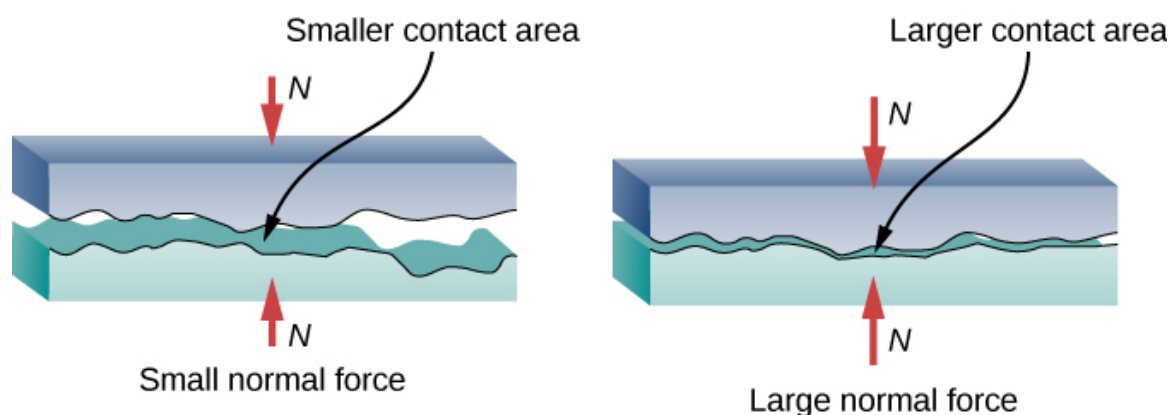
Put a coin on a book and tilt it until the coin slides at a constant velocity down the book. You might need to tap the book lightly to get the coin to move. Measure the angle of tilt

relative to the horizontal and find μ_k . Note that the coin does not start to slide at all until an angle greater than θ is attained, since the coefficient of static friction is larger than the coefficient of kinetic friction. Think about how this may affect the value for μ_k and its uncertainty.

Atomic-Scale Explanations of Friction

The simpler aspects of friction dealt with so far are its macroscopic (large-scale) characteristics. Great strides have been made in the atomic-scale explanation of friction during the past several decades. Researchers are finding that the atomic nature of friction seems to have several fundamental characteristics. These characteristics not only explain some of the simpler aspects of friction—they also hold the potential for the development of nearly friction-free environments that could save hundreds of billions of dollars in energy which is currently being converted (unnecessarily) into heat.

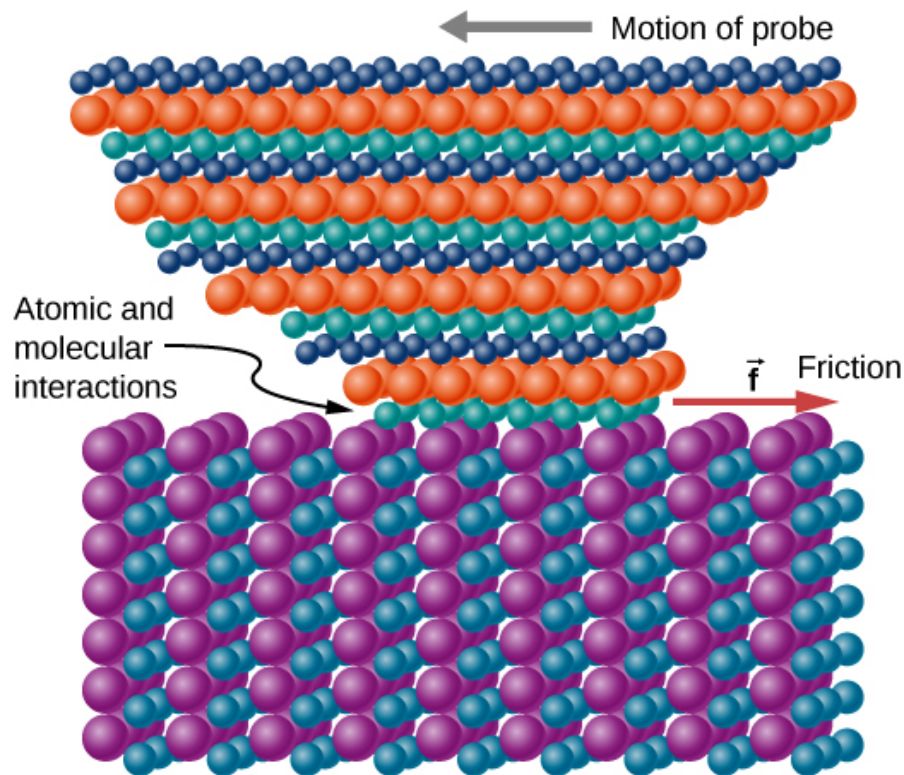
[\[link\]](#) illustrates one macroscopic characteristic of friction that is explained by microscopic (small-scale) research. We have noted that friction is proportional to the normal force, but not to the amount of area in contact, a somewhat counterintuitive notion. When two rough surfaces are in contact, the actual contact area is a tiny fraction of the total area because only high spots touch. When a greater normal force is exerted, the actual contact area increases, and we find that the friction is proportional to this area.



Two rough surfaces in contact have a much smaller area of actual contact than their total area. When the normal force is larger as a result of a larger applied force, the area of actual contact increases, as does friction.

However, the atomic-scale view promises to explain far more than the simpler features of friction. The mechanism for how heat is generated is now being determined. In other

words, why do surfaces get warmer when rubbed? Essentially, atoms are linked with one another to form lattices. When surfaces rub, the surface atoms adhere and cause atomic lattices to vibrate—essentially creating sound waves that penetrate the material. The sound waves diminish with distance, and their energy is converted into heat. Chemical reactions that are related to frictional wear can also occur between atoms and molecules on the surfaces. [\[link\]](#) shows how the tip of a probe drawn across another material is deformed by atomic-scale friction. The force needed to drag the tip can be measured and is found to be related to shear stress, which is discussed in [Static Equilibrium and Elasticity](#). The variation in shear stress is remarkable (more than a factor of 10^{12}) and difficult to predict theoretically, but shear stress is yielding a fundamental understanding of a large-scale phenomenon known since ancient times—friction.



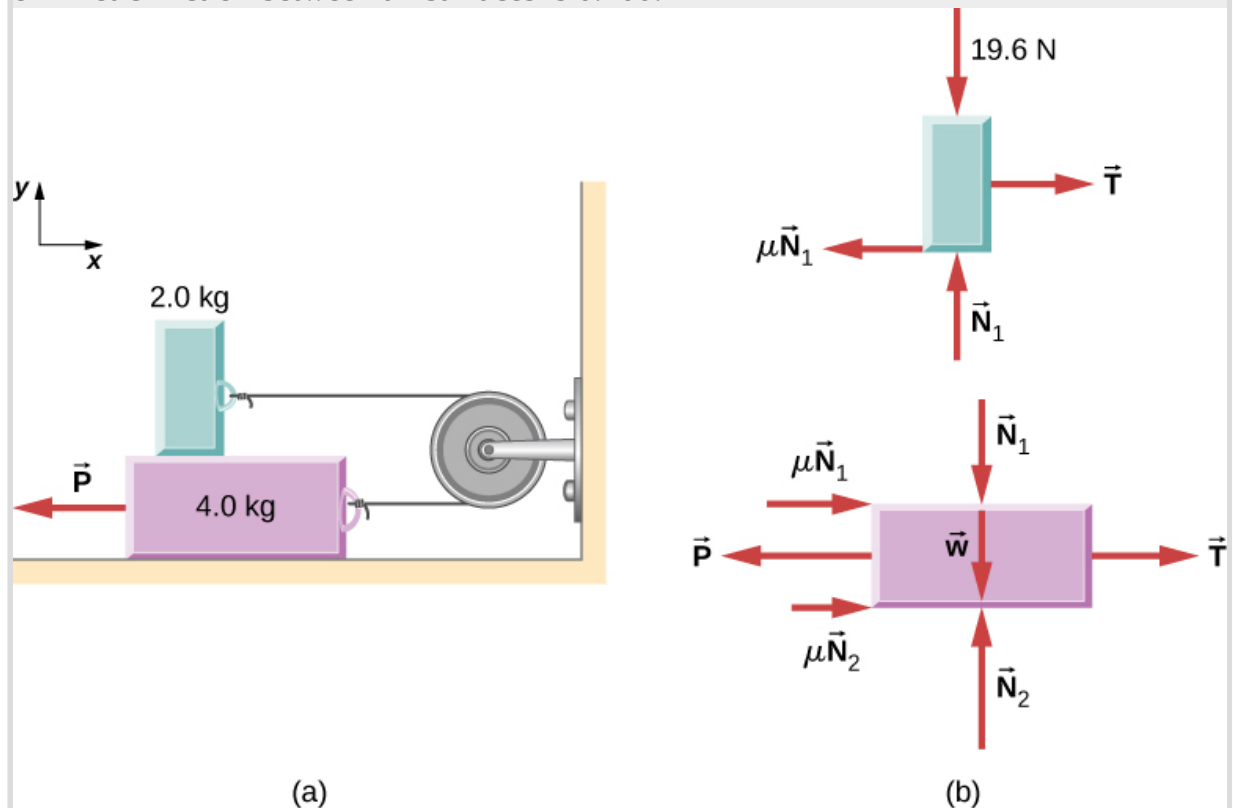
The tip of a probe is deformed sideways by frictional force as the probe is dragged across a surface. Measurements of how the force varies for different materials are yielding fundamental insights into the atomic nature of friction.

Note:

Describe a [model for friction](#) on a molecular level. Describe matter in terms of molecular motion. The description should include diagrams to support the description; how the temperature affects the image; what are the differences and similarities between solid, liquid, and gas particle motion; and how the size and speed of gas molecules relate to everyday objects.

Example:**Sliding Blocks**

The two blocks of [link](#) are attached to each other by a massless string that is wrapped around a frictionless pulley. When the bottom 4.00-kg block is pulled to the left by the constant force \vec{P} , the top 2.00-kg block slides across it to the right. Find the magnitude of the force necessary to move the blocks at constant speed. Assume that the coefficient of kinetic friction between all surfaces is 0.400.



(a) Each block moves at constant velocity. (b) Free-body diagrams for the blocks.

Strategy

We analyze the motions of the two blocks separately. The top block is subjected to a contact force exerted by the bottom block. The components of this force are the normal

force N_1 and the frictional force $-0.400N_1$. Other forces on the top block are the tension T in the string and the weight of the top block itself, 19.6 N. The bottom block is subjected to contact forces due to the top block and due to the floor. The first contact force has components $-N_1$ and $0.400N_1$, which are simply reaction forces to the contact forces that the bottom block exerts on the top block. The components of the contact force of the floor are N_2 and $0.400N_2$. Other forces on this block are $-P$, the tension T , and the weight -39.2 N.

Solution

Since the top block is moving horizontally to the right at constant velocity, its acceleration is zero in both the horizontal and the vertical directions. From Newton's second law,

Equation:

$$\begin{aligned} \sum F_x &= m_1 a_x & \sum F_y &= m_1 a_y \\ T - 0.400N_1 &= 0 & N_1 - 19.6 \text{ N} &= 0. \end{aligned}$$

Solving for the two unknowns, we obtain $N_1 = 19.6$ N and $T = 0.40N_1 = 7.84$ N. The bottom block is also not accelerating, so the application of Newton's second law to this block gives

Equation:

$$\begin{aligned} \sum F_x &= m_2 a_x & \sum F_y &= m_2 a_y \\ T - P + 0.400 N_1 + 0.400 N_2 &= 0 & N_2 - 39.2 \text{ N} - N_1 &= 0. \end{aligned}$$

The values of N_1 and T were found with the first set of equations. When these values are substituted into the second set of equations, we can determine N_2 and P . They are

Equation:

$$N_2 = 58.8 \text{ N} \text{ and } P = 39.2 \text{ N}.$$

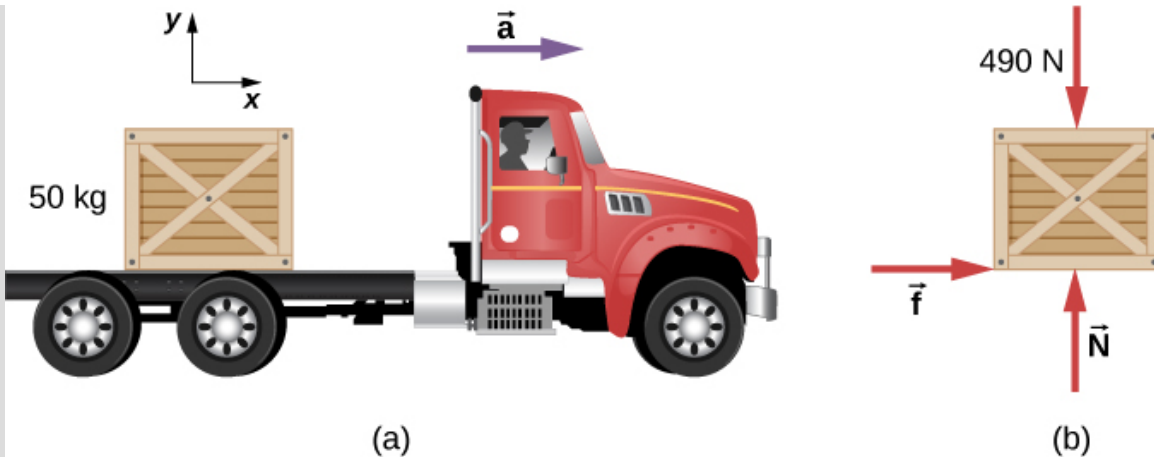
Significance

Understanding what direction in which to draw the friction force is often troublesome. Notice that each friction force labeled in [\[link\]](#) acts in the direction opposite the motion of its corresponding block.

Example:

A Crate on an Accelerating Truck

A 50.0-kg crate rests on the bed of a truck as shown in [\[link\]](#). The coefficients of friction between the surfaces are $\mu_k = 0.300$ and $\mu_s = 0.400$. Find the frictional force on the crate when the truck is accelerating forward relative to the ground at (a) 2.00 m/s^2 , and (b) 5.00 m/s^2 .



(a) A crate rests on the bed of the truck that is accelerating forward. (b) The free-body diagram of the crate.

Strategy

The forces on the crate are its weight and the normal and frictional forces due to contact with the truck bed. We start by *assuming* that the crate is not slipping. In this case, the static frictional force f_s acts on the crate. Furthermore, the accelerations of the crate and the truck are equal.

Solution

- a. Application of Newton's second law to the crate, using the reference frame attached to the ground, yields

Equation:

$$\begin{aligned} \sum F_x &= ma_x & \sum F_y &= ma_y \\ f_s &= (50.0 \text{ kg})(2.00 \text{ m/s}^2) & N - 4.90 \times 10^2 \text{ N} &= (50.0 \text{ kg})(0) \\ &= 1.00 \times 10^2 \text{ N} & N &= 4.90 \times 10^2 \text{ N}. \end{aligned}$$

We can now check the validity of our no-slip assumption. The maximum value of the force of static friction is

Equation:

$$\mu_s N = (0.400)(4.90 \times 10^2 \text{ N}) = 196 \text{ N},$$

whereas the *actual* force of static friction that acts when the truck accelerates forward at 2.00 m/s^2 is only $1.00 \times 10^2 \text{ N}$. Thus, the assumption of no slipping is valid.

- b. If the crate is to move with the truck when it accelerates at 5.0 m/s^2 , the force of static friction must be

Equation:

$$f_s = ma_x = (50.0 \text{ kg})(5.00 \text{ m/s}^2) = 250 \text{ N}.$$

Since this exceeds the maximum of 196 N, the crate must slip. The frictional force is therefore kinetic and is

Equation:

$$f_k = \mu_k N = (0.300)(4.90 \times 10^2 \text{ N}) = 147 \text{ N}.$$

The horizontal acceleration of the crate relative to the ground is now found from

Equation:

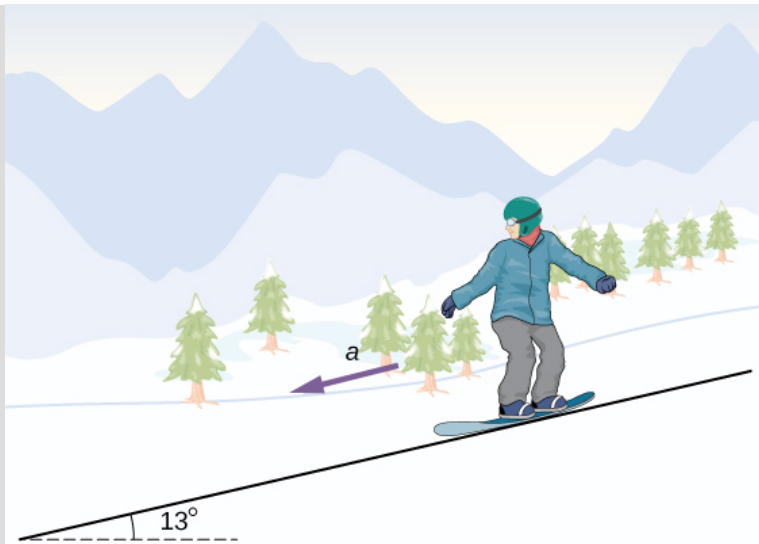
$$\begin{aligned} \sum F_x &= ma_x \\ 147 \text{ N} &= (50.0 \text{ kg})a_x, \\ \text{so } a_x &= 2.94 \text{ m/s}^2. \end{aligned}$$

Significance

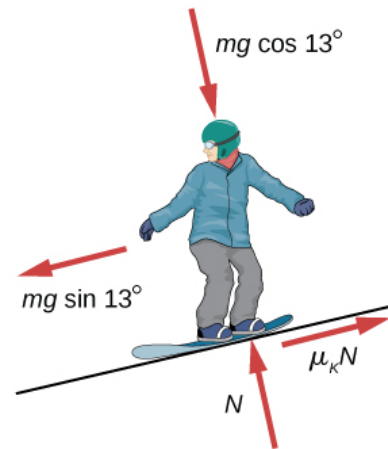
Relative to the ground, the truck is accelerating forward at 5.0 m/s^2 and the crate is accelerating forward at 2.94 m/s^2 . Hence the crate is sliding backward relative to the bed of the truck with an acceleration $2.94 \text{ m/s}^2 - 5.00 \text{ m/s}^2 = -2.06 \text{ m/s}^2$.

Example:**Snowboarding**

Earlier, we analyzed the situation of a downhill skier moving at constant velocity to determine the coefficient of kinetic friction. Now let's do a similar analysis to determine acceleration. The snowboarder of [link](#) glides down a slope that is inclined at $\theta = 13^\circ$ to the horizontal. The coefficient of kinetic friction between the board and the snow is $\mu_k = 0.20$. What is the acceleration of the snowboarder?



(a)



(b)

(a) A snowboarder glides down a slope inclined at 13° to the horizontal. (b) The free-body diagram of the snowboarder.

Strategy

The forces acting on the snowboarder are her weight and the contact force of the slope, which has a component normal to the incline and a component along the incline (force of kinetic friction). Because she moves along the slope, the most convenient reference frame for analyzing her motion is one with the x -axis along and the y -axis perpendicular to the incline. In this frame, both the normal and the frictional forces lie along coordinate axes, the components of the weight are $mg \sin \theta$ along the slope and $mg \cos \theta$ at right angles into the slope, and the only acceleration is along the x -axis ($a_y = 0$).

Solution

We can now apply Newton's second law to the snowboarder:

Equation:

$$\begin{aligned} \sum F_x &= ma_x & \sum F_y &= ma_y \\ mg \sin \theta - \mu_k N &= ma_x & N - mg \cos \theta &= m(0). \end{aligned}$$

From the second equation, $N = mg \cos \theta$. Upon substituting this into the first equation, we find

Equation:

$$\begin{aligned} a_x &= g(\sin \theta - \mu_k \cos \theta) \\ &= g(\sin 13^\circ - 0.20 \cos 13^\circ) = 0.29 \text{ m/s}^2. \end{aligned}$$

Significance

Notice from this equation that if θ is small enough or μ_k is large enough, a_x is negative, that is, the snowboarder slows down.

Note:**Exercise:****Problem:**

Check Your Understanding The snowboarder is now moving down a hill with incline 10.0° . What is the skier's acceleration?

Solution:

-0.23 m/s^2 ; the negative sign indicates that the snowboarder is slowing down.

Summary

- Friction is a contact force that opposes the motion or attempted motion between two systems. Simple friction is proportional to the normal force N supporting the two systems.
- The magnitude of static friction force between two materials stationary relative to each other is determined using the coefficient of static friction, which depends on both materials.
- The kinetic friction force between two materials moving relative to each other is determined using the coefficient of kinetic friction, which also depends on both materials and is always less than the coefficient of static friction.

Conceptual Questions**Exercise:****Problem:**

The glue on a piece of tape can exert forces. Can these forces be a type of simple friction? Explain, considering especially that tape can stick to vertical walls and even to ceilings.

Exercise:

Problem:

When you learn to drive, you discover that you need to let up slightly on the brake pedal as you come to a stop or the car will stop with a jerk. Explain this in terms of the relationship between static and kinetic friction.

Solution:

If you do not let up on the brake pedal, the car's wheels will lock so that they are not rolling; sliding friction is now involved and the sudden change (due to the larger force of static friction) causes the jerk.

Exercise:**Problem:**

When you push a piece of chalk across a chalkboard, it sometimes screeches because it rapidly alternates between slipping and sticking to the board. Describe this process in more detail, in particular, explaining how it is related to the fact that kinetic friction is less than static friction. (The same slip-grab process occurs when tires screech on pavement.)

Exercise:**Problem:**

A physics major is cooking breakfast when she notices that the frictional force between her steel spatula and Teflon frying pan is only 0.200 N. Knowing the coefficient of kinetic friction between the two materials, she quickly calculates the normal force. What is it?

Solution:

5.00 N

Problems**Exercise:****Problem:**

(a) When rebuilding his car's engine, a physics major must exert 3.00×10^2 N of force to insert a dry steel piston into a steel cylinder. What is the normal force between the piston and cylinder? (b) What force would he have to exert if the steel parts were oiled?

Exercise:**Problem:**

(a) What is the maximum frictional force in the knee joint of a person who supports 66.0 kg of her mass on that knee? (b) During strenuous exercise, it is possible to exert forces to the joints that are easily 10 times greater than the weight being supported. What is the maximum force of friction under such conditions? The frictional forces in joints are relatively small in all circumstances except when the joints deteriorate, such as from injury or arthritis. Increased frictional forces can cause further damage and pain.

Solution:

a. 10.0 N; b. 97.0 N

Exercise:**Problem:**

Suppose you have a 120-kg wooden crate resting on a wood floor, with coefficient of static friction 0.500 between these wood surfaces. (a) What maximum force can you exert horizontally on the crate without moving it? (b) If you continue to exert this force once the crate starts to slip, what will its acceleration then be? The coefficient of sliding friction is known to be 0.300 for this situation.

Exercise:**Problem:**

(a) If half of the weight of a small 1.00×10^3 -kg utility truck is supported by its two drive wheels, what is the maximum acceleration it can achieve on dry concrete? (b) Will a metal cabinet lying on the wooden bed of the truck slip if it accelerates at this rate? (c) Solve both problems assuming the truck has four-wheel drive.

Solution:

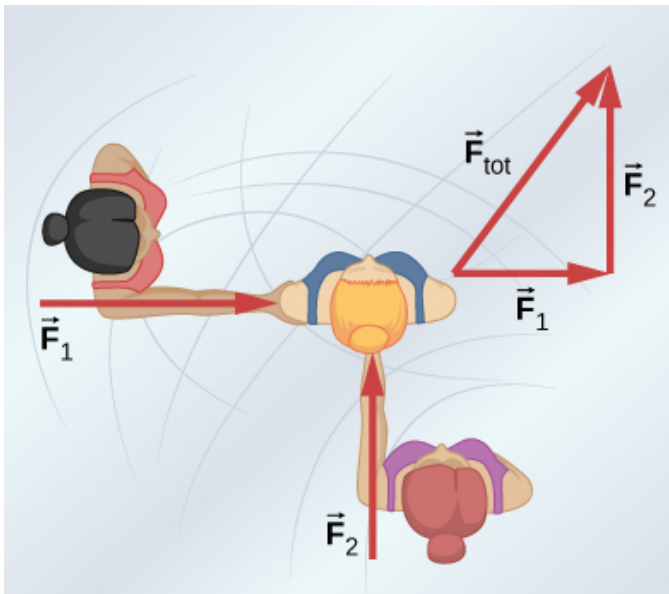
a. 4.9 m/s^2 ; b. The cabinet will not slip. c. The cabinet will slip.

Exercise:**Problem:**

A team of eight dogs pulls a sled with waxed wood runners on wet snow (mush!). The dogs have average masses of 19.0 kg, and the loaded sled with its rider has a mass of 210 kg. (a) Calculate the acceleration of the dogs starting from rest if each dog exerts an average force of 185 N backward on the snow. (b) Calculate the force in the coupling between the dogs and the sled.

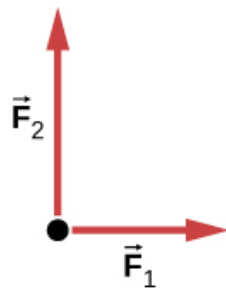
Exercise:**Problem:**

Consider the 65.0-kg ice skater being pushed by two others shown below. (a) Find the direction and magnitude of $\vec{\mathbf{F}}_{\text{tot}}$, the total force exerted on her by the others, given that the magnitudes F_1 and F_2 are 26.4 N and 18.6 N, respectively. (b) What is her initial acceleration if she is initially stationary and wearing steel-bladed skates that point in the direction of $\vec{\mathbf{F}}_{\text{tot}}$? (c) What is her acceleration assuming she is already moving in the direction of $\vec{\mathbf{F}}_{\text{tot}}$? (Remember that friction always acts in the direction opposite that of motion or attempted motion between surfaces in contact.)



(a)

Free-body diagram



(b)

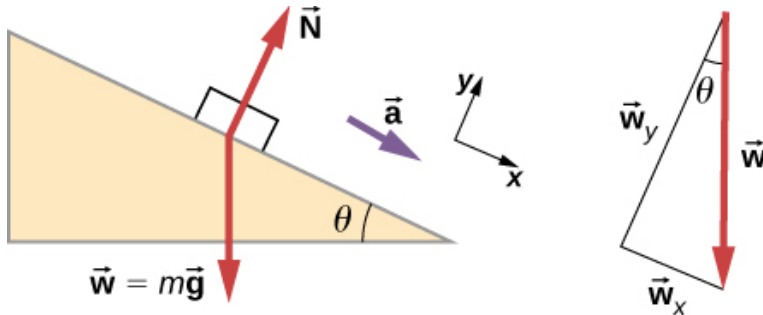
Solution:

a. 32.3 N, 35.2°; b. 0; c. 0.301 m/s² in the direction of \vec{F}_{tot}

Exercise:

Problem:

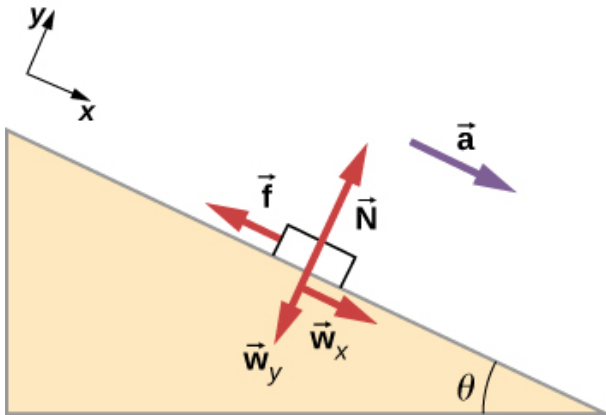
Show that the acceleration of any object down a frictionless incline that makes an angle θ with the horizontal is $a = g \sin \theta$. (Note that this acceleration is independent of mass.)



Exercise:

Problem:

Show that the acceleration of any object down an incline where friction behaves simply (that is, where $f_k = \mu_k N$) is $a = g(\sin \theta - \mu_k \cos \theta)$. Note that the acceleration is independent of mass and reduces to the expression found in the previous problem when friction becomes negligibly small ($\mu_k = 0$).



Solution:

$$\text{net } F_y = 0 \Rightarrow N = mg \cos \theta$$

$$\text{net } F_x = ma$$

$$a = g(\sin \theta - \mu_k \cos \theta)$$

Exercise:

Problem:

Calculate the acceleration opposite to the motion of a snow boarder going up a 5.00° slope, assuming the coefficient of friction for waxed wood on wet snow. The result of the preceding problem may be useful, but be careful to consider the fact that the snow boarder is going uphill.

Exercise:

Problem:

A machine at a post office sends packages out a chute and down a ramp to be loaded into delivery vehicles. (a) Calculate the acceleration of a box heading down a 10.0° slope, assuming the coefficient of friction for a parcel on waxed wood is 0.100. (b) Find the angle of the slope down which this box could move at a constant velocity. You can neglect air resistance in both parts.

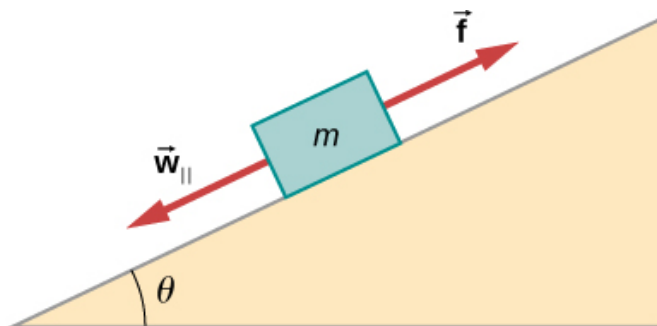
Solution:

a. 0.737 m/s^2 ; b. 5.71°

Exercise:

Problem:

If an object is to rest on an incline without slipping, then friction must equal the component of the weight of the object parallel to the incline. This requires greater and greater friction for steeper slopes. Show that the maximum angle of an incline above the horizontal for which an object will not slide down is $\theta = \tan^{-1} \mu_s$. You may use the result of the previous problem. Assume that $a = 0$ and that static friction has reached its maximum value.



Exercise:**Problem:**

Calculate the maximum acceleration of a car that is heading down a 6.00° slope (one that makes an angle of 6.00° with the horizontal) under the following road conditions. You may assume that the weight of the car is evenly distributed on all four tires and that the coefficient of static friction is involved—that is, the tires are not allowed to slip during the acceleration opposite to the motion. (Ignore rolling.) Calculate for a car: (a) On dry concrete. (b) On wet concrete. (c) On ice, assuming that $\mu_s = 0.100$, the same as for shoes on ice.

Solution:

a. 10.8 m/s^2 ; b. 7.85 m/s^2 ; c. 2.00 m/s^2

Exercise:**Problem:**

Calculate the maximum acceleration of a car that is heading up a 4.00° slope (one that makes an angle of 4.00° with the horizontal) under the following road conditions. Assume that only half the weight of the car is supported by the two drive wheels and that the coefficient of static friction is involved—that is, the tires are not allowed to slip during the acceleration. (Ignore rolling.) (a) On dry concrete. (b) On wet concrete. (c) On ice, assuming that $\mu_s = 0.100$, the same as for shoes on ice.

Exercise:

Problem: Repeat the preceding problem for a car with four-wheel drive.

Solution:

a. 9.09 m/s^2 ; b. 6.16 m/s^2 ; c. 0.294 m/s^2

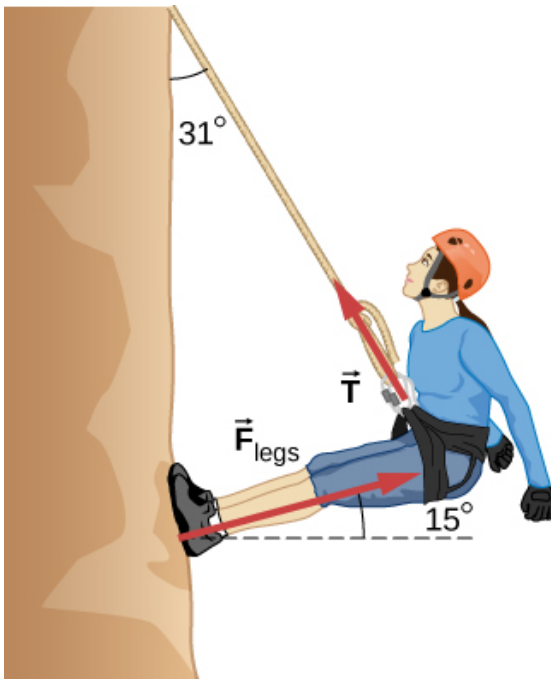
Exercise:

Problem:

A freight train consists of two 8.00×10^5 -kg engines and 45 cars with average masses of 5.50×10^5 kg. (a) What force must each engine exert backward on the track to accelerate the train at a rate of $5.00 \times 10^{-2} \text{ m/s}^2$ if the force of friction is $7.50 \times 10^5 \text{ N}$, assuming the engines exert identical forces? This is not a large frictional force for such a massive system. Rolling friction for trains is small, and consequently, trains are very energy-efficient transportation systems. (b) What is the force in the coupling between the 37th and 38th cars (this is the force each exerts on the other), assuming all cars have the same mass and that friction is evenly distributed among all of the cars and engines?

Exercise:**Problem:**

Consider the 52.0-kg mountain climber shown below. (a) Find the tension in the rope and the force that the mountain climber must exert with her feet on the vertical rock face to remain stationary. Assume that the force is exerted parallel to her legs. Also, assume negligible force exerted by her arms. (b) What is the minimum coefficient of friction between her shoes and the cliff?



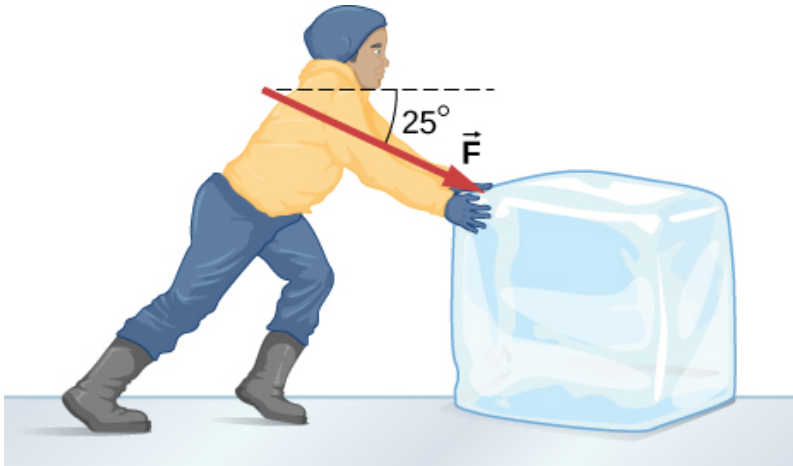
Solution:

a. 272 N, 512 N; b. 0.268

Exercise:

Problem:

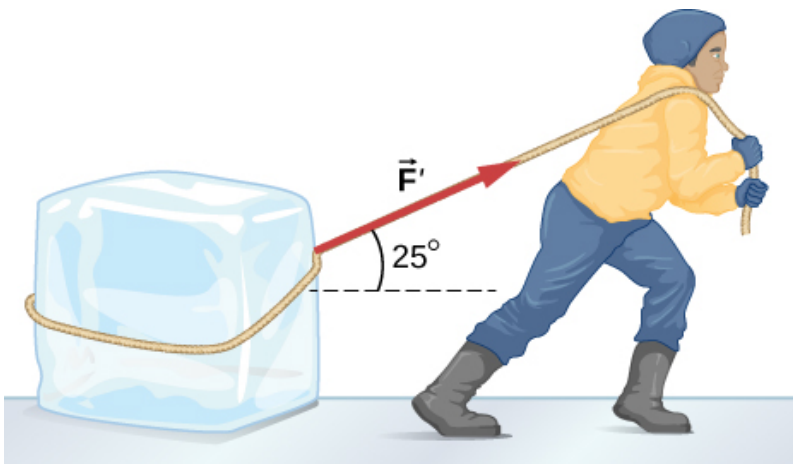
A contestant in a winter sporting event pushes a 45.0-kg block of ice across a frozen lake as shown below. The coefficient of friction of ice can be found in [\[link\]](#). (a) Calculate the minimum force F he must exert to get the block moving. (b) What is its acceleration once it starts to move, if that force is maintained?



Exercise:

Problem:

The contestant now pulls the block of ice with a rope over his shoulder at the same angle above the horizontal as shown below. The coefficient of friction of ice can be found in [\[link\]](#). Calculate the minimum force F he must exert to get the block moving. (b) What is its acceleration once it starts to move, if that force is maintained?



Solution:

a. 46.5 N; b. 0.629 m/s^2

Exercise:**Problem:**

At a post office, a parcel that is a 20.0-kg box slides down a ramp inclined at 30.0° with the horizontal. The coefficient of kinetic friction between the box and plane is 0.0300. (a) Find the acceleration of the box. (b) Find the velocity of the box as it reaches the end of the plane, if the length of the plane is 2 m and the box starts at rest.

Glossary

friction

force that opposes relative motion or attempts at motion between systems in contact

kinetic friction

force that opposes the motion of two systems that are in contact and moving relative to each other

static friction

force that opposes the motion of two systems that are in contact and are not moving relative to each other

Centripetal Force

By the end of the section, you will be able to:

- Explain the equation for centripetal acceleration
- Apply Newton's second law to develop the equation for centripetal force
- Use circular motion concepts in solving problems involving Newton's laws of motion

In [Motion in Two and Three Dimensions](#), we examined the basic concepts of circular motion. An object undergoing circular motion, like one of the race cars shown at the beginning of this chapter, must be accelerating because it is changing the direction of its velocity. We proved that this centrally directed acceleration, called centripetal acceleration, is given by the formula

Equation:

$$a_c = \frac{v^2}{r}$$

where v is the velocity of the object, directed along a tangent line to the curve at any instant. If we know the angular velocity ω , then we can use

Equation:

$$a_c = r\omega^2.$$

Angular velocity gives the rate at which the object is turning through the curve, in units of rad/s. This acceleration acts along the radius of the curved path and is thus also referred to as a radial acceleration.

An acceleration must be produced by a force. Any force or combination of forces can cause a centripetal or radial acceleration. Just a few examples are the tension in the rope on a tether ball, the force of Earth's gravity on the Moon, friction between roller skates and a rink floor, a banked roadway's force on a car, and forces on the tube of a spinning centrifuge. Any net force causing uniform circular motion is called a **centripetal force**. The direction

of a centripetal force is toward the center of curvature, the same as the direction of centripetal acceleration. According to Newton's second law of motion, net force is mass times acceleration: $F_{\text{net}} = ma$. For uniform circular motion, the acceleration is the centripetal acceleration: $a = a_c$. Thus, the magnitude of centripetal force F_c is

Equation:

$$F_c = ma_c.$$

By substituting the expressions for centripetal acceleration a_c ($a_c = \frac{v^2}{r}$; $a_c = r\omega^2$), we get two expressions for the centripetal force F_c in terms of mass, velocity, angular velocity, and radius of curvature:

Note:

Equation:

$$F_c = m \frac{v^2}{r}; \quad F_c = mr\omega^2.$$

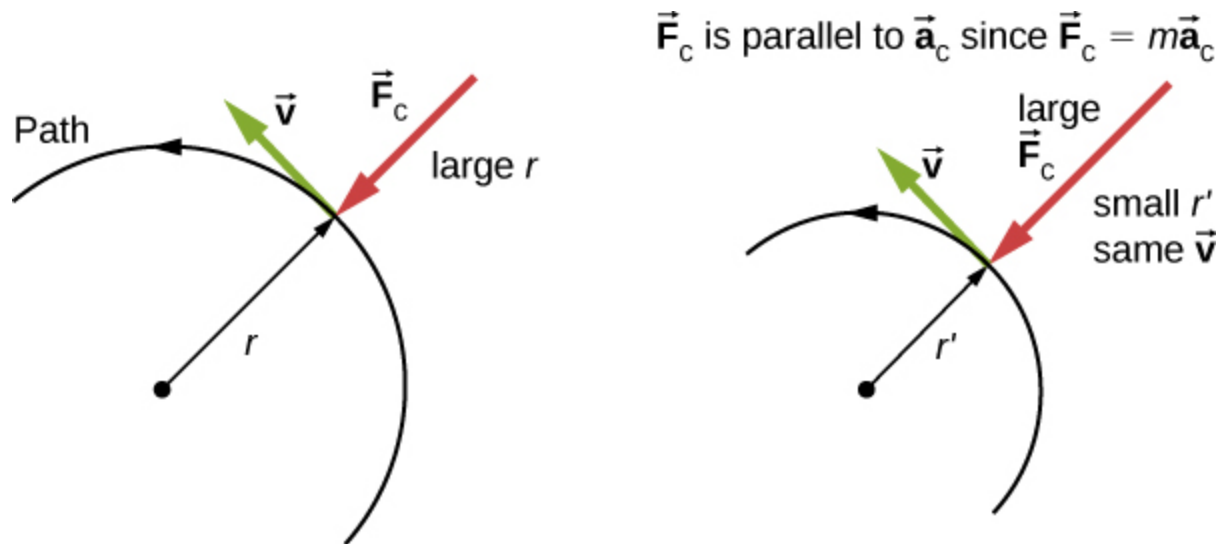
You may use whichever expression for centripetal force is more convenient.

Centripetal force \vec{F}_c is always perpendicular to the path and points to the center of curvature, because \vec{a}_c is perpendicular to the velocity and points to the center of curvature. Note that if you solve the first expression for r , you get

Equation:

$$r = \frac{mv^2}{F_c}.$$

This implies that for a given mass and velocity, a large centripetal force causes a small radius of curvature—that is, a tight curve, as in [\[link\]](#).

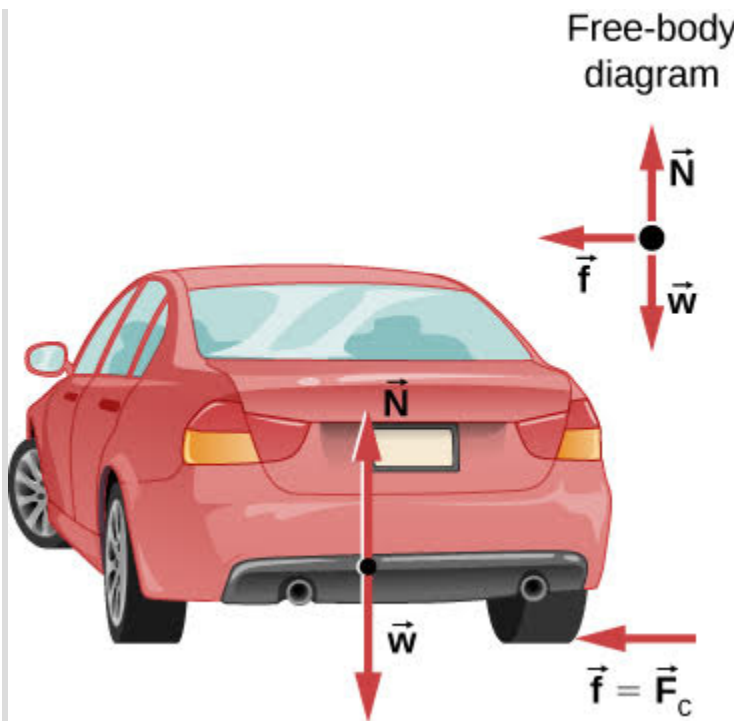


The frictional force supplies the centripetal force and is numerically equal to it. Centripetal force is perpendicular to velocity and causes uniform circular motion. The larger the F_c , the smaller the radius of curvature r and the sharper the curve. The second curve has the same v , but a larger F_c produces a smaller r' .

Example:

What Coefficient of Friction Do Cars Need on a Flat Curve?

(a) Calculate the centripetal force exerted on a 900.0-kg car that negotiates a 500.0-m radius curve at 25.00 m/s. (b) Assuming an unbanked curve, find the minimum static coefficient of friction between the tires and the road, static friction being the reason that keeps the car from slipping ([link](#)).



This car on level ground is moving away and turning to the left. The centripetal force causing the car to turn in a circular path is due to friction between the tires and the road. A minimum coefficient of friction is needed, or the car will move in a larger-radius curve and leave the roadway.

Strategy

- a. We know that $F_c = \frac{mv^2}{r}$. Thus,

Equation:

$$F_c = \frac{mv^2}{r} = \frac{(900.0 \text{ kg})(25.00 \text{ m/s})^2}{(500.0 \text{ m})} = 1125 \text{ N.}$$

- b. [\[link\]](#) shows the forces acting on the car on an unbanked (level ground) curve. Friction is to the left, keeping the car from slipping,

and because it is the only horizontal force acting on the car, the friction is the centripetal force in this case. We know that the maximum static friction (at which the tires roll but do not slip) is $\mu_s N$, where μ_s is the static coefficient of friction and N is the normal force. The normal force equals the car's weight on level ground, so $N = mg$. Thus the centripetal force in this situation is

Equation:

$$F_c \equiv f = \mu_s N = \mu_s mg.$$

Now we have a relationship between centripetal force and the coefficient of friction. Using the equation

Equation:

$$F_c = m \frac{v^2}{r},$$

we obtain

Equation:

$$m \frac{v^2}{r} = \mu_s mg.$$

We solve this for μ_s , noting that mass cancels, and obtain

Equation:

$$\mu_s = \frac{v^2}{rg}.$$

Substituting the knowns,

Equation:

$$\mu_s = \frac{(25.00 \text{ m/s})^2}{(500.0 \text{ m})(9.80 \text{ m/s}^2)} = 0.13.$$

(Because coefficients of friction are approximate, the answer is given to only two digits.)

Significance

The coefficient of friction found in [\[link\]](#)(b) is much smaller than is typically found between tires and roads. The car still negotiates the curve if the coefficient is greater than 0.13, because static friction is a responsive force, able to assume a value less than but no more than $\mu_s N$. A higher coefficient would also allow the car to negotiate the curve at a higher speed, but if the coefficient of friction is less, the safe speed would be less than 25 m/s. Note that mass cancels, implying that, in this example, it does not matter how heavily loaded the car is to negotiate the turn. Mass cancels because friction is assumed proportional to the normal force, which in turn is proportional to mass. If the surface of the road were banked, the normal force would be greater, as discussed next.

Note:

Exercise:

Problem:

Check Your Understanding A car moving at 96.8 km/h travels around a circular curve of radius 182.9 m on a flat country road. What must be the minimum coefficient of static friction to keep the car from slipping?

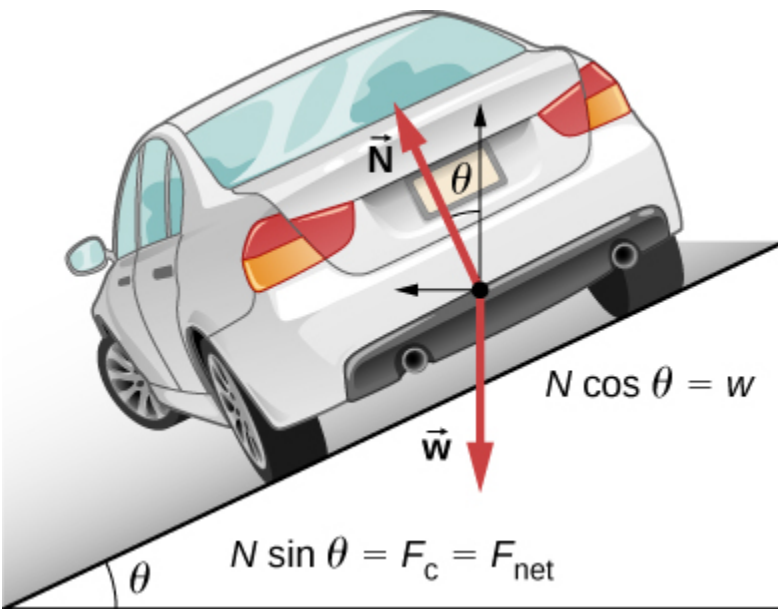
Solution:

0.40

Banked Curves

Let us now consider **banked curves**, where the slope of the road helps you negotiate the curve ([\[link\]](#)). The greater the angle θ , the faster you can take the curve. Race tracks for bikes as well as cars, for example, often have steeply banked curves. In an “ideally banked curve,” the angle θ is such that you can negotiate the curve at a certain speed without the aid of friction

between the tires and the road. We will derive an expression for θ for an ideally banked curve and consider an example related to it.



The car on this banked curve is moving away and turning to the left.

For **ideal banking**, the net external force equals the horizontal centripetal force in the absence of friction. The components of the normal force N in the horizontal and vertical directions must equal the centripetal force and the weight of the car, respectively. In cases in which forces are not parallel, it is most convenient to consider components along perpendicular axes—in this case, the vertical and horizontal directions.

[\[link\]](#) shows a free-body diagram for a car on a frictionless banked curve. If the angle θ is ideal for the speed and radius, then the net external force equals the necessary centripetal force. The only two external forces acting on the car are its weight \vec{w} and the normal force of the road \vec{N} . (A frictionless surface can only exert a force perpendicular to the surface—that is, a normal force.) These two forces must add to give a net external force

that is horizontal toward the center of curvature and has magnitude mv^2/r . Because this is the crucial force and it is horizontal, we use a coordinate system with vertical and horizontal axes. Only the normal force has a horizontal component, so this must equal the centripetal force, that is,

Equation:

$$N \sin \theta = \frac{mv^2}{r}.$$

Because the car does not leave the surface of the road, the net vertical force must be zero, meaning that the vertical components of the two external forces must be equal in magnitude and opposite in direction. From [\[link\]](#), we see that the vertical component of the normal force is $N \cos \theta$, and the only other vertical force is the car's weight. These must be equal in magnitude; thus,

Equation:

$$N \cos \theta = mg.$$

Now we can combine these two equations to eliminate N and get an expression for θ , as desired. Solving the second equation for $N = mg/(\cos \theta)$ and substituting this into the first yields

Equation:

$$\begin{aligned} mg \frac{\sin \theta}{\cos \theta} &= \frac{mv^2}{r} \\ mg \tan \theta &= \frac{mv^2}{r} \\ \tan \theta &= \frac{v^2}{rg}. \end{aligned}$$

Taking the inverse tangent gives

Note:

Equation:

$$\theta = \tan^{-1} \left(\frac{v^2}{rg} \right).$$

This expression can be understood by considering how θ depends on v and r . A large θ is obtained for a large v and a small r . That is, roads must be steeply banked for high speeds and sharp curves. Friction helps, because it allows you to take the curve at greater or lower speed than if the curve were frictionless. Note that θ does not depend on the mass of the vehicle.

Example:**What Is the Ideal Speed to Take a Steeply Banked Tight Curve?**

Curves on some test tracks and race courses, such as Daytona International Speedway in Florida, are very steeply banked. This banking, with the aid of tire friction and very stable car configurations, allows the curves to be taken at very high speed. To illustrate, calculate the speed at which a 100.0-m radius curve banked at 31.0° should be driven if the road were frictionless.

Strategy

We first note that all terms in the expression for the ideal angle of a banked curve except for speed are known; thus, we need only rearrange it so that speed appears on the left-hand side and then substitute known quantities.

Solution

Starting with

Equation:

$$\tan \theta = \frac{v^2}{rg},$$

we get

Equation:

$$v = \sqrt{rg \tan \theta}.$$

Noting that $\tan 31.0^\circ = 0.609$, we obtain

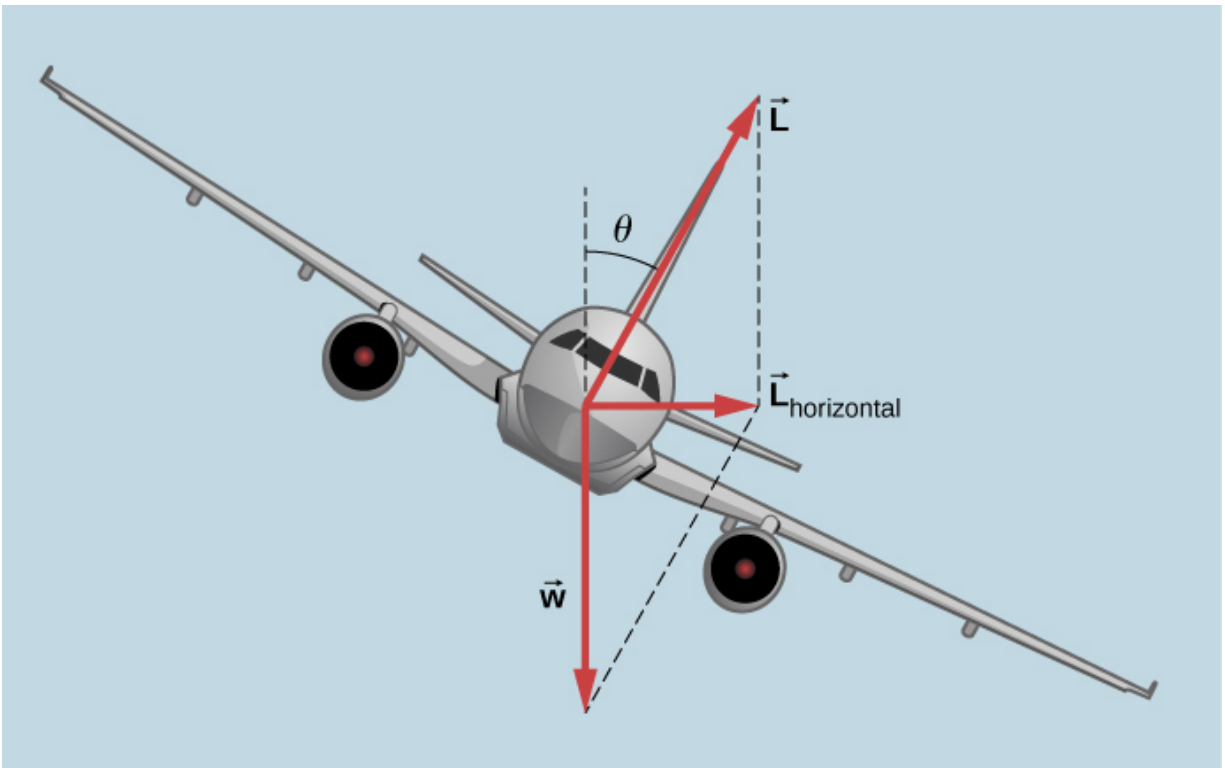
Equation:

$$v = \sqrt{(100.0 \text{ m})(9.80 \text{ m/s}^2)(0.609)} = 24.4 \text{ m/s}.$$

Significance

This is just about 165 km/h, consistent with a very steeply banked and rather sharp curve. Tire friction enables a vehicle to take the curve at significantly higher speeds.

Airplanes also make turns by banking. The lift force, due to the force of the air on the wing, acts at right angles to the wing. When the airplane banks, the pilot is obtaining greater lift than necessary for level flight. The vertical component of lift balances the airplane's weight, and the horizontal component accelerates the plane. The banking angle shown in [\[link\]](#) is given by θ . We analyze the forces in the same way we treat the case of the car rounding a banked curve.



In a banked turn, the horizontal component of lift is unbalanced and accelerates the plane. The normal component of lift balances the plane's weight. The banking angle is given by θ . Compare the vector diagram with that shown in [\[link\]](#).

Note:

Join the [ladybug](#) in an exploration of rotational motion. Rotate the merry-go-round to change its angle or choose a constant angular velocity or angular acceleration. Explore how circular motion relates to the bug's xy-position, velocity, and acceleration using vectors or graphs.

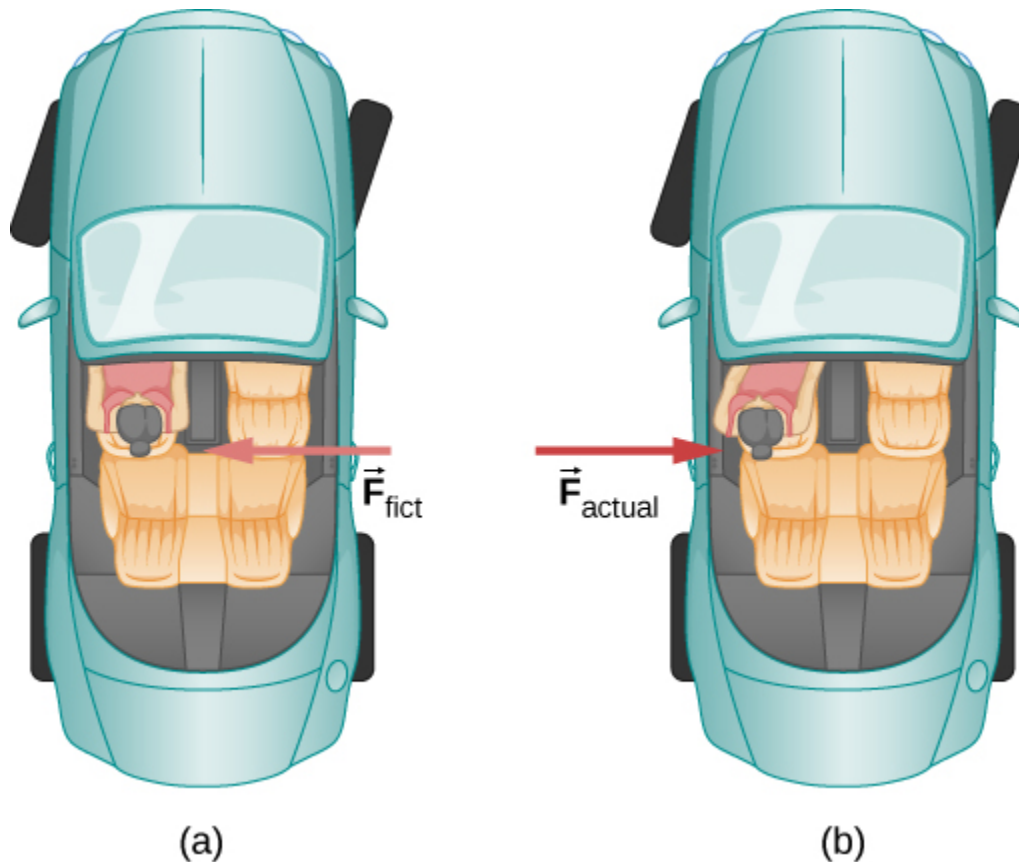
Note:

A circular motion requires a force, the so-called centripetal force, which is directed to the axis of rotation. This simplified [model of a carousel](#)

demonstrates this force.

Inertial Forces and Noninertial (Accelerated) Frames: The Coriolis Force

What do taking off in a jet airplane, turning a corner in a car, riding a merry-go-round, and the circular motion of a tropical cyclone have in common? Each exhibits inertial forces—forces that merely seem to arise from motion, because the observer's frame of reference is accelerating or rotating. When taking off in a jet, most people would agree it feels as if you are being pushed back into the seat as the airplane accelerates down the runway. Yet a physicist would say that *you* tend to remain stationary while the *seat* pushes forward on you. An even more common experience occurs when you make a tight curve in your car—say, to the right ([link](#)). You feel as if you are thrown (that is, *forced*) toward the left relative to the car. Again, a physicist would say that *you* are going in a straight line (recall Newton's first law) but the *car* moves to the right, not that you are experiencing a force from the left.



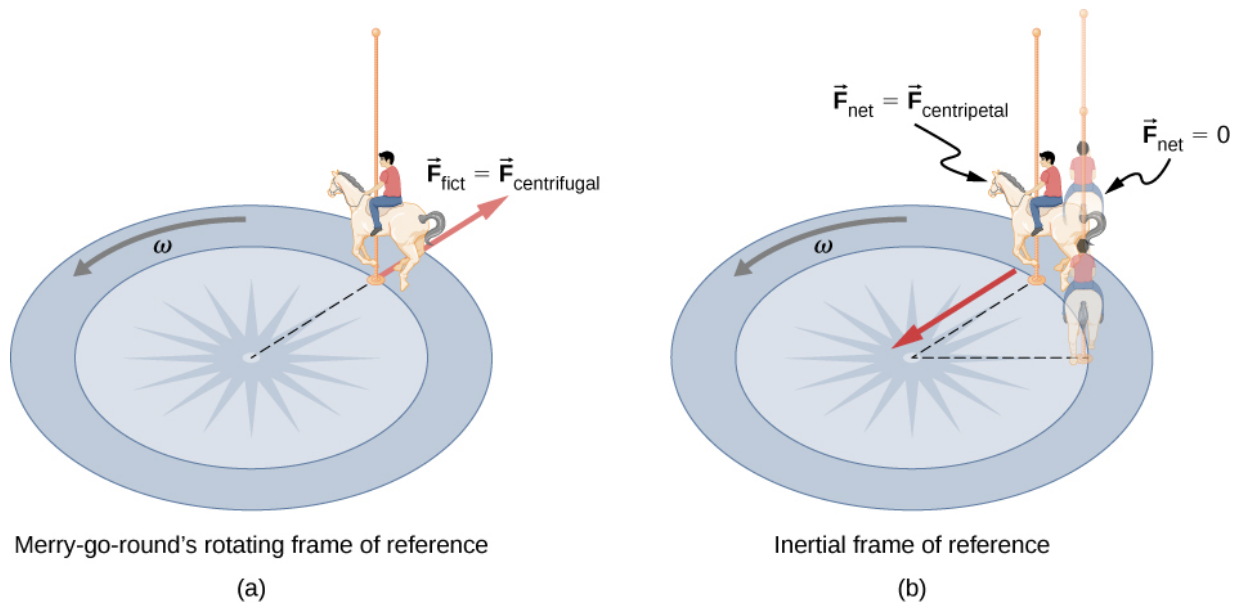
- (a) The car driver feels herself forced to the left relative to the car when she makes a right turn. This is an inertial force arising from the use of the car as a frame of reference. (b) In Earth's frame of reference, the driver moves in a straight line, obeying Newton's first law, and the car moves to the right. There is no force to the left on the driver relative to Earth. Instead, there is a force to the right on the car to make it turn.

We can reconcile these points of view by examining the frames of reference used. Let us concentrate on people in a car. Passengers instinctively use the car as a frame of reference, whereas a physicist might use Earth. The physicist might make this choice because Earth is nearly an inertial frame of reference, in which all forces have an identifiable physical origin. In such a frame of reference, Newton's laws of motion take the form given in

[Newton's Laws of Motion](#). The car is a **noninertial frame of reference** because it is accelerated to the side. The force to the left sensed by car passengers is an **inertial force** having no physical origin (it is due purely to the inertia of the passenger, not to some physical cause such as tension, friction, or gravitation). The car, as well as the driver, is actually accelerating to the right. This inertial force is said to be an inertial force because it does not have a physical origin, such as gravity.

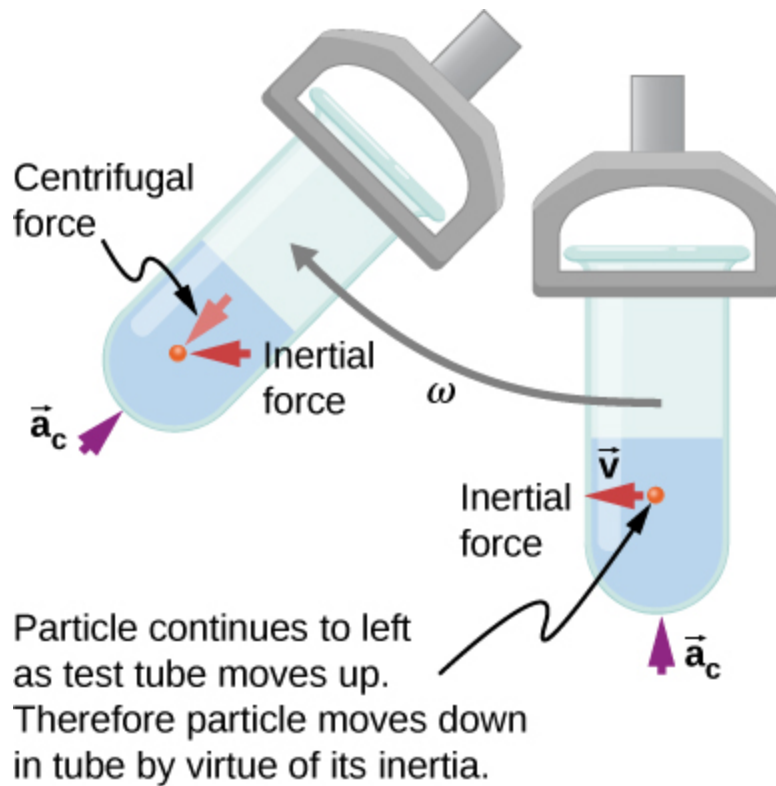
A physicist will choose whatever reference frame is most convenient for the situation being analyzed. There is no problem to a physicist in including inertial forces and Newton's second law, as usual, if that is more convenient, for example, on a merry-go-round or on a rotating planet. Noninertial (accelerated) frames of reference are used when it is useful to do so. Different frames of reference must be considered in discussing the motion of an astronaut in a spacecraft traveling at speeds near the speed of light, as you will appreciate in the study of the special theory of relativity.

Let us now take a mental ride on a merry-go-round—specifically, a rapidly rotating playground merry-go-round ([link](#)). You take the merry-go-round to be your frame of reference because you rotate together. When rotating in that noninertial frame of reference, you feel an inertial force that tends to throw you off; this is often referred to as a *centrifugal force* (not to be confused with centripetal force). Centrifugal force is a commonly used term, but it does not actually exist. You must hang on tightly to counteract your inertia (which people often refer to as centrifugal force). In Earth's frame of reference, there is no force trying to throw you off; we emphasize that centrifugal force is a fiction. You must hang on to make yourself go in a circle because otherwise you would go in a straight line, right off the merry-go-round, in keeping with Newton's first law. But the force you exert acts toward the center of the circle.



- (a) A rider on a merry-go-round feels as if he is being thrown off. This inertial force is sometimes mistakenly called the centrifugal force in an effort to explain the rider's motion in the rotating frame of reference.
- (b) In an inertial frame of reference and according to Newton's laws, it is his inertia that carries him off (the unshaded rider has $F_{\text{net}} = 0$ and heads in a straight line). A force, $F_{\text{centripetal}}$, is needed to cause a circular path.

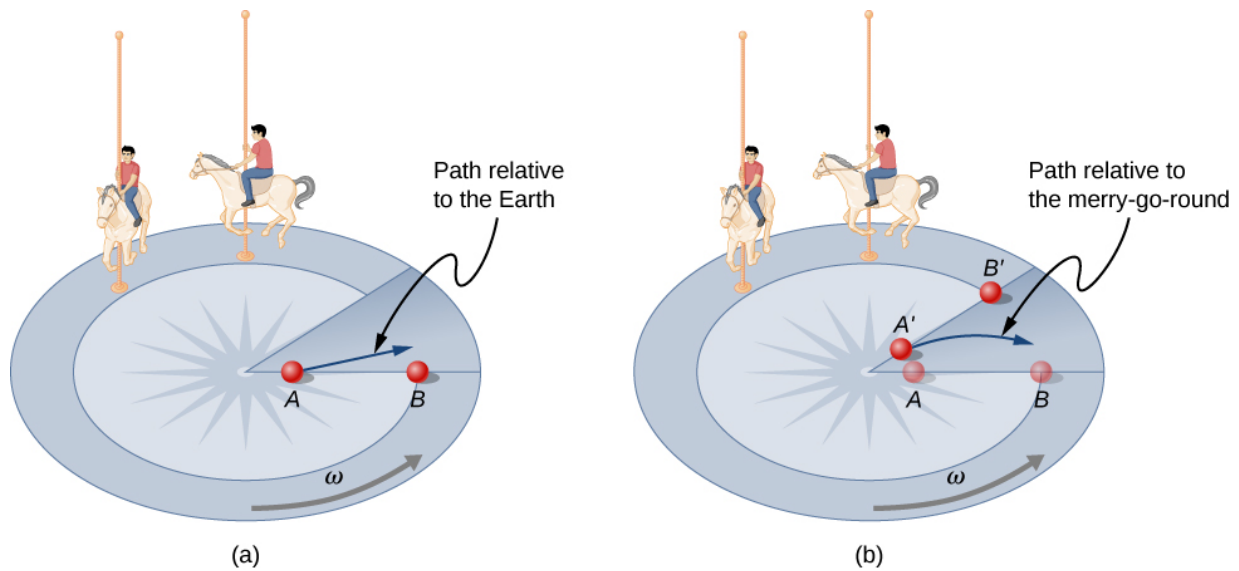
This inertial effect, carrying you away from the center of rotation if there is no centripetal force to cause circular motion, is put to good use in centrifuges ([\[link\]](#)). A centrifuge spins a sample very rapidly, as mentioned earlier in this chapter. Viewed from the rotating frame of reference, the inertial force throws particles outward, hastening their sedimentation. The greater the angular velocity, the greater the centrifugal force. But what really happens is that the inertia of the particles carries them along a line tangent to the circle while the test tube is forced in a circular path by a centripetal force.



Centrifuges use inertia to perform their task. Particles in the fluid sediment settle out because their inertia carries them away from the center of rotation. The large angular velocity of the centrifuge quickens the sedimentation. Ultimately, the particles come into contact with the test tube walls, which then supply the centripetal force needed to make them move in a circle of constant radius.

Let us now consider what happens if something moves in a rotating frame of reference. For example, what if you slide a ball directly away from the center of the merry-go-round, as shown in [\[link\]](#)? The ball follows a straight path relative to Earth (assuming negligible friction) and a path curved to the right on the merry-go-round's surface. A person standing next to the merry-go-round sees the ball moving straight and the merry-go-round rotating

underneath it. In the merry-go-round's frame of reference, we explain the apparent curve to the right by using an inertial force, called the **Coriolis force**, which causes the ball to curve to the right. The Coriolis force can be used by anyone in that frame of reference to explain why objects follow curved paths and allows us to apply Newton's laws in noninertial frames of reference.



Looking down on the counterclockwise rotation of a merry-go-round, we see that a ball slid straight toward the edge follows a path curved to the right. The person slides the ball toward point B , starting at point A .

Both points rotate to the shaded positions (A' and B') shown in the time that the ball follows the curved path in the rotating frame and a straight path in Earth's frame.

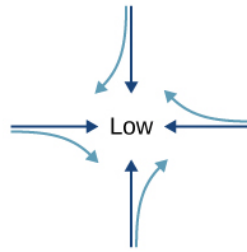
Up until now, we have considered Earth to be an inertial frame of reference with little or no worry about effects due to its rotation. Yet such effects *do* exist—in the rotation of weather systems, for example. Most consequences of Earth's rotation can be qualitatively understood by analogy with the merry-go-round. Viewed from above the North Pole, Earth rotates counterclockwise, as does the merry-go-round in [\[link\]](#). As on the merry-

go-round, any motion in Earth's Northern Hemisphere experiences a Coriolis force to the right. Just the opposite occurs in the Southern Hemisphere; there, the force is to the left. Because Earth's angular velocity is small, the Coriolis force is usually negligible, but for large-scale motions, such as wind patterns, it has substantial effects.

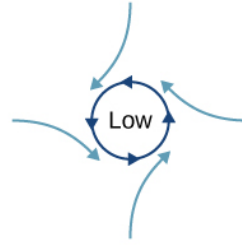
The Coriolis force causes hurricanes in the Northern Hemisphere to rotate in the counterclockwise direction, whereas tropical cyclones in the Southern Hemisphere rotate in the clockwise direction. (The terms hurricane, typhoon, and tropical storm are regionally specific names for cyclones, which are storm systems characterized by low pressure centers, strong winds, and heavy rains.) [\[link\]](#) helps show how these rotations take place. Air flows toward any region of low pressure, and tropical cyclones contain particularly low pressures. Thus winds flow toward the center of a tropical cyclone or a low-pressure weather system at the surface. In the Northern Hemisphere, these inward winds are deflected to the right, as shown in the figure, producing a counterclockwise circulation at the surface for low-pressure zones of any type. Low pressure at the surface is associated with rising air, which also produces cooling and cloud formation, making low-pressure patterns quite visible from space. Conversely, wind circulation around high-pressure zones is clockwise in the Southern Hemisphere but is less visible because high pressure is associated with sinking air, producing clear skies.



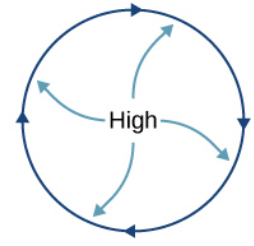
(a)



(b)



(c)



(d)



(e)

(a) The counterclockwise rotation of this Northern Hemisphere hurricane is a major consequence of the Coriolis force. (b) Without the Coriolis force, air would flow straight into a low-pressure zone, such as that found in tropical cyclones. (c) The Coriolis force deflects the winds to the right, producing a counterclockwise rotation. (d) Wind flowing away from a high-pressure zone is also deflected to the right, producing a clockwise rotation. (e) The opposite direction of rotation is produced by the Coriolis force in the Southern Hemisphere, leading to tropical cyclones. (credit a and credit e: modifications of work by NASA)

The rotation of tropical cyclones and the path of a ball on a merry-go-round can just as well be explained by inertia and the rotation of the system underneath. When noninertial frames are used, inertial forces, such as the Coriolis force, must be invented to explain the curved path. There is no identifiable physical source for these inertial forces. In an inertial frame, inertia explains the path, and no force is found to be without an identifiable source. Either view allows us to describe nature, but a view in an inertial

frame is the simplest in the sense that all forces have origins and explanations.

Summary

- Centripetal force \vec{F}_c is a “center-seeking” force that always points toward the center of rotation. It is perpendicular to linear velocity and has the magnitude

Equation:

$$F_c = ma_c.$$

- Rotating and accelerated frames of reference are noninertial. Inertial forces, such as the Coriolis force, are needed to explain motion in such frames.

Conceptual Questions

Exercise:

Problem:

If you wish to reduce the stress (which is related to centripetal force) on high-speed tires, would you use large- or small-diameter tires? Explain.

Exercise:

Problem:

Define centripetal force. Can any type of force (for example, tension, gravitational force, friction, and so on) be a centripetal force? Can any combination of forces be a centripetal force?

Solution:

Centripetal force is defined as any net force causing uniform circular motion. The centripetal force is not a new kind of force. The label “centripetal” refers to *any* force that keeps something turning in a

circle. That force could be tension, gravity, friction, electrical attraction, the normal force, or any other force. Any combination of these could be the source of centripetal force, for example, the centripetal force at the top of the path of a tetherball swung through a vertical circle is the result of both tension and gravity.

Exercise:

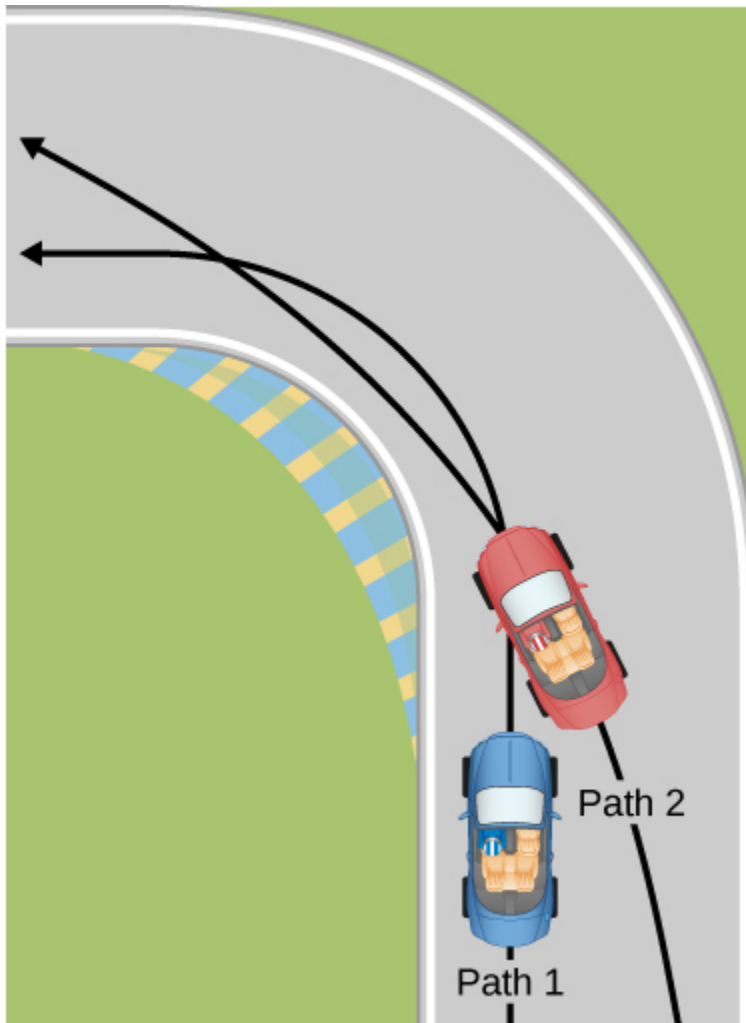
Problem:

If centripetal force is directed toward the center, why do you feel that you are 'thrown' away from the center as a car goes around a curve? Explain.

Exercise:

Problem:

Race car drivers routinely cut corners, as shown below (Path 2). Explain how this allows the curve to be taken at the greatest speed.



Solution:

The driver who cuts the corner (on Path 2) has a more gradual curve, with a larger radius. That one will be the better racing line. If the driver goes too fast around a corner using a racing line, he will still slide off the track; the key is to stay at the maximum value of static friction. So, the driver wants maximum possible speed and maximum friction.

Consider the equation for centripetal force: $F_c = m \frac{v^2}{r}$ where v is speed and r is the radius of curvature. So by decreasing the curvature ($1/r$) of the path that the car takes, we reduce the amount of force the tires have to exert on the road, meaning we can now increase the speed, v . Looking at this from the point of view of the driver on Path 1, we can reason this way: the sharper the turn, the smaller the turning

circle; the smaller the turning circle, the larger is the required centripetal force. If this centripetal force is not exerted, the result is a skid.

Exercise:

Problem:

Many amusement parks have rides that make vertical loops like the one shown below. For safety, the cars are attached to the rails in such a way that they cannot fall off. If the car goes over the top at just the right speed, gravity alone will supply the centripetal force. What other force acts and what is its direction if:

- (a) The car goes over the top at faster than this speed?
- (b) The car goes over the top at slower than this speed?



Exercise:**Problem:**

What causes water to be removed from clothes in a spin-dryer?

Solution:

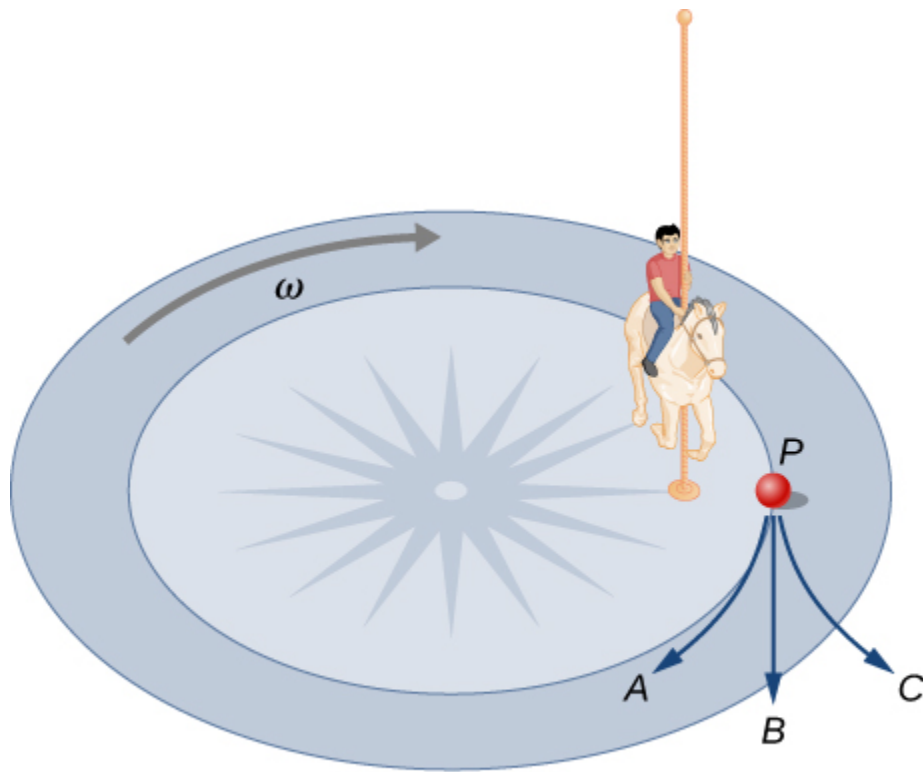
The barrel of the dryer provides a centripetal force on the clothes (including the water droplets) to keep them moving in a circular path. As a water droplet comes to one of the holes in the barrel, it will move in a path tangent to the circle.

Exercise:**Problem:**

As a skater forms a circle, what force is responsible for making his turn? Use a free-body diagram in your answer.

Exercise:**Problem:**

Suppose a child is riding on a merry-go-round at a distance about halfway between its center and edge. She has a lunch box resting on wax paper, so that there is very little friction between it and the merry-go-round. Which path shown below will the lunch box take when she lets go? The lunch box leaves a trail in the dust on the merry-go-round. Is that trail straight, curved to the left, or curved to the right? Explain your answer.



Merry-go-round's rotating
frame of reference

Solution:

If there is no friction, then there is no centripetal force. This means that the lunch box will move along a path tangent to the circle, and thus follows path B. The dust trail will be straight. This is a result of Newton's first law of motion.

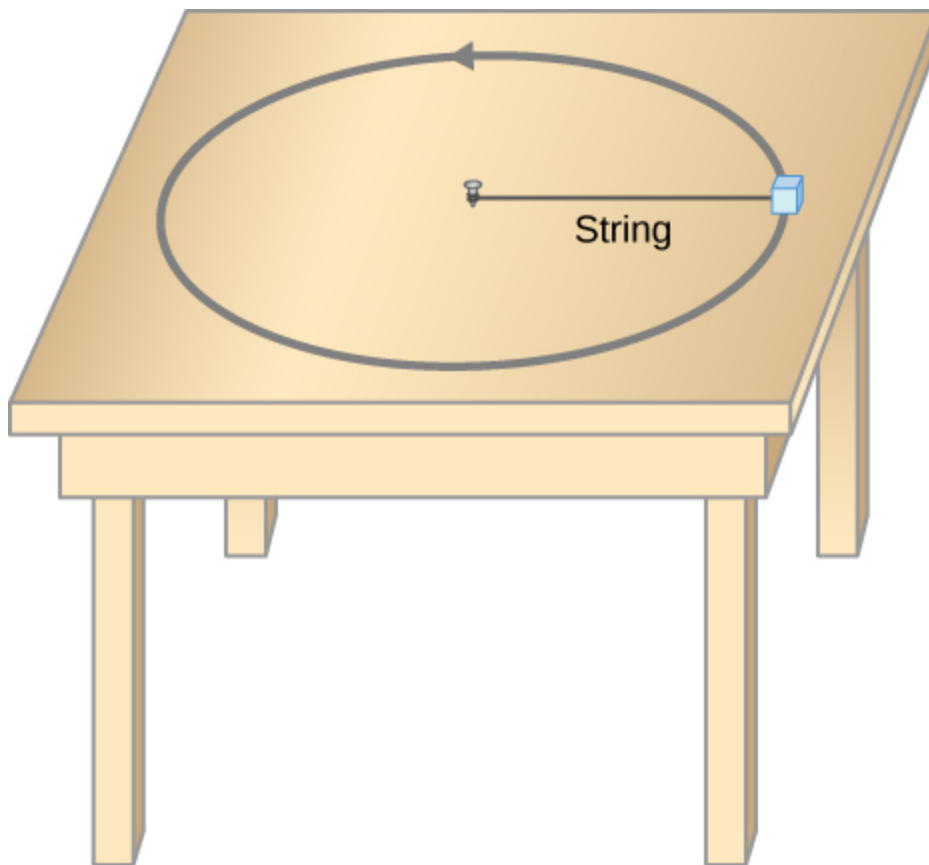
Exercise:**Problem:**

Do you feel yourself thrown to either side when you negotiate a curve that is ideally banked for your car's speed? What is the direction of the force exerted on you by the car seat?

Exercise:

Problem:

Suppose a mass is moving in a circular path on a frictionless table as shown below. In Earth's frame of reference, there is no centrifugal force pulling the mass away from the center of rotation, yet there is a force stretching the string attaching the mass to the nail. Using concepts related to centripetal force and Newton's third law, explain what force stretches the string, identifying its physical origin.



Solution:

There must be a centripetal force to maintain the circular motion; this is provided by the nail at the center. Newton's third law explains the phenomenon. The action force is the force of the string on the mass; the reaction force is the force of the mass on the string. This reaction force causes the string to stretch.

Exercise:**Problem:**

When a toilet is flushed or a sink is drained, the water (and other material) begins to rotate about the drain on the way down. Assuming no initial rotation and a flow initially directly straight toward the drain, explain what causes the rotation and which direction it has in the Northern Hemisphere. (Note that this is a small effect and in most toilets the rotation is caused by directional water jets.) Would the direction of rotation reverse if water were forced up the drain?

Exercise:**Problem:**

A car rounds a curve and encounters a patch of ice with a very low coefficient of kinetic friction. The car slides off the road. Describe the path of the car as it leaves the road.

Solution:

Since the radial friction with the tires supplies the centripetal force, and friction is nearly 0 when the car encounters the ice, the car will obey Newton's first law and go off the road in a straight line path, tangent to the curve. A common misconception is that the car will follow a curved path off the road.

Exercise:**Problem:**

In one amusement park ride, riders enter a large vertical barrel and stand against the wall on its horizontal floor. The barrel is spun up and the floor drops away. Riders feel as if they are pinned to the wall by a force something like the gravitational force. This is an inertial force sensed and used by the riders to explain events in the rotating frame of reference of the barrel. Explain in an inertial frame of reference (Earth is nearly one) what pins the riders to the wall, and identify all forces acting on them.

Exercise:**Problem:**

Two friends are having a conversation. Anna says a satellite in orbit is in free fall because the satellite keeps falling toward Earth. Tom says a satellite in orbit is not in free fall because the acceleration due to gravity is not 9.80 m/s^2 . Who do you agree with and why?

Solution:

Anna is correct. The satellite is freely falling toward Earth due to gravity, even though gravity is weaker at the altitude of the satellite, and g is not 9.80 m/s^2 . Free fall does not depend on the value of g ; that is, you could experience free fall on Mars if you jumped off Olympus Mons (the tallest volcano in the solar system).

Exercise:**Problem:**

A nonrotating frame of reference placed at the center of the Sun is very nearly an inertial one. Why is it not exactly an inertial frame?

Problems**Exercise:****Problem:**

(a) A 22.0-kg child is riding a playground merry-go-round that is rotating at 40.0 rev/min. What centripetal force is exerted if he is 1.25 m from its center? (b) What centripetal force is exerted if the merry-go-round rotates at 3.00 rev/min and he is 8.00 m from its center? (c) Compare each force with his weight.

Solution:

a. 483 N; b. 17.4 N; c. 2.24, 0.0807

Exercise:**Problem:**

Calculate the centripetal force on the end of a 100-m (radius) wind turbine blade that is rotating at 0.5 rev/s. Assume the mass is 4 kg.

Exercise:**Problem:**

What is the ideal banking angle for a gentle turn of 1.20-km radius on a highway with a 105 km/h speed limit (about 65 mi/h), assuming everyone travels at the limit?

Solution:

4.14°

Exercise:**Problem:**

What is the ideal speed to take a 100.0-m-radius curve banked at a 20.0° angle?

Exercise:**Problem:**

(a) What is the radius of a bobsled turn banked at 75.0° and taken at 30.0 m/s, assuming it is ideally banked? (b) Calculate the centripetal acceleration. (c) Does this acceleration seem large to you?

Solution:

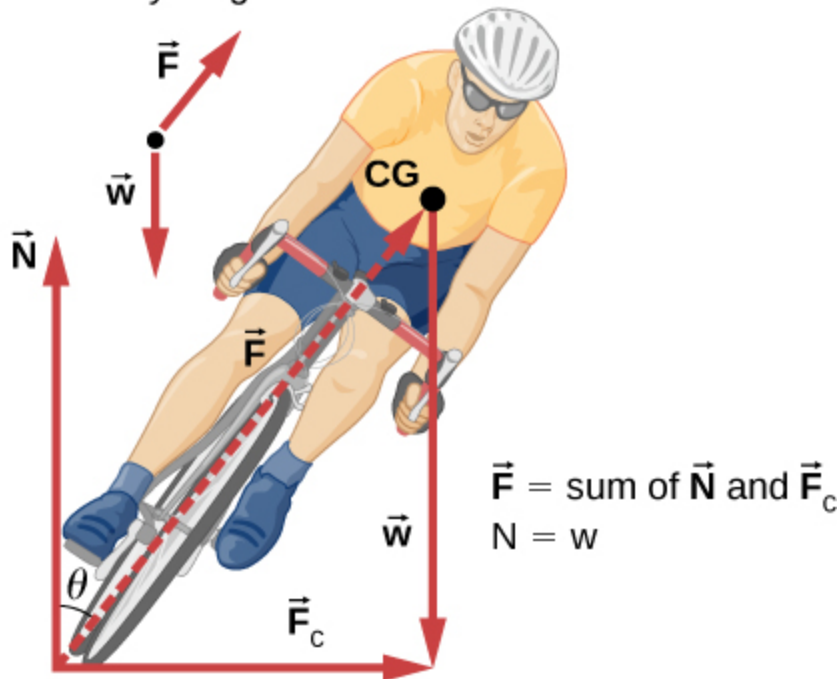
a. 24.6 m; b. 36.6 m/s^2 ; c. 3.73 times g

Exercise:

Problem:

Part of riding a bicycle involves leaning at the correct angle when making a turn, as seen below. To be stable, the force exerted by the ground must be on a line going through the center of gravity. The force on the bicycle wheel can be resolved into two perpendicular components—friction parallel to the road (this must supply the centripetal force) and the vertical normal force (which must equal the system's weight). (a) Show that θ (as defined as shown) is related to the speed v and radius of curvature r of the turn in the same way as for an ideally banked roadway—that is, $\theta = \tan^{-1}(v^2/r g)$. (b) Calculate θ for a 12.0-m/s turn of radius 30.0 m (as in a race).

Free-body diagram



Exercise:

Problem:

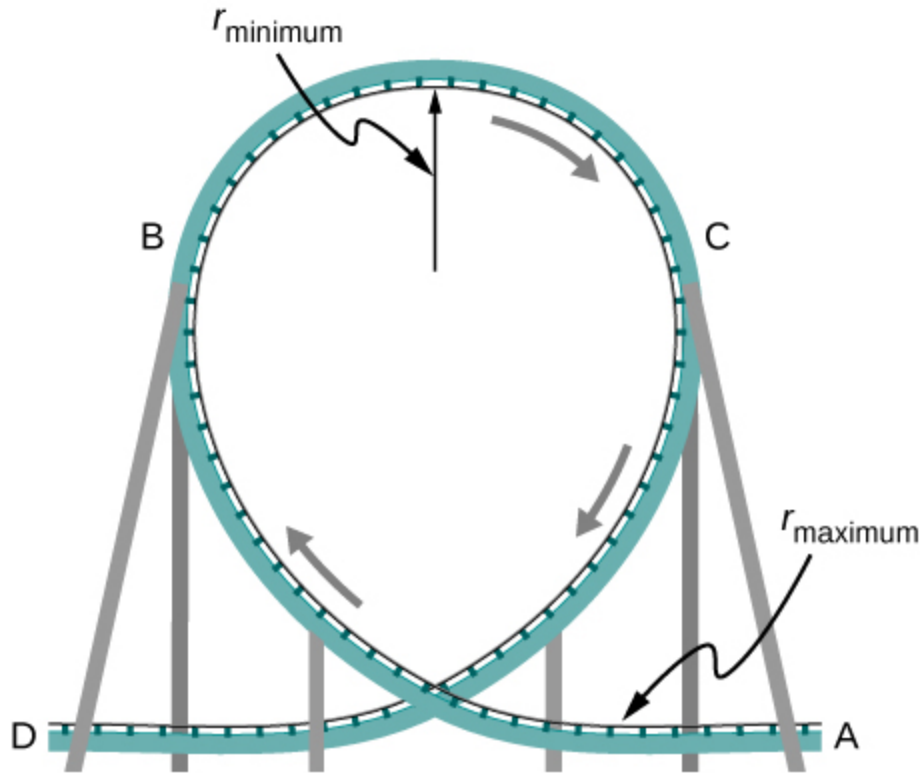
If a car takes a banked curve at less than the ideal speed, friction is needed to keep it from sliding toward the inside of the curve (a problem on icy mountain roads). (a) Calculate the ideal speed to take a 100.0 m radius curve banked at 15.0° . (b) What is the minimum coefficient of friction needed for a frightened driver to take the same curve at 20.0 km/h?

Solution:

a. 16.2 m/s; b. 0.234

Exercise:**Problem:**

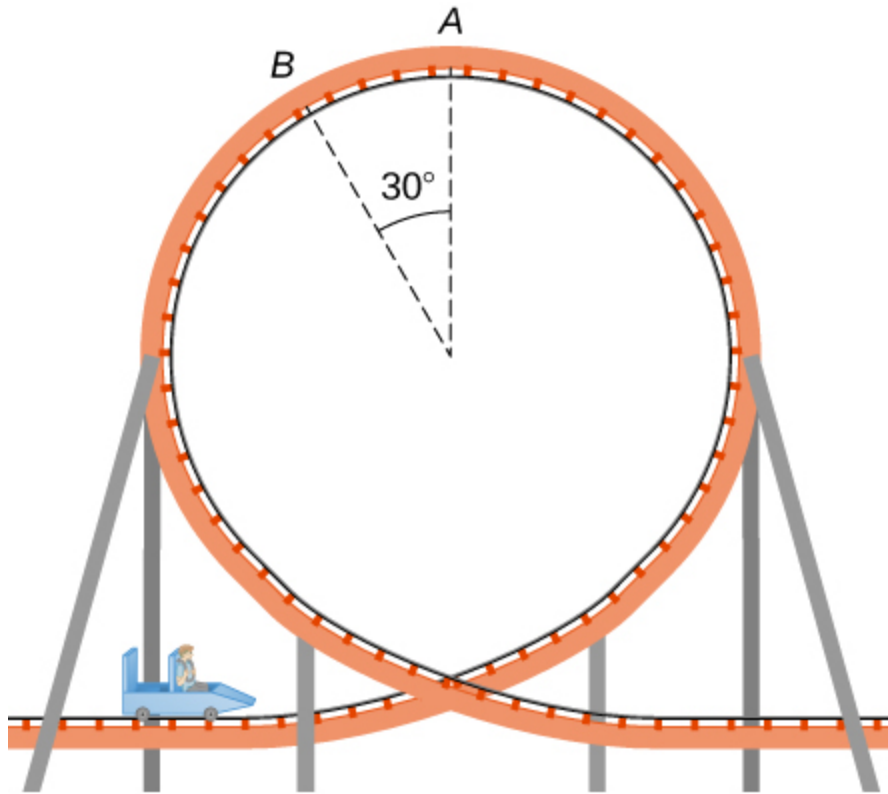
Modern roller coasters have vertical loops like the one shown here. The radius of curvature is smaller at the top than on the sides so that the downward centripetal acceleration at the top will be greater than the acceleration due to gravity, keeping the passengers pressed firmly into their seats. What is the speed of the roller coaster at the top of the loop if the radius of curvature there is 15.0 m and the downward acceleration of the car is $1.50 g$?



Exercise:

Problem:

A child of mass 40.0 kg is in a roller coaster car that travels in a loop of radius 7.00 m. At point A the speed of the car is 10.0 m/s, and at point B, the speed is 10.5 m/s. Assume the child is not holding on and does not wear a seat belt. (a) What is the force of the car seat on the child at point A? (b) What is the force of the car seat on the child at point B? (c) What minimum speed is required to keep the child in his seat at point A?



Solution:

a. 179 N; b. 290 N; c. 8.3 m/s

Exercise:

Problem:

In the simple Bohr model of the ground state of the hydrogen atom, the electron travels in a circular orbit around a fixed proton. The radius of the orbit is 5.28×10^{-11} m, and the speed of the electron is 2.18×10^6 m/s. The mass of an electron is 9.11×10^{-31} kg. What is the force on the electron?

Exercise:

Problem:

Railroad tracks follow a circular curve of radius 500.0 m and are banked at an angle of 5.0° . For trains of what speed are these tracks designed?

Solution:

20.7 m/s

Exercise:**Problem:**

The CERN particle accelerator is circular with a circumference of 7.0 km. (a) What is the acceleration of the protons ($m = 1.67 \times 10^{-27}$ kg) that move around the accelerator at 5% of the speed of light? (The speed of light is $v = 3.00 \times 10^8$ m/s.) (b) What is the force on the protons?

Exercise:**Problem:**

A car rounds an unbanked curve of radius 65 m. If the coefficient of static friction between the road and car is 0.70, what is the maximum speed at which the car can traverse the curve without slipping?

Solution:

21 m/s

Exercise:**Problem:**

A banked highway is designed for traffic moving at 90.0 km/h. The radius of the curve is 310 m. What is the angle of banking of the highway?

Glossary

banked curve

curve in a road that is sloping in a manner that helps a vehicle negotiate the curve

centripetal force

any net force causing uniform circular motion

Coriolis force

inertial force causing the apparent deflection of moving objects when viewed in a rotating frame of reference

ideal banking

sloping of a curve in a road, where the angle of the slope allows the vehicle to negotiate the curve at a certain speed without the aid of friction between the tires and the road; the net external force on the vehicle equals the horizontal centripetal force in the absence of friction

inertial force

force that has no physical origin

noninertial frame of reference

accelerated frame of reference

Drag Force and Terminal Speed

By the end of the section, you will be able to:

- Express the drag force mathematically
- Describe applications of the drag force
- Define terminal velocity
- Determine an object's terminal velocity given its mass

Another interesting force in everyday life is the force of drag on an object when it is moving in a fluid (either a gas or a liquid). You feel the drag force when you move your hand through water. You might also feel it if you move your hand during a strong wind. The faster you move your hand, the harder it is to move. You feel a smaller drag force when you tilt your hand so only the side goes through the air—you have decreased the area of your hand that faces the direction of motion.

Drag Forces

Like friction, the **drag force** always opposes the motion of an object. Unlike simple friction, the drag force is proportional to some function of the velocity of the object in that fluid. This functionality is complicated and depends upon the shape of the object, its size, its velocity, and the fluid it is in. For most large objects such as cyclists, cars, and baseballs not moving too slowly, the magnitude of the drag force F_D is proportional to the square of the speed of the object. We can write this relationship mathematically as $F_D \propto v^2$. When taking into account other factors, this relationship becomes

Note:

Equation:

$$F_D = \frac{1}{2}C\rho Av^2,$$

where C is the drag coefficient, A is the area of the object facing the fluid, and ρ is the density of the fluid. (Recall that density is mass per unit volume.) This equation can also be written in a more generalized fashion as $F_D = bv^n$, where b is a constant equivalent to $0.5C\rho A$. We have set the exponent n for these equations as 2 because when an object is moving at high velocity through air, the magnitude of the drag force is proportional to the square of the speed. As we shall see in [Fluid Mechanics](#), for small particles moving at low speeds in a fluid, the exponent n is equal to 1.

Note:

Drag Force

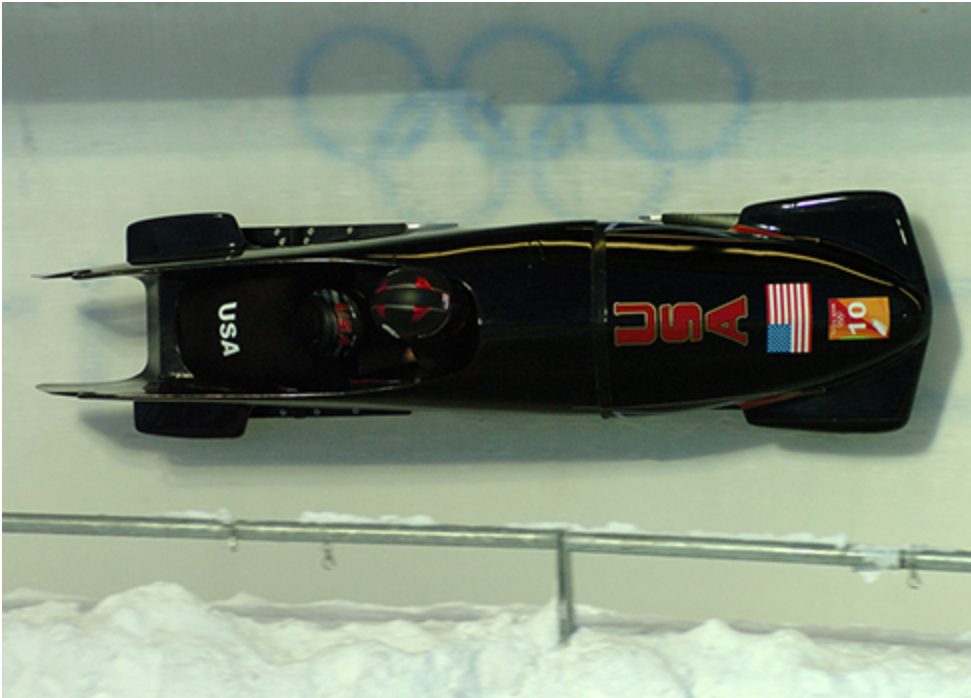
Drag force F_D is proportional to the square of the speed of the object. Mathematically,

Equation:

$$F_D = \frac{1}{2} C \rho A v^2,$$

where C is the drag coefficient, A is the area of the object facing the fluid, and ρ is the density of the fluid.

Athletes as well as car designers seek to reduce the drag force to lower their race times ([\[link\]](#)). Aerodynamic shaping of an automobile can reduce the drag force and thus increase a car's gas mileage.



From racing cars to bobsled racers, aerodynamic shaping is crucial to achieving top speeds. Bobsleds are designed for speed and are shaped like a bullet with tapered fins. (credit: “U.S. Army”/Wikimedia Commons)

The value of the drag coefficient C is determined empirically, usually with the use of a wind tunnel ([link](#)).



NASA researchers test a model plane in a wind tunnel.
(credit: NASA/Ames)

The drag coefficient can depend upon velocity, but we assume that it is a constant here. [\[link\]](#) lists some typical drag coefficients for a variety of objects. Notice that the drag coefficient is a dimensionless quantity. At highway speeds, over 50% of the power of a car is used to overcome air

drag. The most fuel-efficient cruising speed is about 70–80 km/h (about 45–50 mi/h). For this reason, during the 1970s oil crisis in the United States, maximum speeds on highways were set at about 90 km/h (55 mi/h).

Object	C
Airfoil	0.05
Toyota Camry	0.28
Ford Focus	0.32
Honda Civic	0.36
Ferrari Testarossa	0.37
Dodge Ram Pickup	0.43
Sphere	0.45
Hummer H2 SUV	0.64
Skydiver (feet first)	0.70
Bicycle	0.90
Skydiver (horizontal)	1.0
Circular flat plate	1.12

Typical Values of Drag Coefficient C

Substantial research is under way in the sporting world to minimize drag. The dimples on golf balls are being redesigned, as are the clothes that athletes wear. Bicycle racers and some swimmers and runners wear full bodysuits. Australian Cathy Freeman wore a full body suit in the 2000 Sydney Olympics and won a gold medal in the 400-m race. Many swimmers in the 2008 Beijing Olympics wore (Speedo) body suits; it might have made a difference in breaking many world records ([\[link\]](#)). Most elite swimmers (and cyclists) shave their body hair. Such innovations can have the effect of slicing away milliseconds in a race, sometimes making the difference between a gold and a silver medal. One consequence is that careful and precise guidelines must be continuously developed to maintain the integrity of the sport.



Body suits, such as this LZR Racer Suit, have been credited with aiding in many world records after their release in 2008. Smoother “skin” and more

compression forces on a swimmer's body provide at least 10 % less drag. (credit: NASA/Kathy Barnstorff)

Terminal Velocity

Some interesting situations connected to Newton's second law occur when considering the effects of drag forces upon a moving object. For instance, consider a skydiver falling through air under the influence of gravity. The two forces acting on him are the force of gravity and the drag force (ignoring the small buoyant force). The downward force of gravity remains constant regardless of the velocity at which the person is moving. However, as the person's velocity increases, the magnitude of the drag force increases until the magnitude of the drag force is equal to the gravitational force, thus producing a net force of zero. A zero net force means that there is no acceleration, as shown by Newton's second law. At this point, the person's velocity remains constant and we say that the person has reached his **terminal velocity** (v_T). Since F_D is proportional to the speed squared, a heavier skydiver must go faster for F_D to equal his weight. Let's see how this works out more quantitatively.

At the terminal velocity,

Equation:

$$F_{\text{net}} = mg - F_D = ma = 0.$$

Thus,

Equation:

$$mg = F_D.$$

Using the equation for drag force, we have

Equation:

$$mg = \frac{1}{2}C\rho Av_T^2.$$

Solving for the velocity, we obtain

Equation:

$$v_T = \sqrt{\frac{2mg}{\rho CA}}.$$

Assume the density of air is $\rho = 1.21 \text{ kg/m}^3$. A 75-kg skydiver descending head first has a cross-sectional area of approximately $A = 0.18 \text{ m}^2$ and a drag coefficient of approximately $C = 0.70$. We find that

Equation:

$$v_T = \sqrt{\frac{2(75 \text{ kg})(9.80 \text{ m/s}^2)}{(1.21 \text{ kg/m}^3)(0.70)(0.18 \text{ m}^2)}} = 98 \text{ m/s} = 350 \text{ km/h}.$$

This means a skydiver with a mass of 75 kg achieves a terminal velocity of about 350 km/h while traveling in a headfirst position, minimizing the area and his drag. In a spread-eagle position, that terminal velocity may decrease to about 200 km/h as the area increases. This terminal velocity becomes much smaller after the parachute opens.

Example:

Terminal Velocity of a Skydiver

Find the terminal velocity of an 85-kg skydiver falling in a spread-eagle position.

Strategy

At terminal velocity, $F_{\text{net}} = 0$. Thus, the drag force on the skydiver must equal the force of gravity (the person's weight). Using the equation of drag

force, we find $mg = \frac{1}{2}\rho CA v^2$.

Solution

The terminal velocity v_T can be written as

Equation:

$$v_T = \sqrt{\frac{2mg}{\rho CA}} = \sqrt{\frac{2(85 \text{ kg})(9.80 \text{ m/s}^2)}{(1.21 \text{ kg/m}^3)(1.0)(0.70 \text{ m}^2)}} = 44 \text{ m/s}.$$

Significance

This result is consistent with the value for v_T mentioned earlier. The 75-kg skydiver going feet first had a terminal velocity of $v_T = 98 \text{ m/s}$. He weighed less but had a smaller frontal area and so a smaller drag due to the air.

Note:

Exercise:

Problem:

Check Your Understanding Find the terminal velocity of a 50-kg skydiver falling in spread-eagle fashion.

Solution:

34 m/s

The size of the object that is falling through air presents another interesting application of air drag. If you fall from a 5-m-high branch of a tree, you will likely get hurt—possibly fracturing a bone. However, a small squirrel does this all the time, without getting hurt. You do not reach a terminal velocity in such a short distance, but the squirrel does.

The following interesting quote on animal size and terminal velocity is from a 1928 essay by a British biologist, J. B. S. Haldane, titled “On Being the Right Size.”

“To the mouse and any smaller animal, [gravity] presents practically no dangers. You can drop a mouse down a thousand-yard mine shaft; and, on arriving at the bottom, it gets a slight shock and walks away, provided that the ground is fairly soft. A rat is killed, a man is broken, and a horse splashes. For the resistance presented to movement by the air is proportional to the surface of the moving object. Divide an animal’s length, breadth, and height each by ten; its weight is reduced to a thousandth, but its surface only to a hundredth. So the resistance to falling in the case of the small animal is relatively ten times greater than the driving force.”

The above quadratic dependence of air drag upon velocity does not hold if the object is very small, is going very slow, or is in a denser medium than air. Then we find that the drag force is proportional just to the velocity. This relationship is given by Stokes’ law.

Note:

Stokes’ Law

For a spherical object falling in a medium, the drag force is

Equation:

$$F_s = 6\pi r\eta v,$$

where r is the radius of the object, η is the viscosity of the fluid, and v is the object’s velocity.

Good examples of Stokes’ law are provided by microorganisms, pollen, and dust particles. Because each of these objects is so small, we find that many of these objects travel unaided only at a constant (terminal) velocity. Terminal velocities for bacteria (size about $1\ \mu\text{m}$) can be about $2\ \mu\text{m/s}$. To move at a greater speed, many bacteria swim using flagella (organelles

shaped like little tails) that are powered by little motors embedded in the cell.

Sediment in a lake can move at a greater terminal velocity (about $5 \mu\text{m/s}$), so it can take days for it to reach the bottom of the lake after being deposited on the surface.

If we compare animals living on land with those in water, you can see how drag has influenced evolution. Fish, dolphins, and even massive whales are streamlined in shape to reduce drag forces. Birds are streamlined and migratory species that fly large distances often have particular features such as long necks. Flocks of birds fly in the shape of a spearhead as the flock forms a streamlined pattern ([link](#)). In humans, one important example of streamlining is the shape of sperm, which need to be efficient in their use of energy.



Geese fly in a V formation during their long migratory

travels. This shape reduces drag and energy consumption for individual birds, and also allows them a better way to communicate. (credit: modification of work by “Julo”/Wikimedia Commons)

Note:

In lecture demonstrations, we do [measurements of the drag force](#) on different objects. The objects are placed in a uniform airstream created by a fan. Calculate the Reynolds number and the drag coefficient.

The Calculus of Velocity-Dependent Frictional Forces

When a body slides across a surface, the frictional force on it is approximately constant and given by $\mu_k N$. Unfortunately, the frictional force on a body moving through a liquid or a gas does not behave so simply. This drag force is generally a complicated function of the body's velocity. However, for a body moving in a straight line at moderate speeds through a liquid such as water, the frictional force can often be approximated by

Equation:

$$f_R = -bv,$$

where b is a constant whose value depends on the dimensions and shape of the body and the properties of the liquid, and v is the velocity of the body. Two situations for which the frictional force can be represented by this equation are a motorboat moving through water and a small object falling slowly through a liquid.

Let's consider the object falling through a liquid. The free-body diagram of this object with the positive direction downward is shown in [\[link\]](#).

Newton's second law in the vertical direction gives the differential equation
Equation:

$$mg - bv = m \frac{dv}{dt},$$

where we have written the acceleration as dv/dt . As v increases, the frictional force $-bv$ increases until it matches mg . At this point, there is no acceleration and the velocity remains constant at the terminal velocity v_T . From the previous equation,

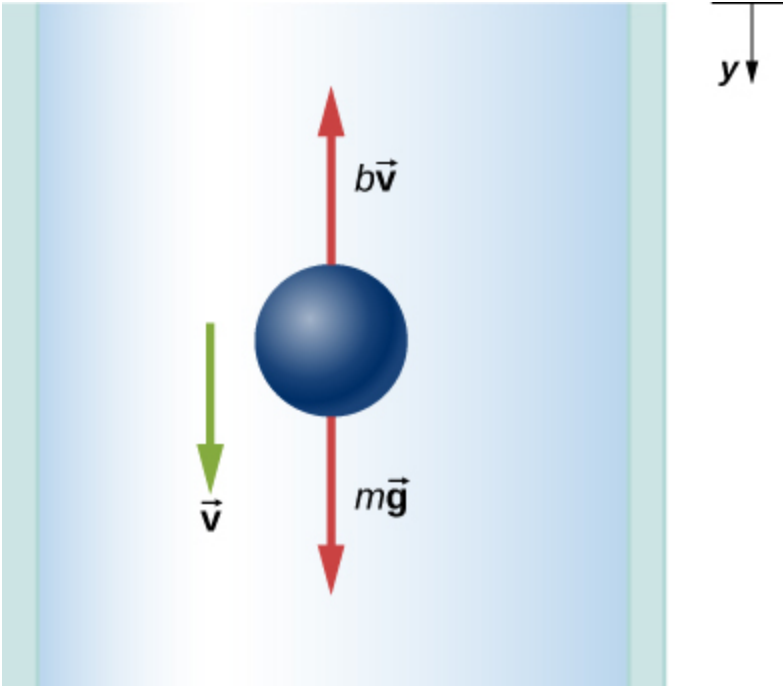
Equation:

$$mg - bv_T = 0,$$

so

Equation:

$$v_T = \frac{mg}{b}.$$



Free-body diagram of an object falling through a resistive medium.

We can find the object's velocity by integrating the differential equation for v . First, we rearrange terms in this equation to obtain

Equation:

$$\frac{dv}{g - (b/m)v} = dt.$$

Assuming that $v = 0$ at $t = 0$, integration of this equation yields

Equation:

$$\int_0^v \frac{dv'}{g - (b/m)v'} = \int_0^t dt',$$

or

Equation:

$$-\frac{m}{b} \ln \left(g - \frac{b}{m} v' \right) \Big|_0^v = t' \Big|_0^t,$$

where v' and t' are dummy variables of integration. With the limits given, we find

Equation:

$$-\frac{m}{b} [\ln \left(g - \frac{b}{m} v \right) - \ln g] = t.$$

Since $\ln A - \ln B = \ln(A/B)$, and $\ln(A/B) = x$ implies $e^x = A/B$, we obtain

Equation:

$$\frac{g - (bv/m)}{g} = e^{-bt/m},$$

and

Equation:

$$v = \frac{mg}{b} (1 - e^{-bt/m}).$$

Notice that as $t \rightarrow \infty$, $v \rightarrow mg/b = v_T$, which is the terminal velocity.

The position at any time may be found by integrating the equation for v . With $v = dy/dt$,

Equation:

$$dy = \frac{mg}{b} (1 - e^{-bt/m}) dt.$$

Assuming $y = 0$ when $t = 0$,

Equation:

$$\int_0^y dy' = \frac{mg}{b} \int_0^t (1 - e^{-bt'/m}) dt',$$

which integrates to

Equation:

$$y = \frac{mg}{b}t + \frac{m^2g}{b^2}(e^{-bt/m} - 1).$$

Example:

Effect of the Resistive Force on a Motorboat

A motorboat is moving across a lake at a speed v_0 when its motor suddenly freezes up and stops. The boat then slows down under the frictional force $f_R = -bv$. (a) What are the velocity and position of the boat as functions of time? (b) If the boat slows down from 4.0 to 1.0 m/s in 10 s, how far does it travel before stopping?

Solution

- a. With the motor stopped, the only horizontal force on the boat is $f_R = -bv$, so from Newton's second law,

Equation:

$$m \frac{dv}{dt} = -bv,$$

which we can write as

Equation:

$$\frac{dv}{v} = -\frac{b}{m}dt.$$

Integrating this equation between the time zero when the velocity is v_0 and the time t when the velocity is v , we have

Equation:

$$\int_0^v \frac{dv'}{v'} = -\frac{b}{m} \int_0^t dt'.$$

Thus,

Equation:

$$\ln \frac{v}{v_0} = -\frac{b}{m} t,$$

which, since $\ln A = x$ implies $e^x = A$, we can write this as

Equation:

$$v = v_0 e^{-bt/m}.$$

Now from the definition of velocity,

Equation:

$$\frac{dx}{dt} = v_0 e^{-bt/m},$$

so we have

Equation:

$$dx = v_0 e^{-bt/m} dt.$$

With the initial position zero, we have

Equation:

$$\int_0^x dx = v_0 \int_0^t e^{-bt'/m} dt',$$

and

Equation:

$$x = -\frac{mv_0}{b} e^{-bt/m} \Big|_0^t = \frac{mv_0}{b} (1 - e^{-bt/m}).$$

As time increases, $e^{-bt/m} \rightarrow 0$, and the position of the boat approaches a limiting value

Equation:

$$x_{\max} = \frac{mv_0}{b}.$$

Although this tells us that the boat takes an infinite amount of time to reach x_{\max} , the boat effectively stops after a reasonable time. For example, at $t = 10m/b$, we have

Equation:

$$v = v_0 e^{-10} \simeq 4.5 \times 10^{-5} v_0,$$

whereas we also have

Equation:

$$x = x_{\max} (1 - e^{-10}) \simeq 0.99995 x_{\max}.$$

Therefore, the boat's velocity and position have essentially reached their final values.

b. With $v_0 = 4.0 \text{ m/s}$ and $v = 1.0 \text{ m/s}$, we have

$$1.0 \text{ m/s} = (4.0 \text{ m/s}) e^{-(b/m)(10 \text{ s})}, \text{ so}$$

Equation:

$$\ln 0.25 = -\ln 4.0 = -\frac{b}{m} (10 \text{ s}),$$

and

Equation:

$$\frac{b}{m} = \frac{1}{10} \ln 4.0 \text{ s}^{-1} = 0.14 \text{ s}^{-1}.$$

Now the boat's limiting position is

Equation:

$$x_{\max} = \frac{mv_0}{b} = \frac{4.0 \text{ m/s}}{0.14 \text{ s}^{-1}} = 29 \text{ m.}$$

Significance

In the both of the previous examples, we found “limiting” values. The terminal velocity is the same as the limiting velocity, which is the velocity of the falling object after a (relatively) long time has passed. Similarly, the limiting distance of the boat is the distance the boat will travel after a long amount of time has passed. Due to the properties of exponential decay, the time involved to reach either of these values is actually not too long (certainly not an infinite amount of time!) but they are quickly found by taking the limit to infinity.

Note:

Exercise:

Problem:

Check Your Understanding Suppose the resistive force of the air on a skydiver can be approximated by $f = -bv^2$. If the terminal velocity of a 100-kg skydiver is 60 m/s, what is the value of b ?

Solution:

0.27 kg/m

Summary

- Drag forces acting on an object moving in a fluid oppose the motion. For larger objects (such as a baseball) moving at a velocity in air, the drag force is determined using the drag coefficient (typical values are given in [\[link\]](#)), the area of the object facing the fluid, and the fluid density.

- For small objects (such as a bacterium) moving in a denser medium (such as water), the drag force is given by Stokes' law.

Key Equations

Magnitude of static friction	$f_s \leq \mu_s N$
Magnitude of kinetic friction	$f_k = \mu_k N$
Centripetal force	$F_c = m \frac{v^2}{r}$ or $F_c = mr\omega^2$
Ideal angle of a banked curve	$\tan \theta = \frac{v^2}{rg}$
Drag force	$F_D = \frac{1}{2} C \rho A v^2$
Stokes' law	$F_s = 6\pi r \eta v$

Conceptual Questions

Exercise:

Problem:

Athletes such as swimmers and bicyclists wear body suits in competition. Formulate a list of pros and cons of such suits.

Solution:

The pros of wearing body suits include: (1) the body suit reduces the drag force on the swimmer and the athlete can move more easily; (2) the tightness of the suit reduces the surface area of the athlete, and

even though this is a small amount, it can make a difference in performance time. The cons of wearing body suits are: (1) The tightness of the suits can induce cramping and breathing problems. (2) Heat will be retained and thus the athlete could overheat during a long period of use.

Exercise:

Problem:

Two expressions were used for the drag force experienced by a moving object in a liquid. One depended upon the speed, while the other was proportional to the square of the speed. In which types of motion would each of these expressions be more applicable than the other one?

Exercise:

Problem:

As cars travel, oil and gasoline leaks onto the road surface. If a light rain falls, what does this do to the control of the car? Does a heavy rain make any difference?

Solution:

The oil is less dense than the water and so rises to the top when a light rain falls and collects on the road. This creates a dangerous situation in which friction is greatly lowered, and so a car can lose control. In a heavy rain, the oil is dispersed and does not affect the motion of cars as much.

Exercise:

Problem:

Why can a squirrel jump from a tree branch to the ground and run away undamaged, while a human could break a bone in such a fall?

Problems

Exercise:**Problem:**

The terminal velocity of a person falling in air depends upon the weight and the area of the person facing the fluid. Find the terminal velocity (in meters per second and kilometers per hour) of an 80.0-kg skydiver falling in a headfirst position with a surface area of 0.140 m^2 .

Solution:

115 m/s or 414 km/h

Exercise:**Problem:**

A 60.0-kg and a 90.0-kg skydiver jump from an airplane at an altitude of $6.00 \times 10^3 \text{ m}$, both falling in a headfirst position. Make some assumption on their frontal areas and calculate their terminal velocities. How long will it take for each skydiver to reach the ground (assuming the time to reach terminal velocity is small)? Assume all values are accurate to three significant digits.

Exercise:**Problem:**

A 560-g squirrel with a surface area of 930 cm^2 falls from a 5.0-m tree to the ground. Estimate its terminal velocity. (Use a drag coefficient for a horizontal skydiver.) What will be the velocity of a 56-kg person hitting the ground, assuming no drag contribution in such a short distance?

Solution:

$$v_T = 11.8 \text{ m/s}; v_2 = 9.9 \text{ m/s}$$

Exercise:

Problem:

To maintain a constant speed, the force provided by a car's engine must equal the drag force plus the force of friction of the road (the rolling resistance). (a) What are the drag forces at 70 km/h and 100 km/h for a Toyota Camry? (Drag area is 0.70 m^2) (b) What is the drag force at 70 km/h and 100 km/h for a Hummer H2? (Drag area is 2.44 m^2) Assume all values are accurate to three significant digits.

Exercise:**Problem:**

By what factor does the drag force on a car increase as it goes from 65 to 110 km/h?

Solution:

$$\left(\frac{110}{65}\right)^2 = 2.86 \text{ times}$$

Exercise:**Problem:**

Calculate the velocity a spherical rain drop would achieve falling from 5.00 km (a) in the absence of air drag (b) with air drag. Take the size across of the drop to be 4 mm, the density to be $1.00 \times 10^3 \text{ kg/m}^3$, and the surface area to be πr^2 .

Exercise:**Problem:**

Using Stokes' law, verify that the units for viscosity are kilograms per meter per second.

Solution:

Stokes' law is $F_s = 6\pi r\eta v$. Solving for the viscosity, $\eta = \frac{F_s}{6\pi r v}$.

Considering only the units, this becomes $[\eta] = \frac{\text{kg}}{\text{m}\cdot\text{s}}$.

Exercise:**Problem:**

Find the terminal velocity of a spherical bacterium (diameter $2.00\text{ }\mu\text{m}$) falling in water. You will first need to note that the drag force is equal to the weight at terminal velocity. Take the density of the bacterium to be $1.10 \times 10^3\text{ kg/m}^3$.

Exercise:**Problem:**

Stokes' law describes sedimentation of particles in liquids and can be used to measure viscosity. Particles in liquids achieve terminal velocity quickly. One can measure the time it takes for a particle to fall a certain distance and then use Stokes' law to calculate the viscosity of the liquid. Suppose a steel ball bearing (density $7.8 \times 10^3\text{ kg/m}^3$, diameter 3.0 mm) is dropped in a container of motor oil. It takes 12 s to fall a distance of 0.60 m . Calculate the viscosity of the oil.

Solution:

$0.76\text{ kg/m} \cdot \text{s}$

Exercise:**Problem:**

Suppose that the resistive force of the air on a skydiver can be approximated by $f = -bv^2$. If the terminal velocity of a 50.0-kg skydiver is 60.0 m/s , what is the value of b ?

Exercise:

Problem:

A small diamond of mass 10.0 g drops from a swimmer's earring and falls through the water, reaching a terminal velocity of 2.0 m/s. (a) Assuming the frictional force on the diamond obeys $f = -bv$, what is b ? (b) How far does the diamond fall before it reaches 90 percent of its terminal speed?

Solution:

a. 0.049 kg/s; b. 0.57 m

Additional Problems**Exercise:****Problem:**

(a) What is the final velocity of a car originally traveling at 50.0 km/h that accelerates opposite to the motion at a rate of 0.400 m/s^2 for 50.0 s? Assume a coefficient of friction of 1.0. (b) What is unreasonable about the result? (c) Which premise is unreasonable, or which premises are inconsistent?

Exercise:**Problem:**

A 75.0-kg woman stands on a bathroom scale in an elevator that accelerates from rest to 30.0 m/s in 2.00 s. (a) Calculate the scale reading in newtons and compare it with her weight. (The scale exerts an upward force on her equal to its reading.) (b) What is unreasonable about the result? (c) Which premise is unreasonable, or which premises are inconsistent?

Solution:

a. 1860 N, 2.53; b. The value (1860 N) is more force than you expect to experience on an elevator. The force of 1860 N is 418 pounds, compared to the force on a typical elevator of 904 N (which is about 203 pounds); this is calculated for a speed from 0 to 10 miles per hour, which is about 4.5 m/s, in 2.00 s). c. The acceleration $a = 1.53 \times g$ is much higher than any standard elevator. The final speed is too large (30.0 m/s is VERY fast)! The time of 2.00 s is not unreasonable for an elevator.

Exercise:

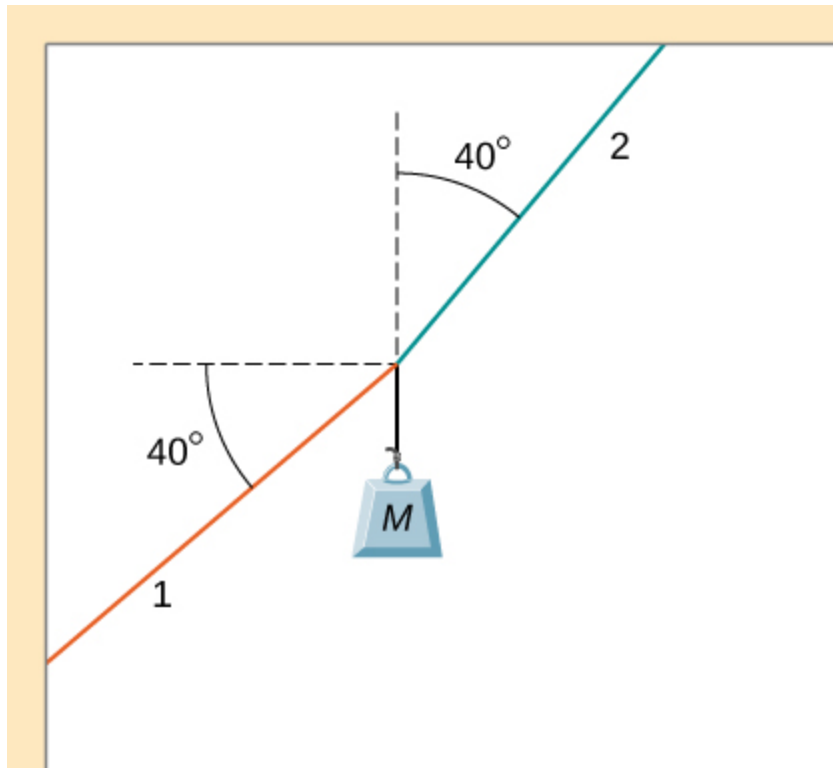
Problem:

(a) Calculate the minimum coefficient of friction needed for a car to negotiate an unbanked 50.0 m radius curve at 30.0 m/s. (b) What is unreasonable about the result? (c) Which premises are unreasonable or inconsistent?

Exercise:

Problem:

As shown below, if $M = 5.50$ kg, what is the tension in string 1?



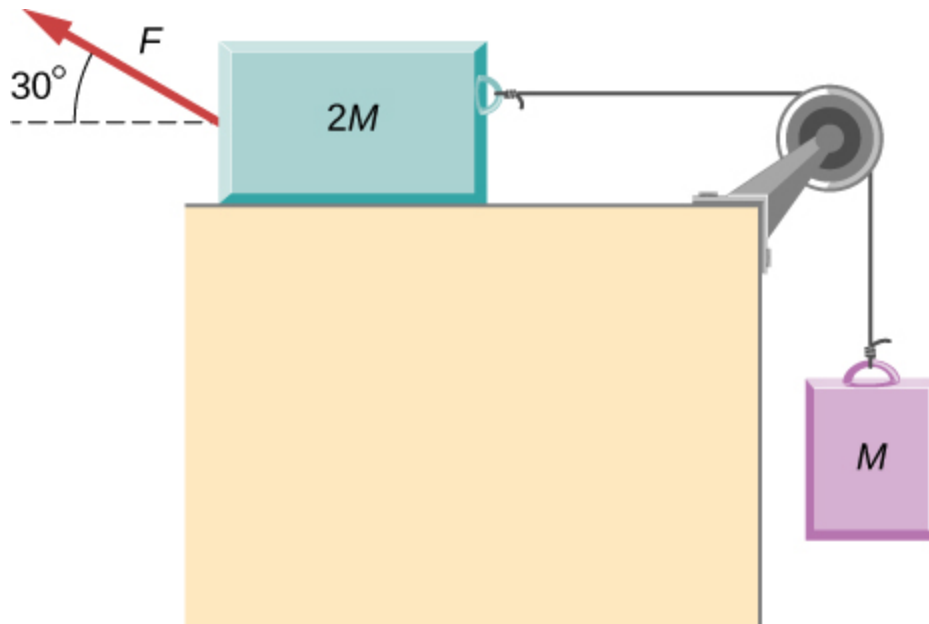
Solution:

199 N

Exercise:

Problem:

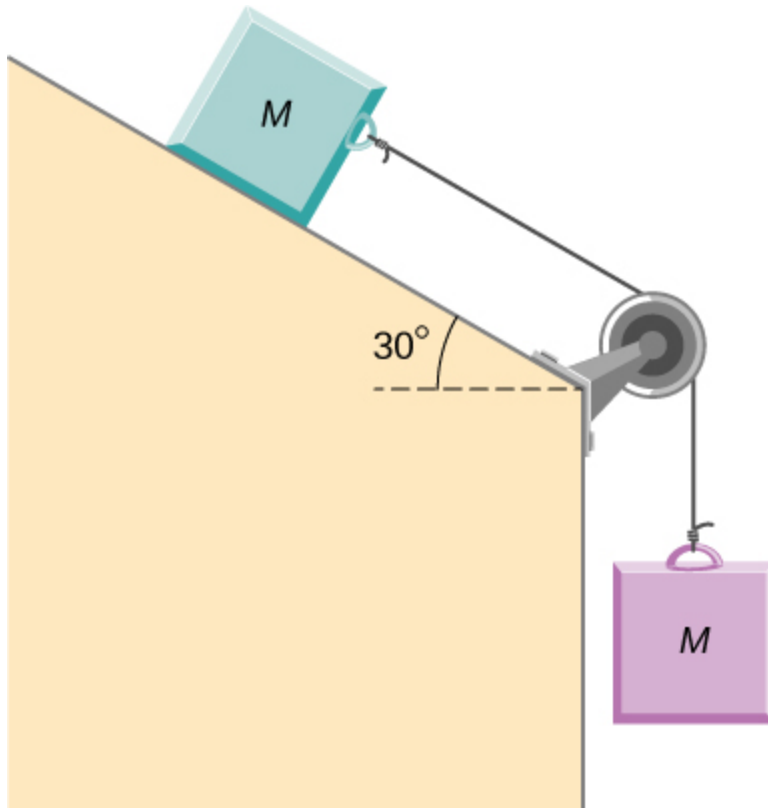
As shown below, if $F = 60.0\text{ N}$ and $M = 4.00\text{ kg}$, what is the magnitude of the acceleration of the suspended object? All surfaces are frictionless.



Exercise:

Problem:

As shown below, if $M = 6.0 \text{ kg}$, what is the tension in the connecting string? The pulley and all surfaces are frictionless.



Solution:

15 N

Exercise:

Problem:

A small space probe is released from a spaceship. The space probe has mass 20.0 kg and contains 90.0 kg of fuel. It starts from rest in deep space, from the origin of a coordinate system based on the spaceship, and burns fuel at the rate of 3.00 kg/s. The engine provides a constant thrust of 120.0 N. (a) Write an expression for the mass of the space probe as a function of time, between 0 and 30 seconds, assuming that the engine ignites fuel beginning at $t = 0$. (b) What is the velocity after 15.0 s? (c) What is the position of the space probe after 15.0 s, with initial position at the origin? (d) Write an expression for the position as a function of time, for $t > 30.0$ s.

Exercise:

Problem:

A half-full recycling bin has mass 3.0 kg and is pushed up a 40.0° incline with constant speed under the action of a 26-N force acting up and parallel to the incline. The incline has friction. What magnitude force must act up and parallel to the incline for the bin to move down the incline at constant velocity?

Solution:

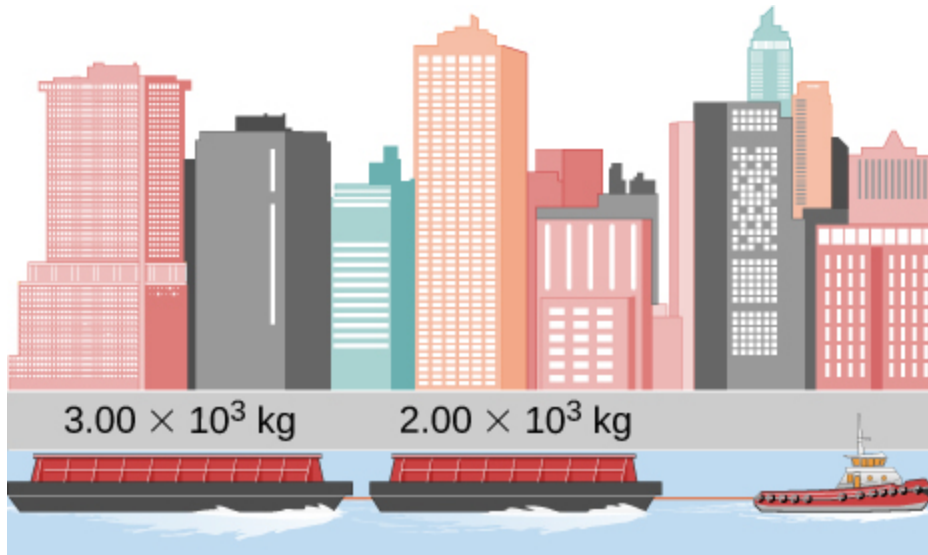
12 N

Exercise:**Problem:**

A child has mass 6.0 kg and slides down a 35° incline with constant speed under the action of a 34-N force acting up and parallel to the incline. What is the coefficient of kinetic friction between the child and the surface of the incline?

Exercise:**Problem:**

The two barges shown here are coupled by a cable of negligible mass. The mass of the front barge is 2.00×10^3 kg and the mass of the rear barge is 3.00×10^3 kg. A tugboat pulls the front barge with a horizontal force of magnitude 20.0×10^3 N, and the frictional forces of the water on the front and rear barges are 8.00×10^3 N and 10.0×10^3 N, respectively. Find the horizontal acceleration of the barges and the tension in the connecting cable.



Solution:

$$a_x = 0.40 \text{ m/s}^2 \text{ and } T = 11.2 \times 10^3 \text{ N}$$

Exercise:

Problem:

If the order of the barges of the preceding exercise is reversed so that the tugboat pulls the 3.00×10^3 -kg barge with a force of $20.0 \times 10^3 \text{ N}$, what are the acceleration of the barges and the tension in the coupling cable?

Exercise:

Problem:

An object with mass m moves along the x -axis. Its position at any time is given by $x(t) = pt^3 + qt^2$ where p and q are constants. Find the net force on this object for any time t .

Solution:

$$m(6pt + 2q)$$

Exercise:

Problem:

A helicopter with mass 2.35×10^4 kg has a position given by $\vec{r}(t) = (0.020 t^3)\hat{i} + (2.2t)\hat{j} - (0.060 t^2)\hat{k}$. Find the net force on the helicopter at $t = 3.0$ s.

Exercise:**Problem:**

Located at the origin, an electric car of mass m is at rest and in equilibrium. A time dependent force of $\vec{F}(t)$ is applied at time $t = 0$, and its components are $F_x(t) = p + nt$ and $F_y(t) = qt$ where p , q , and n are constants. Find the position $\vec{r}(t)$ and velocity $\vec{v}(t)$ as functions of time t .

Solution:

$$\vec{v}(t) = \left(\frac{pt}{m} + \frac{nt^2}{2m} \right) \hat{i} + \left(\frac{qt^2}{2m} \right) \hat{j} \text{ and } \vec{r}(t) = \left(\frac{pt^2}{2m} + \frac{nt^3}{6m} \right) \hat{i} + \left(\frac{qt^3}{6m} \right) \hat{j}$$

Exercise:**Problem:**

A particle of mass m is located at the origin. It is at rest and in equilibrium. A time-dependent force of $\vec{F}(t)$ is applied at time $t = 0$, and its components are $F_x(t) = pt$ and $F_y(t) = n + qt$ where p , q , and n are constants. Find the position $\vec{r}(t)$ and velocity $\vec{v}(t)$ as functions of time t .

Exercise:**Problem:**

A 2.0-kg object has a velocity of $4.0\hat{i}$ m/s at $t = 0$. A constant resultant force of $(2.0\hat{i} + 4.0\hat{j})$ N then acts on the object for 3.0 s. What is the magnitude of the object's velocity at the end of the 3.0-s interval?

Solution:

9.2 m/s

Exercise:**Problem:**

A 1.5-kg mass has an acceleration of $(4.0\hat{\mathbf{i}} - 3.0\hat{\mathbf{j}}) \text{ m/s}^2$. Only two forces act on the mass. If one of the forces is $(2.0\hat{\mathbf{i}} - 1.4\hat{\mathbf{j}}) \text{ N}$, what is the magnitude of the other force?

Exercise:**Problem:**

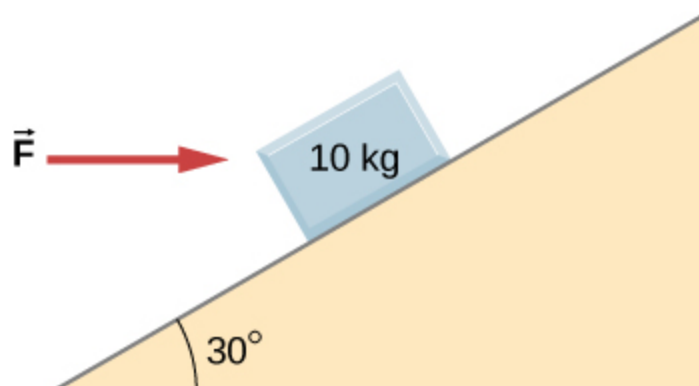
A box is dropped onto a conveyor belt moving at 3.4 m/s. If the coefficient of friction between the box and the belt is 0.27, how long will it take before the box moves without slipping?

Solution:

1.3 s

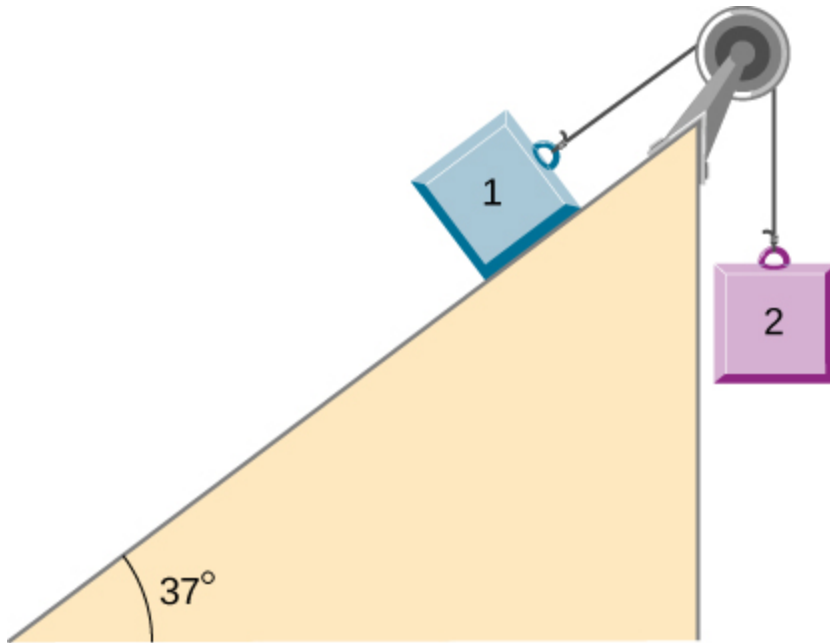
Exercise:**Problem:**

Shown below is a 10.0-kg block being pushed by a horizontal force $\vec{\mathbf{F}}$ of magnitude 200.0 N. The coefficient of kinetic friction between the two surfaces is 0.50. Find the acceleration of the block.



Exercise:**Problem:**

As shown below, the mass of block 1 is $m_1 = 4.0$ kg, while the mass of block 2 is $m_2 = 8.0$ kg. The coefficient of friction between m_1 and the inclined surface is $\mu_k = 0.40$. What is the acceleration of the system?



Solution:

$$3.5 \text{ m/s}^2$$

Exercise:**Problem:**

A student is attempting to move a 30-kg mini-fridge into her dorm room. During a moment of inattention, the mini-fridge slides down a 35 degree incline at constant speed when she applies a force of 25 N acting up and parallel to the incline. What is the coefficient of kinetic friction between the fridge and the surface of the incline?

Exercise:**Problem:**

A crate of mass 100.0 kg rests on a rough surface inclined at an angle of 37.0° with the horizontal. A massless rope to which a force can be applied parallel to the surface is attached to the crate and leads to the top of the incline. In its present state, the crate is just ready to slip and start to move down the plane. The coefficient of friction is 80 % of that for the static case. (a) What is the coefficient of static friction? (b) What is the maximum force that can be applied upward along the plane on the rope and not move the block? (c) With a slightly greater applied force, the block will slide up the plane. Once it begins to move, what is its acceleration and what reduced force is necessary to keep it moving upward at constant speed? (d) If the block is given a slight nudge to get it started down the plane, what will be its acceleration in that direction? (e) Once the block begins to slide downward, what upward force on the rope is required to keep the block from accelerating downward?

Solution:

a. 0.75; b. 1200 N; c. 1.2 m/s^2 and 1080 N; d. -1.2 m/s^2 ; e. 120 N

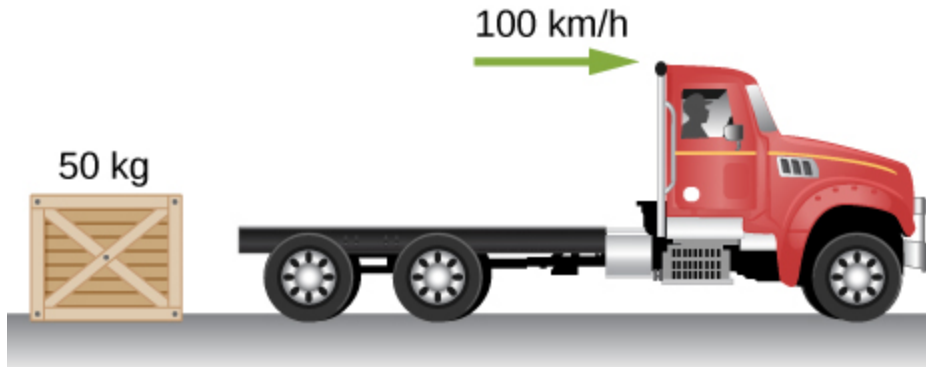
Exercise:**Problem:**

A car is moving at high speed along a highway when the driver makes an emergency braking. The wheels become locked (stop rolling), and the resulting skid marks are 32.0 meters long. If the coefficient of kinetic friction between tires and road is 0.550, and the acceleration was constant during braking, how fast was the car going when the wheels became locked?

Exercise:

Problem:

A crate having mass 50.0 kg falls horizontally off the back of the flatbed truck, which is traveling at 100 km/h. Find the value of the coefficient of kinetic friction between the road and crate if the crate slides 50 m on the road in coming to rest. The initial speed of the crate is the same as the truck, 100 km/h.



Solution:

0.789

Exercise:**Problem:**

A 15-kg sled is pulled across a horizontal, snow-covered surface by a force applied to a rope at 30 degrees with the horizontal. The coefficient of kinetic friction between the sled and the snow is 0.20. (a) If the force is 33 N, what is the horizontal acceleration of the sled? (b) What must the force be in order to pull the sled at constant velocity?

Exercise:

Problem:

A 30.0-g ball at the end of a string is swung in a vertical circle with a radius of 25.0 cm. The tangential velocity is 200.0 cm/s. Find the tension in the string: (a) at the top of the circle, (b) at the bottom of the circle, and (c) at a distance of 12.5 cm from the center of the circle ($r = 12.5$ cm).

Solution:

a. 0.186 N; b. 0.774 N; c. 0.48 N

Exercise:**Problem:**

A particle of mass 0.50 kg starts moves through a circular path in the xy -plane with a position given by $\vec{r}(t) = (4.0 \cos 3t)\hat{i} + (4.0 \sin 3t)\hat{j}$ where r is in meters and t is in seconds. (a) Find the velocity and acceleration vectors as functions of time. (b) Show that the acceleration vector always points toward the center of the circle (and thus represents centripetal acceleration). (c) Find the centripetal force vector as a function of time.

Exercise:**Problem:**

A stunt cyclist rides on the interior of a cylinder 12 m in radius. The coefficient of static friction between the tires and the wall is 0.68. Find the value of the minimum speed for the cyclist to perform the stunt.

Solution:

13 m/s

Exercise:

Problem:

When a body of mass 0.25 kg is attached to a vertical massless spring, it is extended 5.0 cm from its unstretched length of 4.0 cm. The body and spring are placed on a horizontal frictionless surface and rotated about the held end of the spring at 2.0 rev/s. How far is the spring stretched?

Exercise:**Problem:**

A piece of bacon starts to slide down the pan when one side of a pan is raised up 5.0 cm. If the length of the pan from pivot to the raising point is 23.5 cm, what is the coefficient of static friction between the pan and the bacon?

Solution:

0.21

Exercise:**Problem:**

A plumb bob hangs from the roof of a railroad car. The car rounds a circular track of radius 300.0 m at a speed of 90.0 km/h. At what angle relative to the vertical does the plumb bob hang?

Exercise:**Problem:**

An airplane flies at 120.0 m/s and banks at a 30° angle. If its mass is 2.50×10^3 kg, (a) what is the magnitude of the lift force? (b) what is the radius of the turn?

Solution:

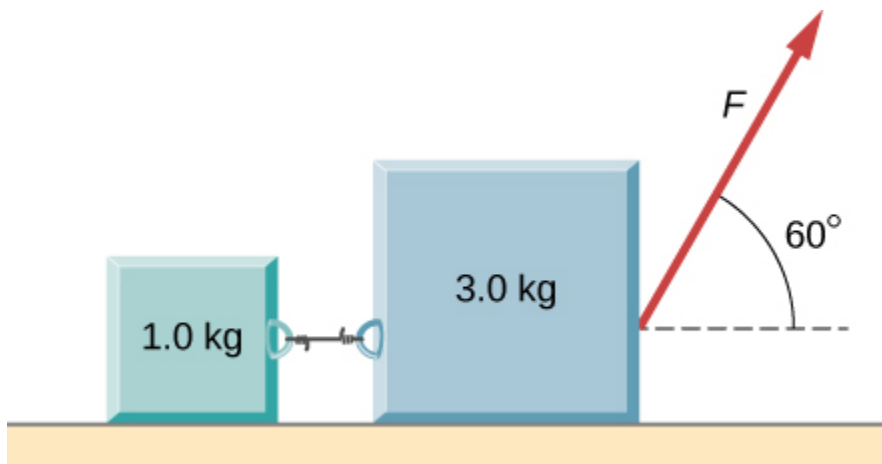
a. 28,300 N; b. 2540 m

Exercise:**Problem:**

The position of a particle is given by $\vec{r}(t) = A (\cos \omega t \hat{i} + \sin \omega t \hat{j})$, where ω is a constant. (a) Show that the particle moves in a circle of radius A . (b) Calculate $d\vec{r}/dt$ and then show that the speed of the particle is a constant $A\omega$. (c) Determine $d^2\vec{r}/dt^2$ and show that a is given by $a_c = r\omega^2$. (d) Calculate the centripetal force on the particle. [Hint: For (b) and (c), you will need to use $(d/dt)(\cos \omega t) = -\omega \sin \omega t$ and $(d/dt)(\sin \omega t) = \omega \cos \omega t$.

Exercise:**Problem:**

Two blocks connected by a string are pulled across a horizontal surface by a force applied to one of the blocks, as shown below. The coefficient of kinetic friction between the blocks and the surface is 0.25. If each block has an acceleration of 2.0 m/s^2 to the right, what is the magnitude F of the applied force?



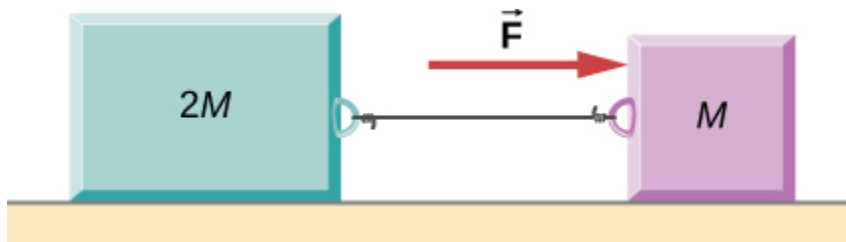
Solution:

25 N

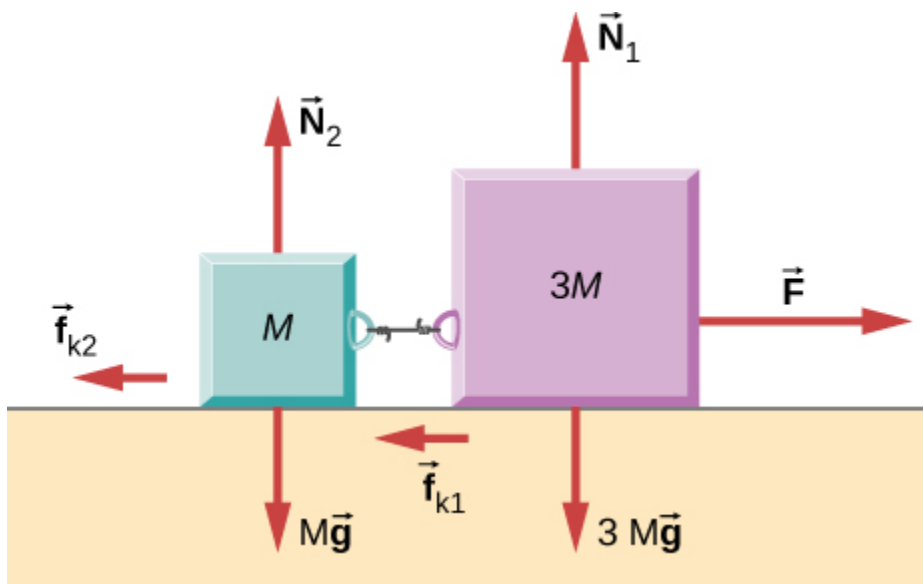
Exercise:

Problem:

As shown below, the coefficient of kinetic friction between the surface and the larger block is 0.20, and the coefficient of kinetic friction between the surface and the smaller block is 0.30. If $F = 10\text{ N}$ and $M = 1.0\text{ kg}$, what is the tension in the connecting string?

**Exercise:****Problem:**

In the figure, the coefficient of kinetic friction between the surface and the blocks is μ_k . If $M = 1.0\text{ kg}$, find an expression for the magnitude of the acceleration of either block (in terms of F , μ_k , and g).



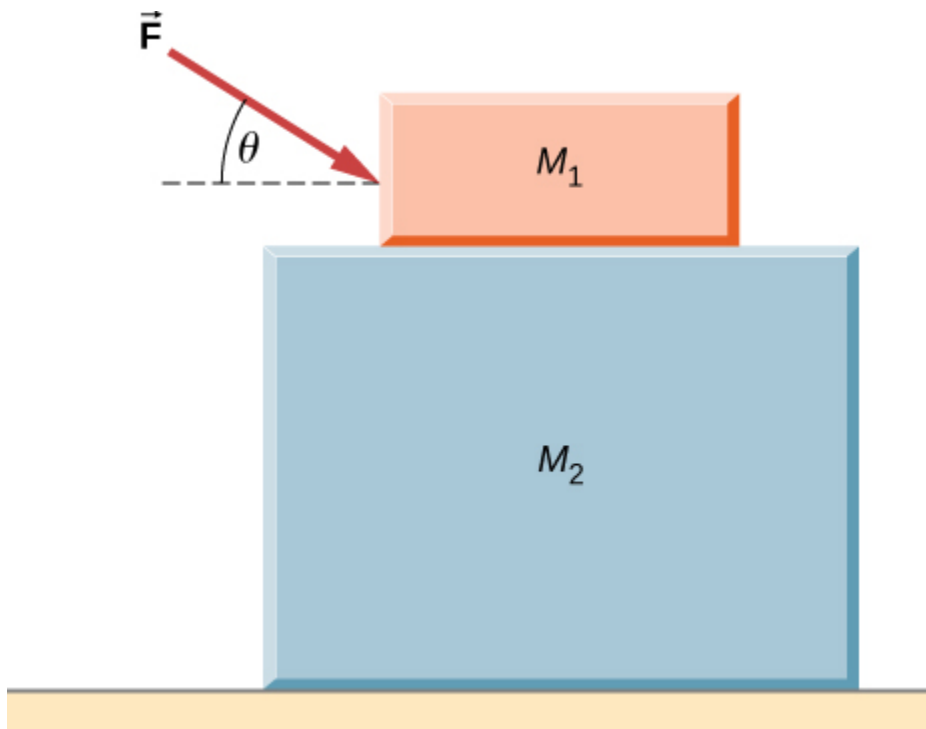
Solution:

$$a = \frac{F}{4} - \mu_k g$$

Exercise:

Problem:

Two blocks are stacked as shown below, and rest on a frictionless surface. There is friction between the two blocks (coefficient of friction μ). An external force is applied to the top block at an angle θ with the horizontal. What is the maximum force F that can be applied for the two blocks to move together?



Exercise:

Problem:

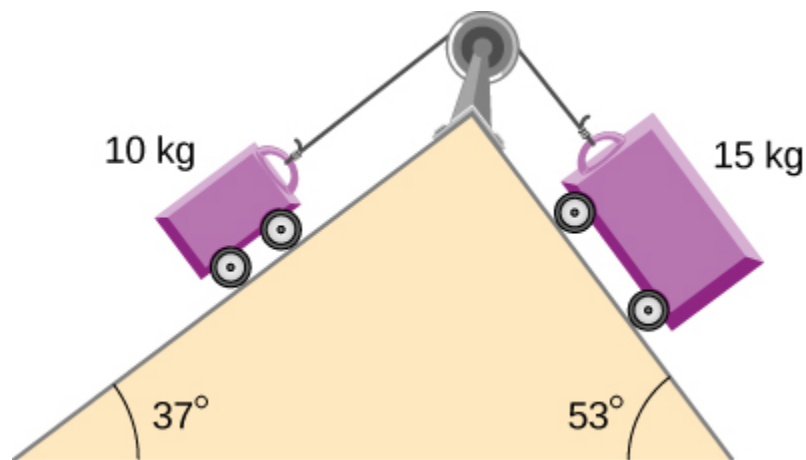
A box rests on the (horizontal) back of a truck. The coefficient of static friction between the box and the surface on which it rests is 0.24. What maximum distance can the truck travel (starting from rest and moving horizontally with constant acceleration) in 3.0 s without having the box slide?

Solution:

11 m

Exercise:**Problem:**

A double-incline plane is shown below. The coefficient of friction on the left surface is 0.30, and on the right surface 0.16. Calculate the acceleration of the system.

**Challenge Problems****Exercise:****Problem:**

In a later chapter, you will find that the weight of a particle varies with altitude such that $w = \frac{mgr_0^2}{r^2}$ where r_0 is the radius of Earth and r is the distance from Earth's center. If the particle is fired vertically with velocity v_0 from Earth's surface, determine its velocity as a function of position r . (Hint: use $a dr = v dv$, the rearrangement mentioned in the text.)

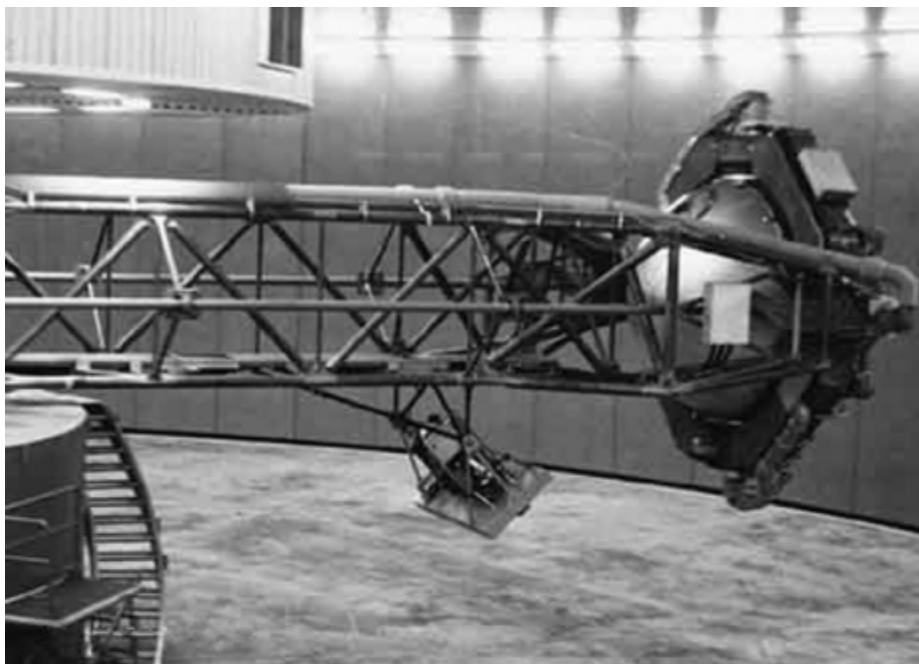
Solution:

$$v = \sqrt{v_0^2 - 2gr_0 \left(1 - \frac{r_0}{r}\right)}$$

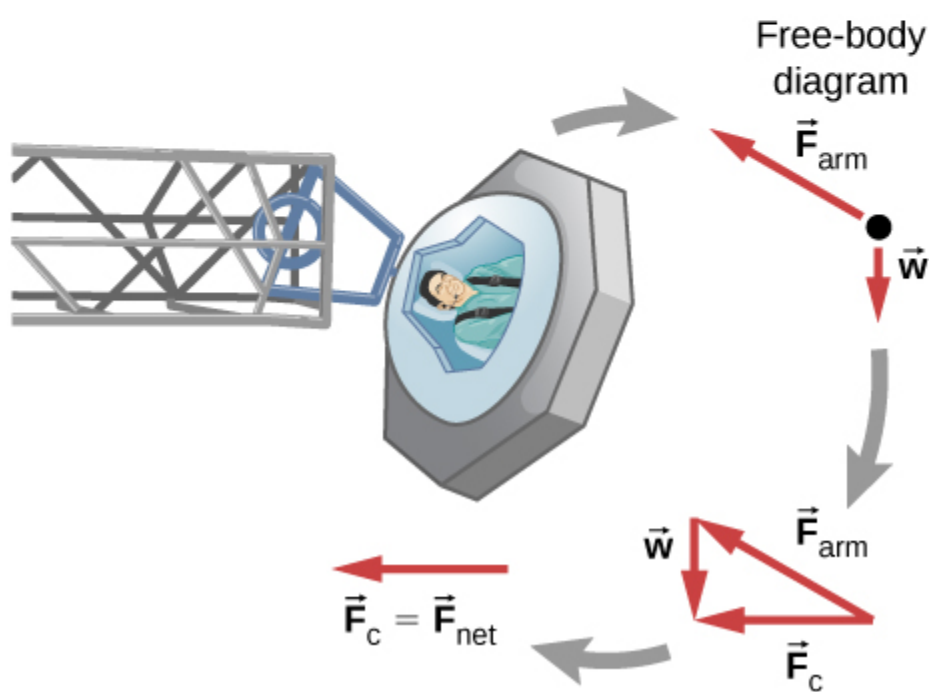
Exercise:

Problem:

A large centrifuge, like the one shown below, is used to expose aspiring astronauts to accelerations similar to those experienced in rocket launches and atmospheric reentries. (a) At what angular velocity is the centripetal acceleration $10g$ if the rider is 15.0 m from the center of rotation? (b) The rider's cage hangs on a pivot at the end of the arm, allowing it to swing outward during rotation as shown in the bottom accompanying figure. At what angle θ below the horizontal will the cage hang when the centripetal acceleration is $10g$? (*Hint:* The arm supplies centripetal force and supports the weight of the cage. Draw a free-body diagram of the forces to see what the angle θ should be.)



(a)



(b)

Exercise:

Problem:

A car of mass 1000.0 kg is traveling along a level road at 100.0 km/h when its brakes are applied. Calculate the stopping distance if the coefficient of kinetic friction of the tires is 0.500. Neglect air resistance. (*Hint: since the distance traveled is of interest rather than the time, x is the desired independent variable and not t . Use the Chain Rule to change the variable: $\frac{dv}{dt} = \frac{dv}{dx} \frac{dx}{dt} = v \frac{dv}{dx}$.*)

Solution:

78.7 m

Exercise:**Problem:**

An airplane flying at 200.0 m/s makes a turn that takes 4.0 min. What bank angle is required? What is the percentage increase in the perceived weight of the passengers?

Exercise:**Problem:**

A skydiver is at an altitude of 1520 m. After 10.0 seconds of free fall, he opens his parachute and finds that the air resistance, F_D , is given by the formula $F_D = -bv$, where b is a constant and v is the velocity. If $b = 0.750$, and the mass of the skydiver is 82.0 kg, first set up differential equations for the velocity and the position, and then find: (a) the speed of the skydiver when the parachute opens, (b) the distance fallen before the parachute opens, (c) the terminal velocity after the parachute opens (find the limiting velocity), and (d) the time the skydiver is in the air after the parachute opens.

Solution:

a. 98 m/s; b. 490 m; c. 107 m/s; d. 9.6 s

Exercise:

Problem:

In a television commercial, a small, spherical bead of mass 4.00 g is released from rest at $t = 0$ in a bottle of liquid shampoo. The terminal speed is observed to be 2.00 cm/s. Find (a) the value of the constant b in the equation $v = \frac{mg}{b}(1 - e^{-bt/m})$, and (b) the value of the resistive force when the bead reaches terminal speed.

Exercise:**Problem:**

A boater and motor boat are at rest on a lake. Together, they have mass 200.0 kg. If the thrust of the motor is a constant force of 40.0 N in the direction of motion, and if the resistive force of the water is numerically equivalent to 2 times the speed v of the boat, set up and solve the differential equation to find: (a) the velocity of the boat at time t ; (b) the limiting velocity (the velocity after a long time has passed).

Solution:

a. $v = 20.0(1 - e^{-0.01t})$; b. $v_{\text{limiting}} = 20 \text{ m/s}$

Glossary

drag force

force that always opposes the motion of an object in a fluid; unlike simple friction, the drag force is proportional to some function of the velocity of the object in that fluid

terminal velocity

constant velocity achieved by a falling object, which occurs when the weight of the object is balanced by the upward drag force

Introduction

class="introduction"

A sprinter exerts her maximum power with the greatest force in the short time her foot is in contact with the ground. This adds to her kinetic energy, preventing her from slowing down during the race.

Pushing back hard on the track generates a reaction force that propels the sprinter forward to win at the finish.

(credit: modification of work by Marie-

Lan
Nguyen)



In this chapter, we discuss some basic physical concepts involved in every physical motion in the universe, going beyond the concepts of force and change in motion, which we discussed in [Motion in Two and Three Dimensions](#) and [Newton's Laws of Motion](#). These concepts are work, kinetic energy, and power. We explain how these quantities are related to one another, which will lead us to a fundamental relationship called the work-energy theorem. In the next chapter, we generalize this idea to the broader principle of conservation of energy.

The application of Newton's laws usually requires solving differential equations that relate the forces acting on an object to the accelerations they produce. Often, an analytic solution is intractable or impossible, requiring lengthy numerical solutions or simulations to get approximate results. In such situations, more general relations, like the work-energy theorem (or the conservation of energy), can still provide useful answers to many questions and require a more modest amount of mathematical calculation. In particular, you will see how the work-energy theorem is useful in relating the speeds of a particle, at different points along its trajectory, to the forces acting on it, even when the trajectory is otherwise too complicated to deal with. Thus, some aspects of motion can be addressed with fewer equations and without vector decompositions.

Work

By the end of this section, you will be able to:

- Represent the work done by any force
- Evaluate the work done for various forces

In physics, **work** is done on an object when energy is transferred to the object. In other words, work is done when a force acts on something that undergoes a displacement from one position to another. Forces can vary as a function of position, and displacements can be along various paths between

two points. We first define the increment of work dW done by a force $\vec{\mathbf{F}}$ acting through an infinitesimal displacement $d\vec{\mathbf{r}}$ as the dot product of these two vectors:

Note:

Equation:

$$dW = \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}} = |\vec{\mathbf{F}}| |d\vec{\mathbf{r}}| \cos \theta.$$

Then, we can add up the contributions for infinitesimal displacements, along a path between two positions, to get the total work.

Note:

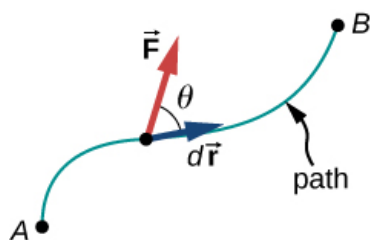
Work Done by a Force

The work done by a force is the integral of the force with respect to displacement along the path of the displacement:

Equation:

$$W_{AB} = \int_{\text{path } AB} \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}}.$$

The vectors involved in the definition of the work done by a force acting on a particle are illustrated in [\[link\]](#).



Vectors used to define work. The force acting on a particle and its infinitesimal displacement are shown at one point along the path between A and B. The infinitesimal work is the dot product of these two vectors; the total work is the integral of the dot product along the path.

We choose to express the dot product in terms of the magnitudes of the vectors and the cosine of the angle between them, because the meaning of the dot product for work can be put into words more directly in terms of magnitudes and angles. We could equally well have expressed the dot product in terms of the various components introduced in [Vectors](#). In two dimensions, these were the x - and y -components in Cartesian coordinates, or the r - and φ -components in polar coordinates; in three dimensions, it was just x -, y -, and z -components. Which choice is more convenient depends on the situation. In words, you can express [\[link\]](#) for the work done by a force acting over a displacement as a product of one component acting parallel to the other component. From the properties of vectors, it doesn't matter if you take the component of the force parallel to the displacement or the component of the displacement parallel to the force—you get the same result either way.

Recall that the magnitude of a force times the cosine of the angle the force makes with a given direction is the component of the force in the given direction. The components of a vector can be positive, negative, or zero, depending on whether the angle between the vector and the component-direction is between 0° and 90° or 90° and 180° , or is equal to 90° . As a result, the work done by a force can be positive, negative, or zero, depending on whether the force is generally in the direction of the displacement, generally opposite to the displacement, or perpendicular to the displacement. The maximum work is done by a given force when it is along the direction of the displacement ($\cos \theta = \pm 1$), and zero work is done when the force is perpendicular to the displacement ($\cos \theta = 0$).

The units of work are units of force multiplied by units of length, which in the SI system is newtons times meters, $\text{N} \cdot \text{m}$. This combination is called a joule, for historical reasons that we will mention later, and is abbreviated as J. In the English system, still used in the United States, the unit of force

is the pound (lb) and the unit of distance is the foot (ft), so the unit of work is the foot-pound (ft · lb).

Work Done by Constant Forces and Contact Forces

The simplest work to evaluate is that done by a force that is constant in magnitude and direction. In this case, we can factor out the force; the remaining integral is just the total displacement, which only depends on the end points A and B , but not on the path between them:

Equation:

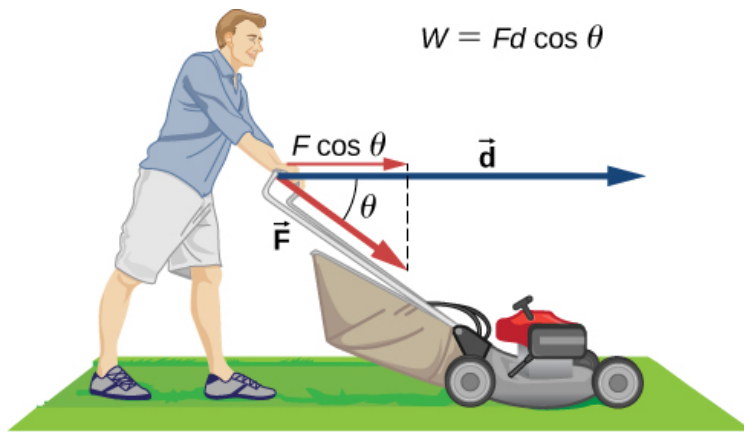
$$W_{AB} = \vec{\mathbf{F}} \cdot \int_A^B d\vec{\mathbf{r}} = \vec{\mathbf{F}} \cdot (\vec{\mathbf{r}}_B - \vec{\mathbf{r}}_A) = \left| \vec{\mathbf{F}} \right| \left| \vec{\mathbf{r}}_B - \vec{\mathbf{r}}_A \right| \cos \theta \quad (\text{constant force}).$$

We can also see this by writing out [\[link\]](#) in Cartesian coordinates and using the fact that the components of the force are constant:

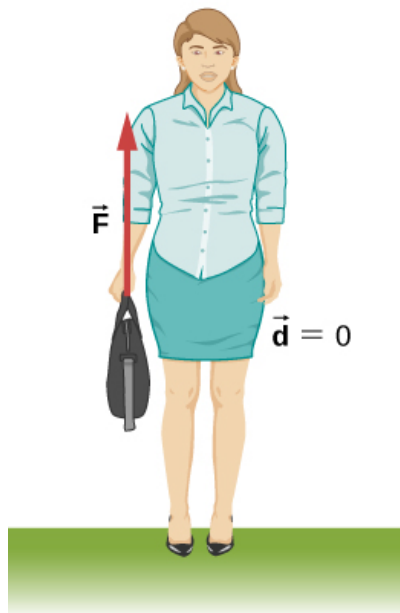
Equation:

$$\begin{aligned} W_{AB} &= \int_{\text{path } AB} \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}} = \int_{\text{path } AB} (F_x dx + F_y dy + F_z dz) = F_x \int_A^B dx + F_y \int_A^B dy + F_z \int_A^B dz \\ &= F_x (x_B - x_A) + F_y (y_B - y_A) + F_z (z_B - z_A) = \vec{\mathbf{F}} \cdot (\vec{\mathbf{r}}_B - \vec{\mathbf{r}}_A). \end{aligned}$$

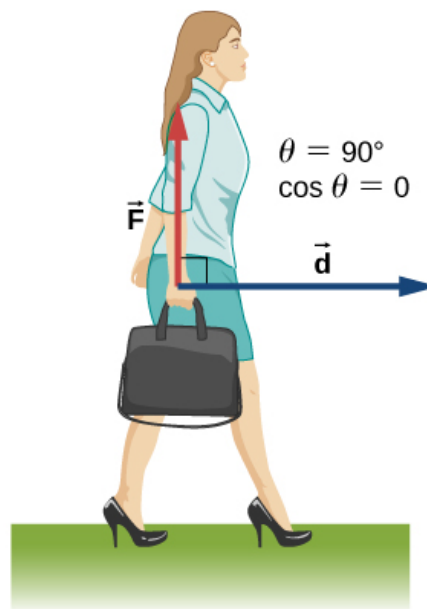
[\[link\]](#)(a) shows a person exerting a constant force $\vec{\mathbf{F}}$ along the handle of a lawn mower, which makes an angle θ with the horizontal. The horizontal displacement of the lawn mower, over which the force acts, is $\vec{\mathbf{d}}$. The work done on the lawn mower is $W = \vec{\mathbf{F}} \cdot \vec{\mathbf{d}} = Fd \cos \theta$, which the figure also illustrates as the horizontal component of the force times the magnitude of the displacement.



(a)



(b)



(c)

Work done by a constant force. (a) A person pushes a lawn mower with a constant force. The component of the force parallel to the displacement is the work done, as shown in the equation in the figure. (b) A person holds a briefcase. No work is done because the displacement is zero. (c) The person in (b) walks horizontally while holding the briefcase. No work is done because $\cos \theta$ is zero.

[\[link\]](#) (b) shows a person holding a briefcase. The person must exert an upward force, equal in magnitude to the weight of the briefcase, but this force does no work, because the displacement over which it acts is zero.

In [\[link\]](#)(c), where the person in (b) is walking horizontally with constant speed, the work done by the person on the briefcase is still zero, but now because the angle between the force exerted and the displacement is 90° ($\vec{\mathbf{F}}$ perpendicular to $\vec{\mathbf{d}}$) and $\cos 90^\circ = 0$.

Example:

Calculating the Work You Do to Push a Lawn Mower

How much work is done on the lawn mower by the person in [\[link\]](#)(a) if he exerts a constant force of 75.0 N at an angle 35° below the horizontal and pushes the mower 25.0 m on level ground?

Strategy

We can solve this problem by substituting the given values into the definition of work done on an object by a constant force, stated in the equation $W = Fd \cos \theta$. The force, angle, and displacement are given, so that only the work W is unknown.

Solution

The equation for the work is

Equation:

$$W = Fd \cos \theta.$$

Substituting the known values gives

Equation:

$$W = (75.0 \text{ N})(25.0 \text{ m})\cos(35.0^\circ) = 1.54 \times 10^3 \text{ J}.$$

Significance

Even though one and a half kilojoules may seem like a lot of work, we will see in [Potential Energy and Conservation of Energy](#) that it's only about as much work as you could do by burning one sixth of a gram of fat.

When you mow the grass, other forces act on the lawn mower besides the force you exert—namely, the contact force of the ground and the gravitational force of Earth. Let's consider the work done by these forces in general. For an object moving on a surface, the displacement $d\vec{\mathbf{r}}$ is tangent to the surface. The part of the contact force on the object that is perpendicular to the surface is the normal force $\vec{\mathbf{N}}$. Since the cosine of the angle between the normal and the tangent to a surface is zero, we have

Equation:

$$dW_N = \vec{\mathbf{N}} \cdot d\vec{\mathbf{r}} = 0.$$

The normal force never does work under these circumstances. (Note that if the displacement $d\vec{\mathbf{r}}$ did have a relative component perpendicular to the surface, the object would either leave the surface or break through it, and there would no longer be any normal contact force. However, if the object is more than a particle, and has an internal structure, the normal contact force can do work on it, for example, by displacing it or deforming its shape. This will be mentioned in the next chapter.)

The part of the contact force on the object that is parallel to the surface is friction, \vec{f} . For this object sliding along the surface, kinetic friction \vec{f}_k is opposite to $d\vec{r}$, relative to the surface, so the work done by kinetic friction is negative. If the magnitude of \vec{f}_k is constant (as it would be if all the other forces on the object were constant), then the work done by friction is

Note:

Equation:

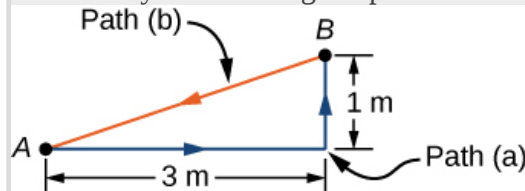
$$W_{\text{fr}} = \int_A^B \vec{f}_k \cdot d\vec{r} = -f_k \int_A^B |dr| = -f_k |l_{AB}|,$$

where $|l_{AB}|$ is the path length on the surface. The force of static friction does no work in the reference frame between two surfaces because there is never displacement between the surfaces. As an external force, static friction can do work. Static friction can keep someone from sliding off a sled when the sled is moving and perform positive work on the person. If you're driving your car at the speed limit on a straight, level stretch of highway, the negative work done by air resistance is balanced by the positive work done by the static friction of the road on the drive wheels. You can pull the rug out from under an object in such a way that it slides backward relative to the rug, but forward relative to the floor. In this case, kinetic friction exerted by the rug on the object could be in the same direction as the displacement of the object, relative to the floor, and do positive work. The bottom line is that you need to analyze each particular case to determine the work done by the forces, whether positive, negative or zero.

Example:

Moving a Couch

You decide to move your couch to a new position on your horizontal living room floor. The normal force on the couch is 1 kN and the coefficient of friction is 0.6. (a) You first push the couch 3 m parallel to a wall and then 1 m perpendicular to the wall (A to B in [link](#)). How much work is done by the frictional force? (b) You don't like the new position, so you move the couch straight back to its original position (B to A in [link](#)). What was the total work done against friction moving the couch away from its original position and back again?



Top view of paths for moving a couch.

Strategy

The magnitude of the force of kinetic friction on the couch is constant, equal to the coefficient of friction times the normal force, $f_K = \mu_K N$. Therefore, the work done by it is $W_{\text{fr}} = -f_K d$, where d is the path length traversed. The segments of the paths are the sides of a right triangle, so the path lengths are easily calculated. In part (b), you can use the fact that the work done against a force is the negative of the work done by the force.

Solution

- a. The work done by friction is

Equation:

$$W = -(0.6)(1 \text{ kN})(3 \text{ m} + 1 \text{ m}) = -2.4 \text{ kJ}.$$

- b. The length of the path along the hypotenuse is $\sqrt{10} \text{ m}$, so the total work done against friction is

Equation:

$$W = (0.6)(1 \text{ kN})(3 \text{ m} + 1 \text{ m} + \sqrt{10} \text{ m}) = 4.3 \text{ kJ}.$$

Significance

The total path over which the work of friction was evaluated began and ended at the same point (it was a closed path), so that the total displacement of the couch was zero. However, the total work was not zero. The reason is that forces like friction are classified as nonconservative forces, or dissipative forces, as we discuss in the next chapter.

Note:**Exercise:****Problem:**

Check Your Understanding Can kinetic friction ever be a constant force for all paths?

Solution:

No, only its magnitude can be constant; its direction must change, to be always opposite the relative displacement along the surface.

The other force on the lawn mower mentioned above was Earth's gravitational force, or the weight of the mower. Near the surface of Earth, the gravitational force on an object of mass m has a constant magnitude, mg , and constant direction, vertically down. Therefore, the work done by gravity on an object is the dot product of its weight and its displacement. In many cases, it is convenient to express the dot product for gravitational work in terms of the x -, y -, and z -components of the vectors. A typical coordinate system has the x -axis horizontal and the y -axis vertically up. Then the gravitational force is $-mg\hat{\mathbf{j}}$, so the work done by gravity, over any path from A to B , is

Note:

Equation:

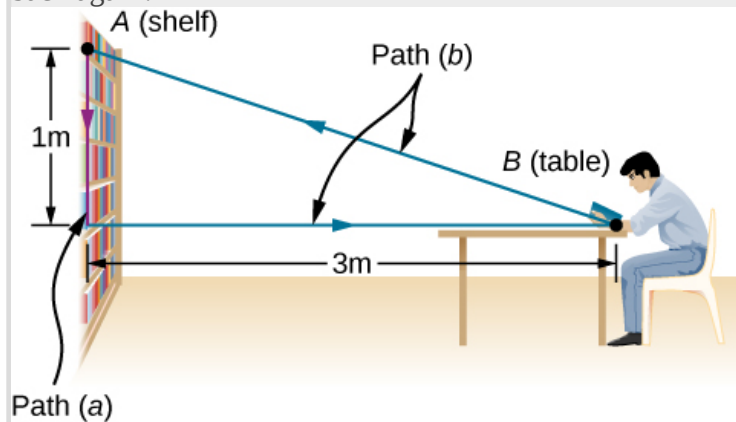
$$W_{\text{grav},AB} = -mg\hat{\mathbf{j}} \cdot (\vec{\mathbf{r}}_B - \vec{\mathbf{r}}_A) = -mg(y_B - y_A).$$

The work done by a constant force of gravity on an object depends only on the object's weight and the difference in height through which the object is displaced. Gravity does negative work on an object that moves upward ($y_B > y_A$), or, in other words, you must do positive work against gravity to lift an object upward. Alternately, gravity does positive work on an object that moves downward ($y_B < y_A$), or you do negative work against gravity to “lift” an object downward, controlling its descent so it doesn't drop to the ground. (“Lift” is used as opposed to “drop”.)

Example:

Shelving a Book

You lift an oversized library book, weighing 20 N, 1 m vertically down from a shelf, and carry it 3 m horizontally to a table ([link](#)). How much work does gravity do on the book? (b) When you're finished, you move the book in a straight line back to its original place on the shelf. What was the total work done against gravity, moving the book away from its original position on the shelf and back again?



Side view of the paths for moving a book to and from a shelf.

Strategy

We have just seen that the work done by a constant force of gravity depends only on the weight of the object moved and the difference in height for the path taken, $W_{AB} = -mg(y_B - y_A)$. We can evaluate the difference in height to answer (a) and (b).

Solution

- a. Since the book starts on the shelf and is lifted down $y_B - y_A = -1$ m, we have

Equation:

$$W = -(20 \text{ N})(-1 \text{ m}) = 20 \text{ J}.$$

- b. There is zero difference in height for any path that begins and ends at the same place on the shelf, so $W = 0$.

Significance

Gravity does positive work (20 J) when the book moves down from the shelf. The gravitational force between two objects is an attractive force, which does positive work when the objects get closer together. Gravity does zero work (0 J) when the book moves horizontally from the shelf to the table and negative work (−20 J) when the book moves from the table back to the shelf. The total work done by gravity is zero $[20 \text{ J} + 0 \text{ J} + (-20 \text{ J}) = 0]$. Unlike friction or other dissipative forces, described in [\[link\]](#), the total work done against gravity, over any closed path, is zero. Positive work is done against gravity on the upward parts of a closed path, but an equal amount of negative work is done against gravity on the downward parts. In other words, work done *against* gravity, lifting an object *up*, is “given back” when the object comes back down. Forces like gravity (those that do zero work over any closed path) are classified as conservative forces and play an important role in physics.

Note:

Exercise:

Problem:

Check Your Understanding Can Earth’s gravity ever be a constant force for all paths?

Solution:

No, it’s only approximately constant near Earth’s surface.

Work Done by Forces that Vary

In general, forces may vary in magnitude and direction at points in space, and paths between two points may be curved. The infinitesimal work done by a variable force can be expressed in terms of the components of the force and the displacement along the path,

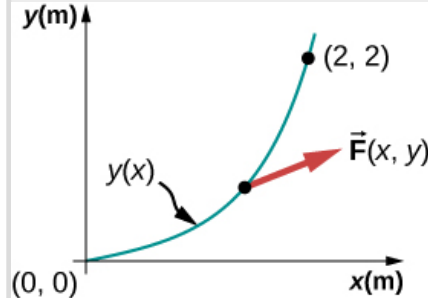
Equation:

$$dW = F_x dx + F_y dy + F_z dz.$$

Here, the components of the force are functions of position along the path, and the displacements depend on the equations of the path. (Although we chose to illustrate dW in Cartesian coordinates, other coordinates are better suited to some situations.) [\[link\]](#) defines the total work as a line integral, or the limit of a sum of infinitesimal amounts of work. The physical concept of work is straightforward: you calculate the work for tiny displacements and add them up. Sometimes the mathematics can seem complicated, but the following example demonstrates how cleanly they can operate.

Example:**Work Done by a Variable Force over a Curved Path**

An object moves along a parabolic path $y = (0.5 \text{ m}^{-1})x^2$ from the origin $A = (0, 0)$ to the point $B = (2 \text{ m}, 2 \text{ m})$ under the action of a force $\vec{F} = (5 \text{ N/m})y\hat{i} + (10 \text{ N/m})x\hat{j}$ ([link](#)). Calculate the work done.



The parabolic path of a particle acted on by a given force.

Strategy

The components of the force are given functions of x and y . We can use the equation of the path to express y and dy in terms of x and dx ; namely,

Equation:

$$y = (0.5 \text{ m}^{-1})x^2 \text{ and } dy = 2(0.5 \text{ m}^{-1})x dx.$$

Then, the integral for the work is just a definite integral of a function of x .

Solution

The infinitesimal element of work is

Equation:

$$\begin{aligned} dW &= F_x dx + F_y dy = (5 \text{ N/m})y dx + (10 \text{ N/m})x dy \\ &= (5 \text{ N/m})(0.5 \text{ m}^{-1})x^2 dx + (10 \text{ N/m})2(0.5 \text{ m}^{-1})x^2 dx = (12.5 \text{ N/m}^2)x^2 dx. \end{aligned}$$

The integral of x^2 is $x^3/3$, so

Equation:

$$W = \int_0^{2 \text{ m}} (12.5 \text{ N/m}^2)x^2 dx = (12.5 \text{ N/m}^2) \frac{x^3}{3} \bigg|_0^{2 \text{ m}} = (12.5 \text{ N/m}^2) \left(\frac{8}{3} \right) = 33.3 \text{ J}.$$

Significance

This integral was not hard to do. You can follow the same steps, as in this example, to calculate line integrals representing work for more complicated forces and paths. In this example, everything was given in terms of x - and y -components, which are easiest to use in evaluating the work in this case. In other situations, magnitudes and angles might be easier.

Note:

Exercise:

Problem:

Check Your Understanding Find the work done by the same force in [\[link\]](#) over a cubic path, $y = (0.25 \text{ m}^{-2})x^3$, between the same points $A = (0, 0)$ and $B = (2 \text{ m}, 2 \text{ m})$.

Solution:

$$W = 35 \text{ J}$$

You saw in [\[link\]](#) that to evaluate a line integral, you could reduce it to an integral over a single variable or parameter. Usually, there are several ways to do this, which may be more or less convenient, depending on the particular case. In [\[link\]](#), we reduced the line integral to an integral over x , but we could equally well have chosen to reduce everything to a function of y . We didn't do that because the functions in y involve the square root and fractional exponents, which may be less familiar, but for illustrative purposes, we do this now. Solving for x and dx , in terms of y , along the parabolic path, we get

Equation:

$$x = \sqrt{y/(0.5 \text{ m}^{-1})} = \sqrt{(2 \text{ m})y} \text{ and } dx = \sqrt{(2 \text{ m})} \times \frac{1}{2} dy / \sqrt{y} = dy / \sqrt{(2 \text{ m}^{-1})y}.$$

The components of the force, in terms of y , are

Equation:

$$F_x = (5 \text{ N/m})y \text{ and } F_y = (10 \text{ N/m})x = (10 \text{ N/m})\sqrt{(2 \text{ m})y},$$

so the infinitesimal work element becomes

Equation:

$$\begin{aligned} dW &= F_x dx + F_y dy = \frac{(5 \text{ N/m})y dy}{\sqrt{(2 \text{ m}^{-1})y}} + (10 \text{ N/m})\sqrt{(2 \text{ m})y} dy \\ &= (5 \text{ N} \cdot \text{m}^{-1/2}) \left(\frac{1}{\sqrt{2}} + 2\sqrt{2} \right) \sqrt{y} dy = (17.7 \text{ N} \cdot \text{m}^{-1/2}) y^{1/2} dy. \end{aligned}$$

The integral of $y^{1/2}$ is $\frac{2}{3} y^{3/2}$, so the work done from A to B is

Equation:

$$W = \int_0^{2 \text{ m}} (17.7 \text{ N} \cdot \text{m}^{-1/2}) y^{1/2} dy = (17.7 \text{ N} \cdot \text{m}^{-1/2}) \frac{2}{3} (2 \text{ m})^{3/2} = 33.3 \text{ J}.$$

As expected, this is exactly the same result as before.

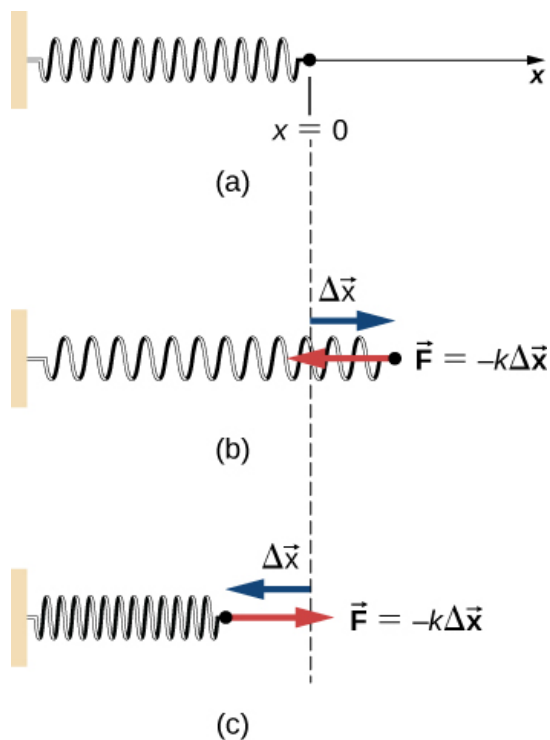
One very important and widely applicable variable force is the force exerted by a perfectly elastic spring, which satisfies Hooke's law $\vec{F} = -k\Delta\vec{x}$, where k is the spring constant, and $\Delta\vec{x} = \vec{x} - \vec{x}_{\text{eq}}$ is the displacement from the spring's unstretched (equilibrium) position ([Newton's Laws of Motion](#)). Note that the unstretched position is only the same as the equilibrium position if no other forces are acting (or, if they are, they cancel one another). Forces between molecules, or in any system undergoing small displacements from a stable equilibrium, behave approximately like a spring force.

To calculate the work done by a spring force, we can choose the x -axis along the length of the spring, in the direction of increasing length, as in [\[link\]](#), with the origin at the equilibrium position $x_{\text{eq}} = 0$. (Then positive x corresponds to a stretch and negative x to a compression.) With this choice of coordinates, the spring force has only an x -component, $F_x = -kx$, and the work done when x changes from x_A to x_B is

Note:

Equation:

$$W_{\text{spring},AB} = \int_A^B F_x dx = -k \int_A^B x dx = -k \frac{x^2}{2} \Big|_A^B = -\frac{1}{2}k(x_B^2 - x_A^2).$$

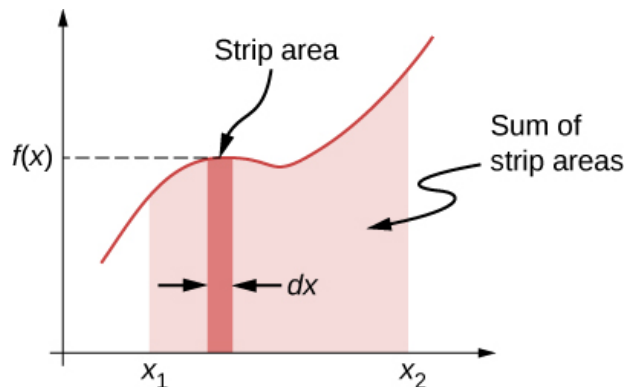


(a) The spring exerts no force at its equilibrium position. The spring

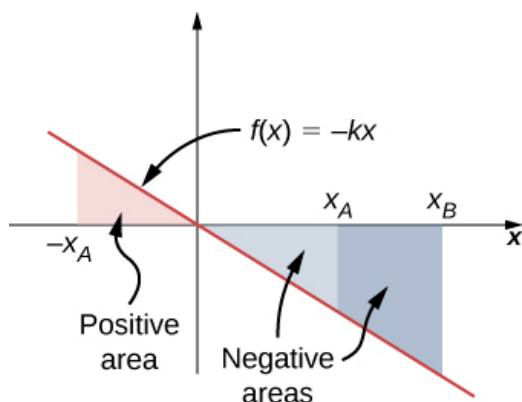
exerts a force in the opposite direction to (b) an extension or stretch, and (c) a compression.

Notice that W_{AB} depends only on the starting and ending points, A and B , and is independent of the actual path between them, as long as it starts at A and ends at B . That is, the actual path could involve going back and forth before ending.

Another interesting thing to notice about [\[link\]](#) is that, for this one-dimensional case, you can readily see the correspondence between the work done by a force and the area under the curve of the force versus its displacement. Recall that, in general, a one-dimensional integral is the limit of the sum of infinitesimals, $f(x)dx$, representing the area of strips, as shown in [\[link\]](#). In [\[link\]](#), since $F = -kx$ is a straight line with slope $-k$, when plotted versus x , the “area” under the line is just an algebraic combination of triangular “areas,” where “areas” above the x -axis are positive and those below are negative, as shown in [\[link\]](#). The magnitude of one of these “areas” is just one-half the triangle’s base, along the x -axis, times the triangle’s height, along the force axis. (There are quotation marks around “area” because this base-height product has the units of work, rather than square meters.)



A curve of $f(x)$ versus x showing the area of an infinitesimal strip, $f(x)dx$, and the sum of such areas, which is the integral of $f(x)$ from x_1 to x_2 .



Curve of the spring force $f(x) = -kx$ versus x , showing areas under the line, between x_A and x_B , for both positive and negative values of x_A . When x_A is negative, the total area under the curve for the integral in [\[link\]](#) is the sum of positive and negative triangular areas. When x_A is positive, the total area under the curve is the difference between two negative triangles.

Example:

Work Done by a Spring Force

A perfectly elastic spring requires 0.54 J of work to stretch 6 cm from its equilibrium position, as in [\[link\]\(b\)](#). (a) What is its spring constant k ? (b) How much work is required to stretch it an additional 6 cm?

Strategy

Work “required” means work done against the spring force, which is the negative of the work in [\[link\]](#), that is

Equation:

$$W = \frac{1}{2}k(x_B^2 - x_A^2).$$

For part (a), $x_A = 0$ and $x_B = 6\text{cm}$; for part (b), $x_B = 6\text{cm}$ and $x_B = 12\text{cm}$. In part (a), the work is given and you can solve for the spring constant; in part (b), you can use the value of k , from part (a), to solve for the work.

Solution

a. $W = 0.54\text{ J} = \frac{1}{2}k[(6\text{ cm})^2 - 0]$, so $k = 3\text{ N/cm}$.

$$\text{b. } W = \frac{1}{2}(3 \text{ N/cm})[(12 \text{ cm})^2 - (6 \text{ cm})^2] = 1.62 \text{ J.}$$

Significance

Since the work done by a spring force is independent of the path, you only needed to calculate the difference in the quantity $\frac{1}{2}kx^2$ at the end points. Notice that the work required to stretch the spring from 0 to 12 cm is four times that required to stretch it from 0 to 6 cm, because that work depends on the square of the amount of stretch from equilibrium, $\frac{1}{2}kx^2$. In this circumstance, the work to stretch the spring from 0 to 12 cm is also equal to the work for a composite path from 0 to 6 cm followed by an additional stretch from 6 cm to 12 cm. Therefore, $4W(0 \text{ cm to } 6 \text{ cm}) = W(0 \text{ cm to } 6 \text{ cm}) + W(6 \text{ cm to } 12 \text{ cm})$, or $W(6 \text{ cm to } 12 \text{ cm}) = 3W(0 \text{ cm to } 6 \text{ cm})$, as we found above.

Note:

Exercise:

Problem:

Check Your Understanding The spring in [\[link\]](#) is compressed 6 cm from its equilibrium length. (a) Does the spring force do positive or negative work and (b) what is the magnitude?

Solution:

a. The spring force is the opposite direction to a compression (as it is for an extension), so the work it does is negative. b. The work done depends on the square of the displacement, which is the same for $x = \pm 6 \text{ cm}$, so the magnitude is 0.54 J.

Summary

- The infinitesimal increment of work done by a force, acting over an infinitesimal displacement, is the dot product of the force and the displacement.
- The work done by a force, acting over a finite path, is the integral of the infinitesimal increments of work done along the path.
- The work done *against* a force is the negative of the work done *by* the force.
- The work done by a normal or frictional contact force must be determined in each particular case.
- The work done by the force of gravity, on an object near the surface of Earth, depends only on the weight of the object and the difference in height through which it moved.
- The work done by a spring force, acting from an initial position to a final position, depends only on the spring constant and the squares of those positions.

Conceptual Questions

Exercise:

Problem:

Give an example of something we think of as work in everyday circumstances that is not work in the scientific sense. Is energy transferred or changed in form in your example? If so, explain how this is accomplished without doing work.

Solution:

When you push on the wall, this “feels” like work; however, there is no displacement so there is no physical work. Energy is consumed, but no energy is transferred.

Exercise:**Problem:**

Give an example of a situation in which there is a force and a displacement, but the force does no work. Explain why it does no work.

Exercise:**Problem:**

Describe a situation in which a force is exerted for a long time but does no work. Explain.

Solution:

If you continue to push on a wall without breaking through the wall, you continue to exert a force with no displacement, so no work is done.

Exercise:**Problem:**

A body moves in a circle at constant speed. Does the centripetal force that accelerates the body do any work? Explain.

Exercise:**Problem:**

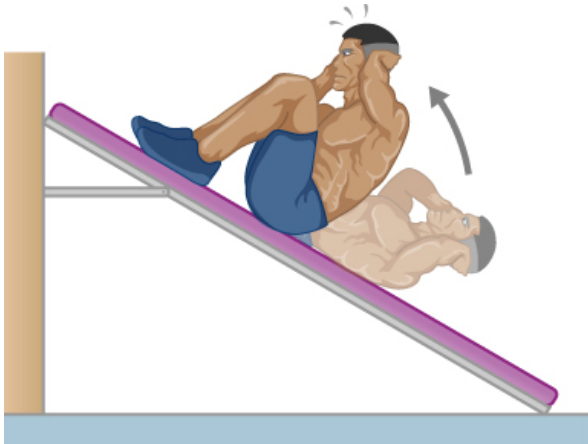
Suppose you throw a ball upward and catch it when it returns at the same height. How much work does the gravitational force do on the ball over its entire trip?

Solution:

The total displacement of the ball is zero, so no work is done.

Exercise:**Problem:**

Why is it more difficult to do sit-ups while on a slant board than on a horizontal surface? (See below.)



Exercise:

Problem:

As a young man, Tarzan climbed up a vine to reach his tree house. As he got older, he decided to build and use a staircase instead. Since the work of the gravitational force mg is path independent, what did the King of the Apes gain in using stairs?

Solution:

Both require the same gravitational work, but the stairs allow Tarzan to take this work over a longer time interval and hence gradually exert his energy, rather than dramatically by climbing a vine.

Problems

Exercise:

Problem:

How much work does a supermarket checkout attendant do on a can of soup he pushes 0.600 m horizontally with a force of 5.00 N?

Solution:

3.00 J

Exercise:**Problem:**

A 75.0-kg person climbs stairs, gaining 2.50 m in height. Find the work done to accomplish this task.

Exercise:**Problem:**

(a) Calculate the work done on a 1500-kg elevator car by its cable to lift it 40.0 m at constant speed, assuming friction averages 100 N. (b) What is the work done on the lift by the gravitational force in this process? (c) What is the total work done on the lift?

Solution:

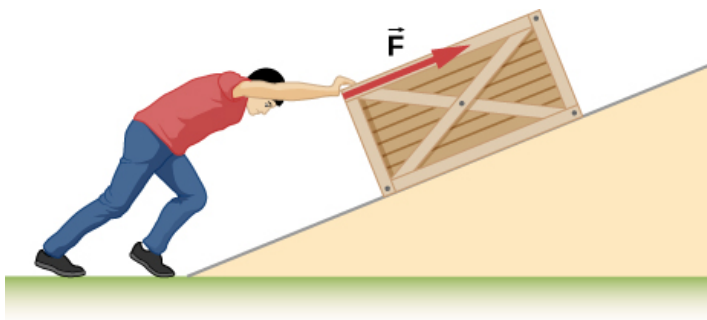
a. 593 kJ; b. -589 kJ; c. 0 J

Exercise:**Problem:**

Suppose a car travels 108 km at a speed of 30.0 m/s, and uses 2.0 gal of gasoline. Only 30% of the gasoline goes into useful work by the force that keeps the car moving at constant speed despite friction. (The energy content of gasoline is about 140 MJ/gal.) (a) What is the magnitude of the force exerted to keep the car moving at constant speed? (b) If the required force is directly proportional to speed, how many gallons will be used to drive 108 km at a speed of 28.0 m/s?

Exercise:**Problem:**

Calculate the work done by an 85.0-kg man who pushes a crate 4.00 m up along a ramp that makes an angle of 20.0° with the horizontal (see below). He exerts a force of 500 N on the crate parallel to the ramp and moves at a constant speed. Be certain to include the work he does on the crate and on his body to get up the ramp.

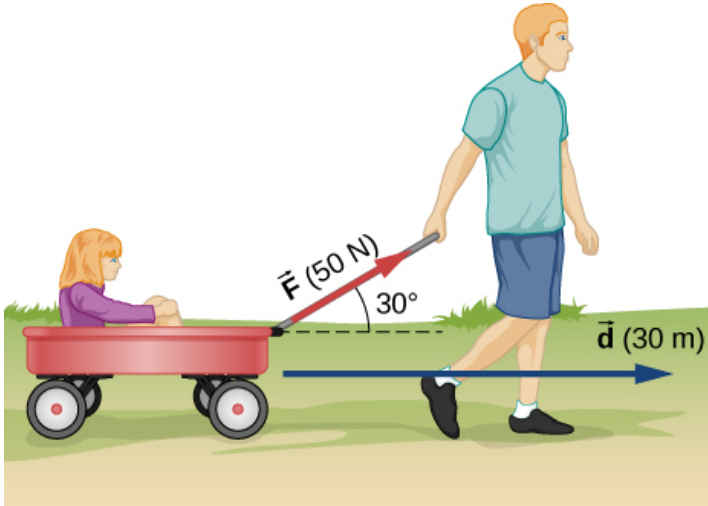
**Solution:**

3.14 kJ

Exercise:

Problem:

How much work is done by the boy pulling his sister 30.0 m in a wagon as shown below? Assume no friction acts on the wagon.



Exercise:

Problem:

A shopper pushes a grocery cart 20.0 m at constant speed on level ground, against a 35.0 N frictional force. He pushes in a direction 25.0° below the horizontal. (a) What is the work done on the cart by friction? (b) What is the work done on the cart by the gravitational force? (c) What is the work done on the cart by the shopper? (d) Find the force the shopper exerts, using energy considerations. (e) What is the total work done on the cart?

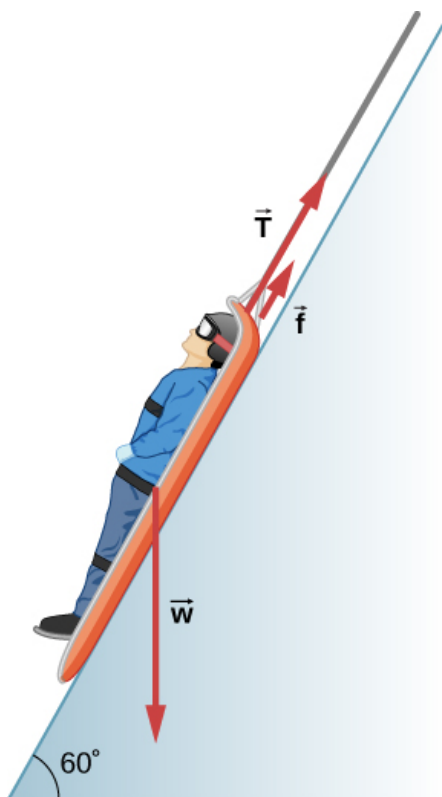
Solution:

a. -700 J ; b. 0 J ; c. 700 J ; d. 38.6 N ; e. 0 J

Exercise:

Problem:

Suppose the ski patrol lowers a rescue sled and victim, having a total mass of 90.0 kg, down a 60.0° slope at constant speed, as shown below. The coefficient of friction between the sled and the snow is 0.100. (a) How much work is done by friction as the sled moves 30.0 m along the hill? (b) How much work is done by the rope on the sled in this distance? (c) What is the work done by the gravitational force on the sled? (d) What is the total work done?



Exercise:

Problem:

A constant 20-N force pushes a small ball in the direction of the force over a distance of 5.0 m. What is the work done by the force?

Solution:

100 J

Exercise:

Problem:

A toy cart is pulled a distance of 6.0 m in a straight line across the floor. The force pulling the cart has a magnitude of 20 N and is directed at 37° above the horizontal. What is the work done by this force?

Exercise:

Problem:

A 5.0-kg box rests on a horizontal surface. The coefficient of kinetic friction between the box and surface is $\mu_K = 0.50$. A horizontal force pulls the box at constant velocity for 10 cm. Find the work done by (a) the applied horizontal force, (b) the frictional force, and (c) the net force.

Solution:

a. 2.45 J; b. -2.45 J; c. 0 J

Exercise:

Problem:

A sled plus passenger with total mass 50 kg is pulled 20 m across the snow ($\mu_k = 0.20$) at constant velocity by a force directed 25° above the horizontal. Calculate (a) the work of the applied force, (b) the work of friction, and (c) the total work.

Exercise:

Problem:

Suppose that the sled plus passenger of the preceding problem is pushed 20 m across the snow at constant velocity by a force directed 30° below the horizontal. Calculate (a) the work of the applied force, (b) the work of friction, and (c) the total work.

Solution:

a. 2.22 kJ; b. -2.22 kJ; c. 0 J

Exercise:

Problem:

How much work does the force $F(x) = (-2.0/x)$ N do on a particle as it moves from $x = 2.0$ m to $x = 5.0$ m?

Exercise:

Problem:

How much work is done against the gravitational force on a 5.0-kg briefcase when it is carried from the ground floor to the roof of the Empire State Building, a vertical climb of 380 m?

Solution:

18.6 kJ

Exercise:

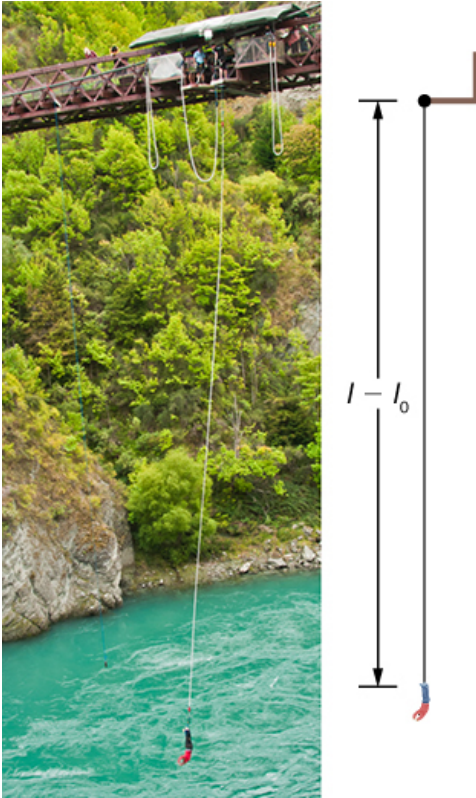
Problem:

It takes 500 J of work to compress a spring 10 cm. What is the force constant of the spring?

Exercise:

Problem:

A bungee cord is essentially a very long rubber band that can stretch up to four times its unstretched length. However, its spring constant varies over its stretch [see Menz, P.G. "The Physics of Bungee Jumping." *The Physics Teacher* (November 1993) 31: 483-487]. Take the length of the cord to be along the x -direction and define the stretch x as the length of the cord l minus its un-stretched length l_0 ; that is, $x = l - l_0$ (see below). Suppose a particular bungee cord has a spring constant, for $0 \leq x \leq 4.88$ m, of $k_1 = 204$ N/m and for $x \geq 4.88$ m, of $k_2 = 111$ N/m. (Recall that the spring constant is the slope of the force $F(x)$ versus its stretch x .) (a) What is the tension in the cord when the stretch is 16.7 m (the maximum desired for a given jump)? (b) How much work must be done against the elastic force of the bungee cord to stretch it 16.7 m?



(credit: modification of work by
Graeme Churchard)

Solution:

a. 2.32 kN; b. 22.0 kJ

Exercise:

Problem:

A bungee cord exerts a nonlinear elastic force of magnitude $F(x) = k_1x + k_2x^3$, where x is the distance the cord is stretched, $k_1 = 204 \text{ N/m}$ and $k_2 = -0.233 \text{ N/m}^3$. How much work must be done on the cord to stretch it 16.7 m?

Exercise:**Problem:**

Engineers desire to model the magnitude of the elastic force of a bungee cord using the equation

$$F(x) = a \left[\frac{x+9 \text{ m}}{9 \text{ m}} - \left(\frac{9 \text{ m}}{x+9 \text{ m}} \right)^2 \right],$$

where x is the stretch of the cord along its length and a is a constant. If it takes 22.0 kJ of work to stretch the cord by 16.7 m, determine the value of the constant a .

Solution:

835 N

Exercise:**Problem:**

A particle moving in the xy -plane is subject to a force

$$\vec{F}(x, y) = (50 \text{ N/m}) \left(x\hat{\mathbf{i}} + \frac{y^2}{3\text{m}}\hat{\mathbf{j}} \right)$$

where x and y are in meters. Calculate the work done on the particle by this force, as it moves in a straight line from the point (3 m, 4 m) to the point (6 m, 8 m).

Exercise:**Problem:**

A particle moves along a curved path $y(x) = (10 \text{ m})\{1 + \cos[(0.1 \text{ m}^{-1})x]\}$, from $x = 0$ to $x = 10\pi \text{ m}$, subject to a tangential force of variable magnitude $F(x) = (10 \text{ N})\sin[(0.1 \text{ m}^{-1})x]$. How much work does the force do? (*Hint: Consult a table of integrals or use a numerical integration program.*)

Solution:

257 J

Glossary**work**

done when a force acts on something that undergoes a displacement from one position to another

Kinetic Energy

By the end of this section, you will be able to:

- Calculate the kinetic energy of a particle given its mass and its velocity or momentum
- Evaluate the kinetic energy of a body, relative to different frames of reference

It's plausible to suppose that the greater the velocity of a body, the greater effect it could have on other bodies. This does not depend on the direction of the velocity, only its magnitude. At the end of the seventeenth century, a quantity was introduced into mechanics to explain collisions between two perfectly elastic bodies, in which one body makes a head-on collision with an identical body at rest. The first body stops, and the second body moves off with the initial velocity of the first body. (If you have ever played billiards or croquet, or seen a model of Newton's Cradle, you have observed this type of collision.) The idea behind this quantity was related to the forces acting on a body and was referred to as "the energy of motion." Later on, during the eighteenth century, the name **kinetic energy** was given to energy of motion.

With this history in mind, we can now state the classical definition of kinetic energy. Note that when we say "classical," we mean non-relativistic, that is, at speeds much less than the speed of light. At speeds comparable to the speed of light, the special theory of relativity requires a different expression for the kinetic energy of a particle, as discussed in [Relativity](#).

Since objects (or systems) of interest vary in complexity, we first define the kinetic energy of a particle with mass m .

Note:

Kinetic Energy

The kinetic energy of a particle is one-half the product of the particle's mass m and the square of its speed v :

Equation:

$$K = \frac{1}{2}mv^2.$$

We then extend this definition to any system of particles by adding up the kinetic energies of all the constituent particles:

Equation:

$$K = \sum \frac{1}{2}mv^2.$$

Note that just as we can express Newton's second law in terms of either the rate of change of momentum or mass times the rate of change of velocity, so the kinetic energy of a particle can be expressed in terms of its mass and momentum ($\vec{p} = m\vec{v}$), instead of its mass and velocity. Since $v = p/m$, we see that

Equation:

$$K = \frac{1}{2}m\left(\frac{p}{m}\right)^2 = \frac{p^2}{2m}$$

also expresses the kinetic energy of a single particle. Sometimes, this expression is more convenient to use than [\[link\]](#).

The units of kinetic energy are mass times the square of speed, or $\text{kg} \cdot \text{m}^2/\text{s}^2$. But the units of force are mass times acceleration, $\text{kg} \cdot \text{m}/\text{s}^2$, so the units of kinetic energy are also the units of force times distance, which are the units of work, or joules. You will see in the next section that work and kinetic energy have the same units, because they are different forms of the same, more general, physical property.

Example:

Kinetic Energy of an Object

(a) What is the kinetic energy of an 80-kg athlete, running at 10 m/s? (b) The Chicxulub crater in Yucatan, one of the largest existing impact craters on Earth, is thought to have been created by an asteroid, traveling at 22 km/s and releasing 4.2×10^{23} J of kinetic energy upon impact. What was its mass? (c) In nuclear reactors, thermal neutrons, traveling at about 2.2 km/s, play an important role. What is the kinetic energy of such a particle?

Strategy

To answer these questions, you can use the definition of kinetic energy in [\[link\]](#). You also have to look up the mass of a neutron.

Solution

Don't forget to convert km into m to do these calculations, although, to save space, we omitted showing these conversions.

$$\text{a. } K = \frac{1}{2}(80 \text{ kg})(10 \text{ m/s})^2 = 4.0 \text{ kJ.}$$

$$\text{b. } m = 2K/v^2 = 2(4.2 \times 10^{23} \text{ J})/(22 \text{ km/s})^2 = 1.7 \times 10^{15} \text{ kg.}$$

$$\text{c. } K = \frac{1}{2}(1.68 \times 10^{-27} \text{ kg})(2.2 \text{ km/s})^2 = 4.1 \times 10^{-21} \text{ J.}$$

Significance

In this example, we used the way mass and speed are related to kinetic energy, and we encountered a very wide range of values for the kinetic energies. Different units are commonly used for such very large and very small values. The energy of the impactor in part (b) can be compared to the explosive yield of TNT and nuclear explosions,

1 megaton = 4.18×10^{15} J. The Chicxulub asteroid's kinetic energy was about a hundred million megatons. At the other extreme, the energy of subatomic particle is expressed in electron-volts, $1 \text{ eV} = 1.6 \times 10^{-19}$ J. The thermal neutron in part (c) has a kinetic energy of about one fortieth of an electron-volt.

Note:

Exercise:

Problem:

Check Your Understanding (a) A car and a truck are each moving with the same kinetic energy. Assume that the truck has more mass than the car. Which has the greater speed? (b) A car and a truck are each moving with the same speed. Which has the greater kinetic energy?

Solution:

a. the car; b. the truck

Because velocity is a relative quantity, you can see that the value of kinetic energy must depend on your frame of reference. You can generally choose a frame of reference that is suited to the purpose of your analysis and that simplifies your calculations. One such frame of reference is the one in which the observations of the system are made (likely an external frame). Another choice is a frame that is attached to, or moves with, the system (likely an internal frame). The equations for relative motion, discussed in [Motion in Two and Three Dimensions](#), provide a link to calculating the kinetic energy of an object with respect to different frames of reference.

Example:**Kinetic Energy Relative to Different Frames**

A 75.0-kg person walks down the central aisle of a subway car at a speed of 1.50 m/s relative to the car, whereas the train is moving at 15.0 m/s relative to the tracks. (a) What is the person's kinetic energy relative to the car? (b) What is the person's kinetic energy relative to the tracks? (c) What is the person's kinetic energy relative to a frame moving with the person?

Strategy

Since speeds are given, we can use $\frac{1}{2}mv^2$ to calculate the person's kinetic energy. However, in part (a), the person's speed is relative to the subway car (as given); in part (b), it is relative to the tracks; and in part (c), it is

zero. If we denote the car frame by C, the track frame by T, and the person by P, the relative velocities in part (b) are related by $\vec{v}_{PT} = \vec{v}_{PC} + \vec{v}_{CT}$. We can assume that the central aisle and the tracks lie along the same line, but the direction the person is walking relative to the car isn't specified, so we will give an answer for each possibility, $v_{PT} = v_{CT} \pm v_{PC}$, as shown in [\[link\]](#).



The possible motions of a person walking in a train are (a) toward the front of the car and (b) toward the back of the car.

Solution

a. $K = \frac{1}{2} (75.0 \text{ kg})(1.50 \text{ m/s})^2 = 84.4 \text{ J}.$

b. $v_{PT} = (15.0 \pm 1.50) \text{ m/s}.$ Therefore, the two possible values for kinetic energy relative to the car are

Equation:

$$K = \frac{1}{2} (75.0 \text{ kg})(13.5 \text{ m/s})^2 = 6.83 \text{ kJ}$$

and

Equation:

$$K = \frac{1}{2} (75.0 \text{ kg})(16.5 \text{ m/s})^2 = 10.2 \text{ kJ}.$$

c. In a frame where $v_P = 0$, $K = 0$ as well.

Significance

You can see that the kinetic energy of an object can have very different values, depending on the frame of reference. However, the kinetic energy

of an object can never be negative, since it is the product of the mass and the square of the speed, both of which are always positive or zero.

Note:

Exercise:

Problem:

Check Your Understanding You are rowing a boat parallel to the banks of a river. Your kinetic energy relative to the banks is less than your kinetic energy relative to the water. Are you rowing with or against the current?

Solution:

against

The kinetic energy of a particle is a single quantity, but the kinetic energy of a system of particles can sometimes be divided into various types, depending on the system and its motion. For example, if all the particles in a system have the same velocity, the system is undergoing translational motion and has translational kinetic energy. If an object is rotating, it could have rotational kinetic energy, or if it's vibrating, it could have vibrational kinetic energy. The kinetic energy of a system, relative to an internal frame of reference, may be called internal kinetic energy. The kinetic energy associated with random molecular motion may be called thermal energy. These names will be used in later chapters of the book, when appropriate. Regardless of the name, every kind of kinetic energy is the same physical quantity, representing energy associated with motion.

Example:

Special Names for Kinetic Energy

(a) A player lobs a mid-court pass with a 624-g basketball, which covers 15 m in 2 s. What is the basketball's horizontal translational kinetic energy while in flight? (b) An average molecule of air, in the basketball in part (a), has a mass of 29 u, and an average speed of 500 m/s, relative to the basketball. There are about 3×10^{23} molecules inside it, moving in random directions, when the ball is properly inflated. What is the average translational kinetic energy of the random motion of all the molecules inside, relative to the basketball? (c) How fast would the basketball have to travel relative to the court, as in part (a), so as to have a kinetic energy equal to the amount in part (b)?

Strategy

In part (a), first find the horizontal speed of the basketball and then use the definition of kinetic energy in terms of mass and speed, $K = \frac{1}{2}mv^2$. Then in part (b), convert unified units to kilograms and then use $K = \frac{1}{2}mv^2$ to get the average translational kinetic energy of one molecule, relative to the basketball. Then multiply by the number of molecules to get the total result. Finally, in part (c), we can substitute the amount of kinetic energy in part (b), and the mass of the basketball in part (a), into the definition $K = \frac{1}{2}mv^2$, and solve for v .

Solution

- a. The horizontal speed is (15 m)/(2 s), so the horizontal kinetic energy of the basketball is

Equation:

$$\frac{1}{2}(0.624 \text{ kg})(7.5 \text{ m/s})^2 = 17.6 \text{ J}.$$

- b. The average translational kinetic energy of a molecule is

Equation:

$$\frac{1}{2}(29 \text{ u})(1.66 \times 10^{-27} \text{ kg/u})(500 \text{ m/s})^2 = 6.02 \times 10^{-21} \text{ J},$$

and the total kinetic energy of all the molecules is

Equation:

$$(3 \times 10^{23})(6.02 \times 10^{-21} \text{ J}) = 1.80 \text{ kJ}.$$

$$\text{c. } v = \sqrt{2(1.8 \text{ kJ})/(0.624 \text{ kg})} = 76.0 \text{ m/s}.$$

Significance

In part (a), this kind of kinetic energy can be called the horizontal kinetic energy of an object (the basketball), relative to its surroundings (the court). If the basketball were spinning, all parts of it would have not just the average speed, but it would also have rotational kinetic energy. Part (b) reminds us that this kind of kinetic energy can be called internal or thermal kinetic energy. Notice that this energy is about a hundred times the energy in part (a). How to make use of thermal energy will be the subject of the chapters on thermodynamics. In part (c), since the energy in part (b) is about 100 times that in part (a), the speed should be about 10 times as big, which it is (76 compared to 7.5 m/s).

Summary

- The kinetic energy of a particle is the product of one-half its mass and the square of its speed, for non-relativistic speeds.
- The kinetic energy of a system is the sum of the kinetic energies of all the particles in the system.
- Kinetic energy is relative to a frame of reference, is always positive, and is sometimes given special names for different types of motion.

Conceptual Questions

Exercise:

Problem:

A particle of m has a velocity of $v_x \hat{\mathbf{i}} + v_y \hat{\mathbf{j}} + v_z \hat{\mathbf{k}}$. Is its kinetic energy given by $m(v_x^2 \hat{\mathbf{i}} + v_y^2 \hat{\mathbf{j}} + v_z^2 \hat{\mathbf{k}})/2$? If not, what is the correct expression?

Exercise:**Problem:**

One particle has mass m and a second particle has mass $2m$. The second particle is moving with speed v and the first with speed $2v$. How do their kinetic energies compare?

Solution:

The first particle has a kinetic energy of $4(\frac{1}{2}mv^2)$ whereas the second particle has a kinetic energy of $2(\frac{1}{2}mv^2)$, so the first particle has twice the kinetic energy of the second particle.

Exercise:**Problem:**

A person drops a pebble of mass m_1 from a height h , and it hits the floor with kinetic energy K . The person drops another pebble of mass m_2 from a height of $2h$, and it hits the floor with the same kinetic energy K . How do the masses of the pebbles compare?

Problems**Exercise:****Problem:**

Compare the kinetic energy of a 20,000-kg truck moving at 110 km/h with that of an 80.0-kg astronaut in orbit moving at 27,500 km/h.

Exercise:**Problem:**

(a) How fast must a 3000-kg elephant move to have the same kinetic energy as a 65.0-kg sprinter running at 10.0 m/s? (b) Discuss how the larger energies needed for the movement of larger animals would relate to metabolic rates.

Solution:

a. 1.47 m/s; b. answers may vary

Exercise:**Problem:**

Estimate the kinetic energy of a 90,000-ton aircraft carrier moving at a speed of at 30 knots. You will need to look up the definition of a nautical mile to use in converting the unit for speed, where 1 knot equals 1 nautical mile per hour. Furthermore for this problem, 1 ton is equivalent to 2,000 pounds.

Exercise:**Problem:**

Calculate the kinetic energies of (a) a 2000.0-kg automobile moving at 100.0 km/h; (b) an 80.-kg runner sprinting at 10. m/s; and (c) a 9.1×10^{-31} -kg electron moving at 2.0×10^7 m/s.

Solution:

a. 772 kJ; b. 4.0 kJ; c. 1.8×10^{-16} J

Exercise:**Problem:**

A 5.0-kg body has three times the kinetic energy of an 8.0-kg body. Calculate the ratio of the speeds of these bodies.

Exercise:**Problem:**

An 8.0-g bullet has a speed of 800 m/s. (a) What is its kinetic energy? (b) What is its kinetic energy if the speed is halved?

Solution:

a. 2.6 kJ; b. 640 J

Glossary

kinetic energy

energy of motion, one-half an object's mass times the square of its speed

Work-Energy Theorem

By the end of this section, you will be able to:

- Apply the work-energy theorem to find information about the motion of a particle, given the forces acting on it
- Use the work-energy theorem to find information about the forces acting on a particle, given information about its motion

We have discussed how to find the work done on a particle by the forces that act on it, but how is that work manifested in the motion of the particle? According to Newton's second law of motion, the sum of all the forces acting on a particle, or the net force, determines the rate of change in the momentum of the particle, or its motion. Therefore, we should consider the work done by all the forces acting on a particle, or the **net work**, to see what effect it has on the particle's motion.

Let's start by looking at the net work done on a particle as it moves over an infinitesimal displacement, which is the dot product of the net force and the displacement: $dW_{\text{net}} = \vec{\mathbf{F}}_{\text{net}} \cdot d\vec{\mathbf{r}}$. Newton's second law tells us that $\vec{\mathbf{F}}_{\text{net}} = m(d\vec{\mathbf{v}}/dt)$, so $dW_{\text{net}} = m(d\vec{\mathbf{v}}/dt) \cdot d\vec{\mathbf{r}}$. For the mathematical functions describing the motion of a physical particle, we can rearrange the differentials dt , etc., as algebraic quantities in this expression, that is,

Equation:

$$dW_{\text{net}} = m \left(\frac{d\vec{\mathbf{v}}}{dt} \right) \cdot d\vec{\mathbf{r}} = m d\vec{\mathbf{v}} \cdot \left(\frac{d\vec{\mathbf{r}}}{dt} \right) = m \vec{\mathbf{v}} \cdot d\vec{\mathbf{v}},$$

where we substituted the velocity for the time derivative of the displacement and used the commutative property of the dot product [[link](#)]. Since derivatives and integrals of scalars are probably more familiar to you at this point, we express the dot product in terms of Cartesian coordinates before we integrate between any two points A and B on the particle's trajectory. This gives us the net work done on the particle:

Equation:

$$\begin{aligned}
 W_{\text{net},AB} &= \int_A^B (mv_x dv_x + mv_y dv_y + mv_z dv_z) \\
 &= \frac{1}{2}m \left| v_x^2 + v_y^2 + v_z^2 \right|_A^B = \left| \frac{1}{2}mv^2 \right|_A^B = K_B - K_A.
 \end{aligned}$$

In the middle step, we used the fact that the square of the velocity is the sum of the squares of its Cartesian components, and in the last step, we used the definition of the particle's kinetic energy. This important result is called the **work-energy theorem** ([\[link\]](#)).

Note:

Work-Energy Theorem

The net work done on a particle equals the change in the particle's kinetic energy:

Equation:

$$W_{\text{net}} = K_B - K_A.$$



Horse pulls are common events at state fairs. The work done by the horses pulling on the load results in a change in kinetic energy of the load, ultimately going faster. (credit: modification of work by “Jassen”/ Flickr)

According to this theorem, when an object slows down, its final kinetic energy is less than its initial kinetic energy, the change in its kinetic energy is negative, and so is the net work done on it. If an object speeds up, the net work done on it is positive. When calculating the net work, you must include all the forces that act on an object. If you leave out any forces that act on an object, or if you include any forces that don't act on it, you will get a wrong result.

The importance of the work-energy theorem, and the further generalizations to which it leads, is that it makes some types of calculations much simpler to accomplish than they would be by trying to solve Newton's second law.

For example, in [Newton's Laws of Motion](#), we found the speed of an object sliding down a frictionless plane by solving Newton's second law for the acceleration and using kinematic equations for constant acceleration, obtaining

Equation:

$$v_f^2 = v_i^2 + 2g(s_f - s_i)\sin \theta,$$

where s is the displacement down the plane.

We can also get this result from the work-energy theorem in [\[link\]](#). Since only two forces are acting on the object-gravity and the normal force-and the normal force doesn't do any work, the net work is just the work done by gravity. The work dW is the dot product of the force of gravity or

$\vec{F} = -mg\hat{j}$ and the displacement $\vec{dr} = dx\hat{i} + dy\hat{j}$. After taking the dot product and integrating from an initial position y_i to a final position y_f , one finds the net work as

Equation:

$$W_{\text{net}} = W_{\text{grav}} = -mg(y_f - y_i),$$

where y is positive up. The work-energy theorem says that this equals the change in kinetic energy:

Equation:

$$-mg(y_f - y_i) = \frac{1}{2}m(v_f^2 - v_i^2).$$

Using a right triangle, we can see that $(y_f - y_i) = (s_f - s_i)\sin \theta$, so the result for the final speed is the same.

What is gained by using the work-energy theorem? The answer is that for a frictionless plane surface, not much. However, Newton's second law is easy to solve only for this particular case, whereas the work-energy theorem gives the final speed for any shaped frictionless surface. For an arbitrary

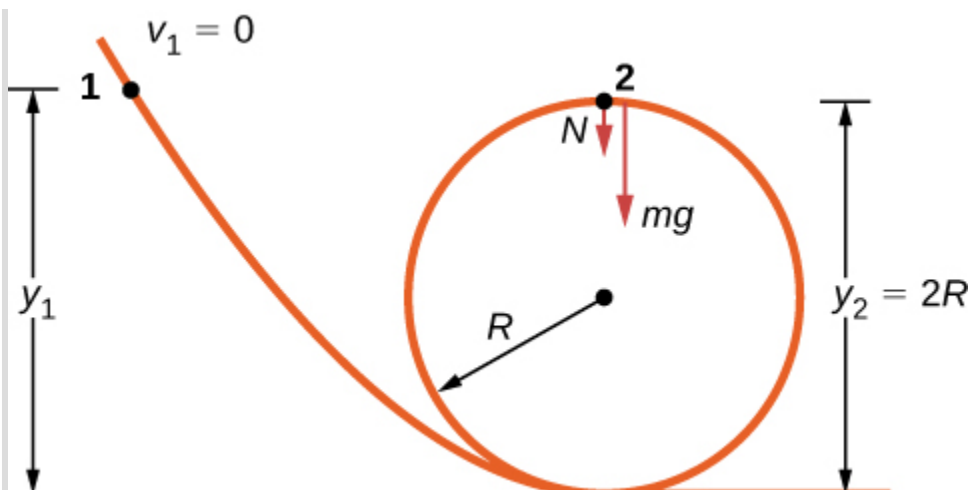
curved surface, the normal force is not constant, and Newton's second law may be difficult or impossible to solve analytically. Constant or not, for motion along a surface, the normal force never does any work, because it's perpendicular to the displacement. A calculation using the work-energy theorem avoids this difficulty and applies to more general situations.

Note:**Work-Energy Theorem**

1. Draw a free-body diagram for each force on the object.
2. Determine whether or not each force does work over the displacement in the diagram. Be sure to keep any positive or negative signs in the work done.
3. Add up the total amount of work done by each force.
4. Set this total work equal to the change in kinetic energy and solve for any unknown parameter.
5. Check your answers. If the object is traveling at a constant speed or zero acceleration, the total work done should be zero and match the change in kinetic energy. If the total work is positive, the object must have sped up or increased kinetic energy. If the total work is negative, the object must have slowed down or decreased kinetic energy.

Example:**Loop-the-Loop**

The frictionless track for a toy car includes a loop-the-loop of radius R . How high, measured from the bottom of the loop, must the car be placed to start from rest on the approaching section of track and go all the way around the loop?



A frictionless track for a toy car has a loop-the-loop in it. How high must the car start so that it can go around the loop without falling off?

Strategy

The free-body diagram at the final position of the object is drawn in [\[link\]](#). The gravitational work is the only work done over the displacement that is not zero. Since the weight points in the same direction as the net vertical displacement, the total work done by the gravitational force is positive. From the work-energy theorem, the starting height determines the speed of the car at the top of the loop,

Equation:

$$-mg(y_2 - y_1) = \frac{1}{2}mv_2^2,$$

where the notation is shown in the accompanying figure. At the top of the loop, the normal force and gravity are both down and the acceleration is centripetal, so

Equation:

$$a_{\text{top}} = \frac{F}{m} = \frac{N + mg}{m} = \frac{v_2^2}{R}.$$

The condition for maintaining contact with the track is that there must be some normal force, however slight; that is, $N > 0$. Substituting for v_2^2 and N , we can find the condition for y_1 .

Solution

Implement the steps in the strategy to arrive at the desired result:

Equation:

$$N = -mg + \frac{mv_2^2}{R} = \frac{-mgR + 2mg(y_1 - R)}{R} > 0 \quad \text{or} \quad y_1 > \frac{5R}{2}.$$

Significance

On the surface of the loop, the normal component of gravity and the normal contact force must provide the centripetal acceleration of the car going around the loop. The tangential component of gravity slows down or speeds up the car. A child would find out how high to start the car by trial and error, but now that you know the work-energy theorem, you can predict the minimum height (as well as other more useful results) from physical principles. By using the work-energy theorem, you did not have to solve a differential equation to determine the height.

Note:

Exercise:

Problem:

Check Your Understanding Suppose the radius of the loop-the-loop in [\[link\]](#) is 15 cm and the toy car starts from rest at a height of 45 cm above the bottom. What is its speed at the top of the loop?

Solution:

$$\sqrt{3} \text{ m/s}$$

Note:

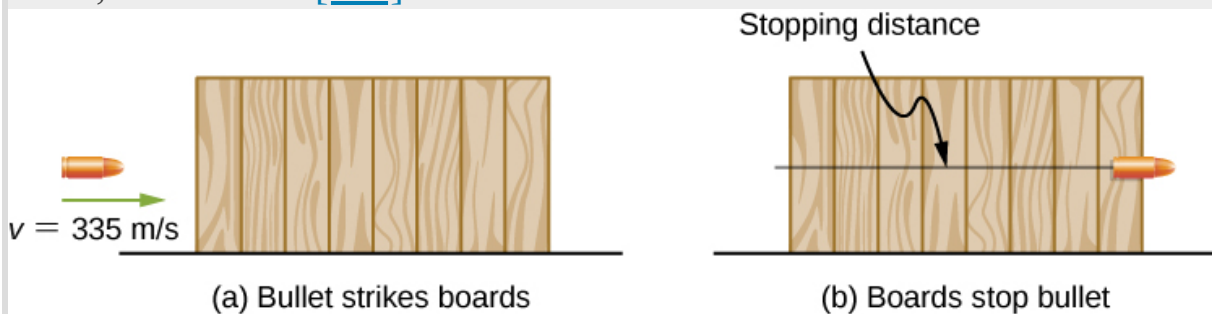
Visit Carleton College's site to see a [video](#) of a looping rollercoaster.

In situations where the motion of an object is known, but the values of one or more of the forces acting on it are not known, you may be able to use the work-energy theorem to get some information about the forces. Work depends on the force and the distance over which it acts, so the information is provided via their product.

Example:

Determining a Stopping Force

A bullet has a mass of 40 grains (2.60 g) and a muzzle velocity of 1100 ft./s (335 m/s). It can penetrate eight 1-inch pine boards, each with thickness 0.75 inches. What is the average stopping force exerted by the wood, as shown in [\[link\]](#)?



The boards exert a force to stop the bullet. As a result, the boards do work and the bullet loses kinetic energy.

Strategy

We can assume that under the general conditions stated, the bullet loses all its kinetic energy penetrating the boards, so the work-energy theorem says its initial kinetic energy is equal to the average stopping force times the distance penetrated. The change in the bullet's kinetic energy and the net work done stopping it are both negative, so when you write out the work-energy theorem, with the net work equal to the average force times the

stopping distance, that's what you get. The total thickness of eight 1-inch pine boards that the bullet penetrates is $8 \times \frac{3}{4} \text{ in.} = 6 \text{ in.} = 15.2 \text{ cm.}$

Solution

Applying the work-energy theorem, we get

Equation:

$$W_{\text{net}} = -F_{\text{ave}}\Delta s_{\text{stop}} = -K_{\text{initial}},$$

so

Equation:

$$F_{\text{ave}} = \frac{\frac{1}{2}mv^2}{\Delta s_{\text{stop}}} = \frac{\frac{1}{2}(2.6 \times 10^{-3}\text{kg})(335 \text{ m/s})^2}{0.152 \text{ m}} = 960 \text{ N.}$$

Significance

We could have used Newton's second law and kinematics in this example, but the work-energy theorem also supplies an answer to less simple situations. The penetration of a bullet, fired vertically upward into a block of wood, is discussed in one section of Asif Shakur's recent article ["Bullet-Block Science Video Puzzle." *The Physics Teacher* (January 2015) 53(1): 15-16]. If the bullet is fired dead center into the block, it loses all its kinetic energy and penetrates slightly farther than if fired off-center. The reason is that if the bullet hits off-center, it has a little kinetic energy after it stops penetrating, because the block rotates. The work-energy theorem implies that a smaller change in kinetic energy results in a smaller penetration. You will understand more of the physics in this interesting article after you finish reading [Angular Momentum](#).

Note:

Learn more about work and energy in this [PhET simulation](#) called "the ramp." Try changing the force pushing the box and the frictional force along the incline. The work and energy plots can be examined to note the total work done and change in kinetic energy of the box.

Summary

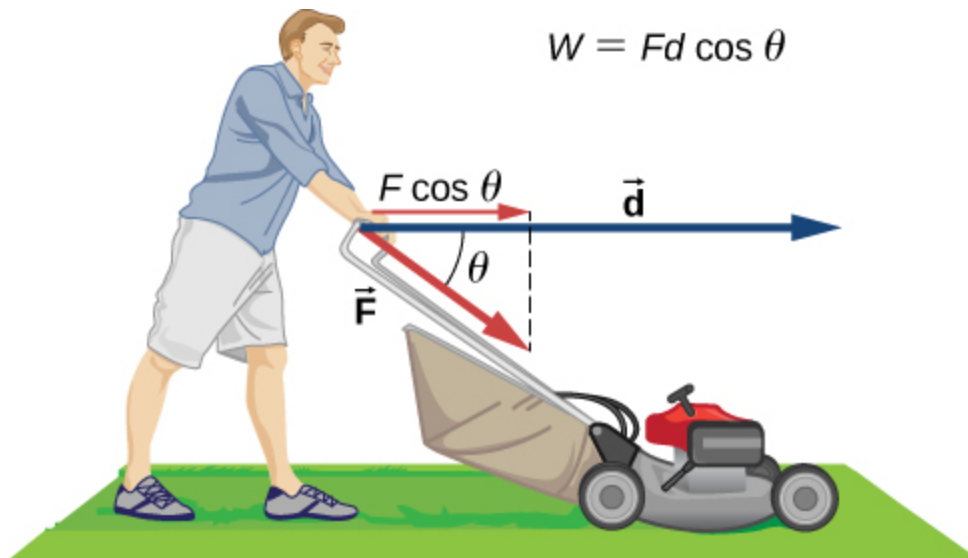
- Because the net force on a particle is equal to its mass times the derivative of its velocity, the integral for the net work done on the particle is equal to the change in the particle's kinetic energy. This is the work-energy theorem.
- You can use the work-energy theorem to find certain properties of a system, without having to solve the differential equation for Newton's second law.

Conceptual Questions

Exercise:

Problem:

The person shown below does work on the lawn mower. Under what conditions would the mower gain energy from the person pushing the mower? Under what conditions would it lose energy?



Solution:

The mower would gain energy if $-90^\circ < \theta < 90^\circ$. It would lose energy if $90^\circ < \theta < 270^\circ$. The mower may also lose energy due to

friction with the grass while pushing; however, we are not concerned with that energy loss for this problem.

Exercise:

Problem:

Work done on a system puts energy into it. Work done by a system removes energy from it. Give an example for each statement.

Exercise:

Problem:

Two marbles of masses m and $2m$ are dropped from a height h . Compare their kinetic energies when they reach the ground.

Solution:

The second marble has twice the kinetic energy of the first because kinetic energy is directly proportional to mass, like the work done by gravity.

Exercise:

Problem:

Compare the work required to accelerate a car of mass 2000 kg from 30.0 to 40.0 km/h with that required for an acceleration from 50.0 to 60.0 km/h.

Exercise:

Problem:

Suppose you are jogging at constant velocity. Are you doing any work on the environment and vice versa?

Solution:

Unless the environment is nearly frictionless, you are doing some positive work on the environment to cancel out the frictional work

against you, resulting in zero total work producing a constant velocity.

Exercise:

Problem:

Two forces act to double the speed of a particle, initially moving with kinetic energy of 1 J. One of the forces does 4 J of work. How much work does the other force do?

Problems

Exercise:

Problem:

(a) Calculate the force needed to bring a 950-kg car to rest from a speed of 90.0 km/h in a distance of 120 m (a fairly typical distance for a non-panic stop). (b) Suppose instead the car hits a concrete abutment at full speed and is brought to a stop in 2.00 m. Calculate the force exerted on the car and compare it with the force found in part (a).

Exercise:

Problem:

A car's bumper is designed to withstand a 4.0-km/h (1.1-m/s) collision with an immovable object without damage to the body of the car. The bumper cushions the shock by absorbing the force over a distance. Calculate the magnitude of the average force on a bumper that collapses 0.200 m while bringing a 900-kg car to rest from an initial speed of 1.1 m/s.

Solution:

2.72 kN

Exercise:

Problem:

Boxing gloves are padded to lessen the force of a blow. (a) Calculate the force exerted by a boxing glove on an opponent's face, if the glove and face compress 7.50 cm during a blow in which the 7.00-kg arm and glove are brought to rest from an initial speed of 10.0 m/s. (b) Calculate the force exerted by an identical blow in the days when no gloves were used, and the knuckles and face would compress only 2.00 cm. Assume the change in mass by removing the glove is negligible. (c) Discuss the magnitude of the force with glove on. Does it seem high enough to cause damage even though it is lower than the force with no glove?

Exercise:**Problem:**

Using energy considerations, calculate the average force a 60.0-kg sprinter exerts backward on the track to accelerate from 2.00 to 8.00 m/s in a distance of 25.0 m, if he encounters a headwind that exerts an average force of 30.0 N against him.

Solution:

102 N

Exercise:**Problem:**

A 5.0-kg box has an acceleration of 2.0 m/s^2 when it is pulled by a horizontal force across a surface with $\mu_K = 0.50$. Find the work done over a distance of 10 cm by (a) the horizontal force, (b) the frictional force, and (c) the net force. (d) What is the change in kinetic energy of the box?

Exercise:

Problem:

A constant 10-N horizontal force is applied to a 20-kg cart at rest on a level floor. If friction is negligible, what is the speed of the cart when it has been pushed 8.0 m?

Solution:

2.8 m/s

Exercise:**Problem:**

In the preceding problem, the 10-N force is applied at an angle of 45° below the horizontal. What is the speed of the cart when it has been pushed 8.0 m?

Exercise:**Problem:**

Compare the work required to stop a 100-kg crate sliding at 1.0 m/s and an 8.0-g bullet traveling at 500 m/s.

Solution:

$$W(\text{bullet}) = 20 \times W(\text{crate})$$

Exercise:**Problem:**

A wagon with its passenger sits at the top of a hill. The wagon is given a slight push and rolls 100 m down a 10° incline to the bottom of the hill. What is the wagon's speed when it reaches the end of the incline. Assume that the retarding force of friction is negligible.

Exercise:

Problem:

An 8.0-g bullet with a speed of 800 m/s is shot into a wooden block and penetrates 20 cm before stopping. What is the average force of the wood on the bullet? Assume the block does not move.

Solution:

12.8 kN

Exercise:**Problem:**

A 2.0-kg block starts with a speed of 10 m/s at the bottom of a plane inclined at 37° to the horizontal. The coefficient of sliding friction between the block and plane is $\mu_k = 0.30$. (a) Use the work-energy principle to determine how far the block slides along the plane before momentarily coming to rest. (b) After stopping, the block slides back down the plane. What is its speed when it reaches the bottom? (*Hint:* For the round trip, only the force of friction does work on the block.)

Exercise:**Problem:**

When a 3.0-kg block is pushed against a massless spring of force constant $4.5 \times 10^3 \text{ N/m}$, the spring is compressed 8.0 cm. The block is released, and it slides 2.0 m (from the point at which it is released) across a horizontal surface before friction stops it. What is the coefficient of kinetic friction between the block and the surface?

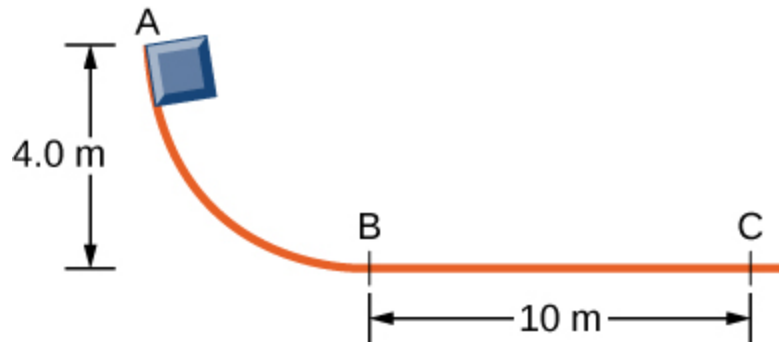
Solution:

0.25

Exercise:

Problem:

A small block of mass 200 g starts at rest at A, slides to B where its speed is $v_B = 8.0 \text{ m/s}$, then slides along the horizontal surface a distance 10 m before coming to rest at C. (See below.) (a) What is the work of friction along the curved surface? (b) What is the coefficient of kinetic friction along the horizontal surface?

**Exercise:****Problem:**

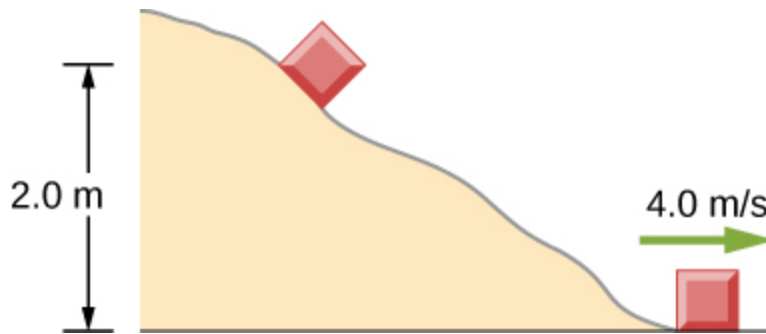
A small object is placed at the top of an incline that is essentially frictionless. The object slides down the incline onto a rough horizontal surface, where it stops in 5.0 s after traveling 60 m. (a) What is the speed of the object at the bottom of the incline and its acceleration along the horizontal surface? (b) What is the height of the incline?

Solution:

a. 24 m/s, -4.8 m/s^2 ; b. 29.4 m

Exercise:**Problem:**

When released, a 100-g block slides down the path shown below, reaching the bottom with a speed of 4.0 m/s. How much work does the force of friction do?



Exercise:

Problem:

A 0.22LR-caliber bullet like that mentioned in [\[link\]](#) is fired into a door made of a single thickness of 1-inch pine boards. How fast would the bullet be traveling after it penetrated through the door?

Solution:

310 m/s

Exercise:

Problem:

A sled starts from rest at the top of a snow-covered incline that makes a 22° angle with the horizontal. After sliding 75 m down the slope, its speed is 14 m/s. Use the work-energy theorem to calculate the coefficient of kinetic friction between the runners of the sled and the snowy surface.

Glossary

net work

work done by all the forces acting on an object

work-energy theorem

net work done on a particle is equal to the change in its kinetic energy

Power

By the end of this section, you will be able to:

- Relate the work done during a time interval to the power delivered
- Find the power expended by a force acting on a moving body

The concept of work involves force and displacement; the work-energy theorem relates the net work done on a body to the difference in its kinetic energy, calculated between two points on its trajectory. None of these quantities or relations involves time explicitly, yet we know that the time available to accomplish a particular amount of work is frequently just as important to us as the amount itself. In the chapter-opening figure, several sprinters may have achieved the same velocity at the finish, and therefore did the same amount of work, but the winner of the race did it in the least amount of time.

We express the relation between work done and the time interval involved in doing it, by introducing the concept of power. Since work can vary as a function of time, we first define **average power** as the work done during a time interval, divided by the interval,

Equation:

$$P_{\text{ave}} = \frac{\Delta W}{\Delta t}.$$

Then, we can define the **instantaneous power** (frequently referred to as just plain **power**).

Note:

Power

Power is defined as the rate of doing work, or the limit of the average power for time intervals approaching zero,

Equation:

$$P = \frac{dW}{dt}.$$

If the power is constant over a time interval, the average power for that interval equals the instantaneous power, and the work done by the agent supplying the power is $W = P\Delta t$. If the power during an interval varies with time, then the work done is the time integral of the power,

Equation:

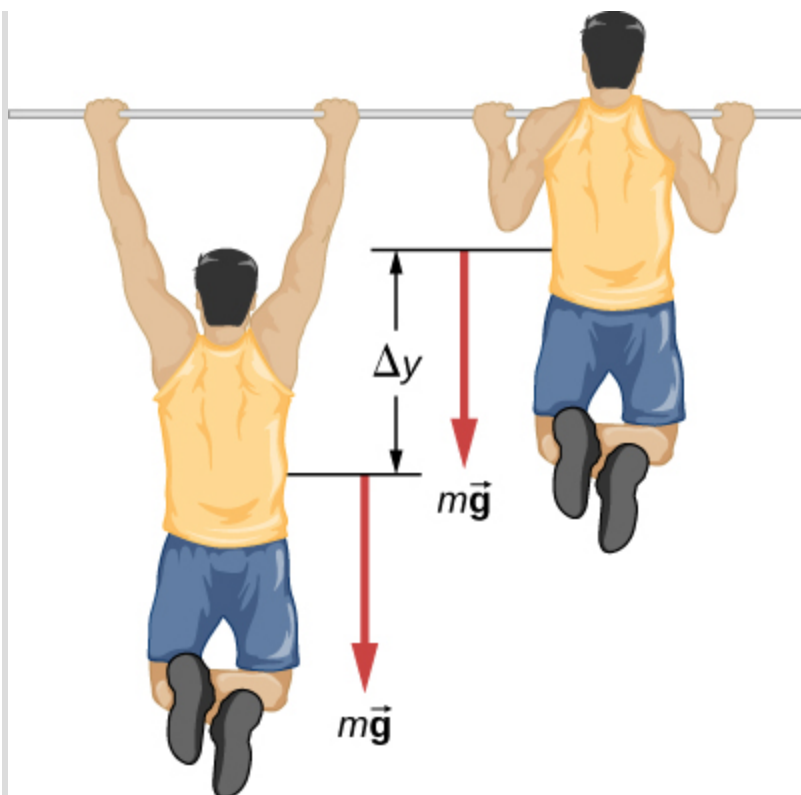
$$W = \int P dt.$$

The work-energy theorem relates how work can be transformed into kinetic energy. Since there are other forms of energy as well, as we discuss in the next chapter, we can also define power as the rate of transfer of energy. Work and energy are measured in units of joules, so power is measured in units of joules per second, which has been given the SI name watts, abbreviation W: $1 \text{ J/s} = 1 \text{ W}$. Another common unit for expressing the power capability of everyday devices is horsepower: $1 \text{ hp} = 746 \text{ W}$.

Example:

Pull-Up Power

An 80-kg army trainee does pull-ups on a horizontal bar ([link](#)). It takes the trainee 0.8 seconds to raise the body from a lower position to where the chin is above the bar. How much power do the trainee's muscles supply moving his body from the lower position to where the chin is above the bar? (*Hint: Make reasonable estimates for any quantities needed.*)



What is the power expended in doing ten pull-ups in ten seconds?

Strategy

The work done against gravity, going up or down a distance Δy , is $mg\Delta y$. Let's assume that $\Delta y = 2\text{ft} \approx 60\text{ cm}$. Also, assume that the arms comprise 10% of the body mass and are not included in the moving mass. With these assumptions, we can calculate the work done.

Solution

The result we get, applying our assumptions, is

Equation:

$$P = \frac{mg(\Delta y)}{t} = \frac{0.9 (80\text{ kg})(9.8\text{ m/s}^2)(0.60\text{ m})}{0.8\text{ s}} = 529\text{ W}.$$

Significance

This is typical for power expenditure in strenuous exercise; in everyday units, it's somewhat more than one horsepower ($1\text{ hp} = 746\text{ W}$).

Note:

Exercise:

Problem:

Check Your Understanding Estimate the power expended by a weightlifter raising a 150-kg barbell 2 m in 3 s.

Solution:

980 W

The power involved in moving a body can also be expressed in terms of the forces acting on it. If a force $\vec{\mathbf{F}}$ acts on a body that is displaced $d\vec{\mathbf{r}}$ in a time dt , the power expended by the force is

Note:

Equation:

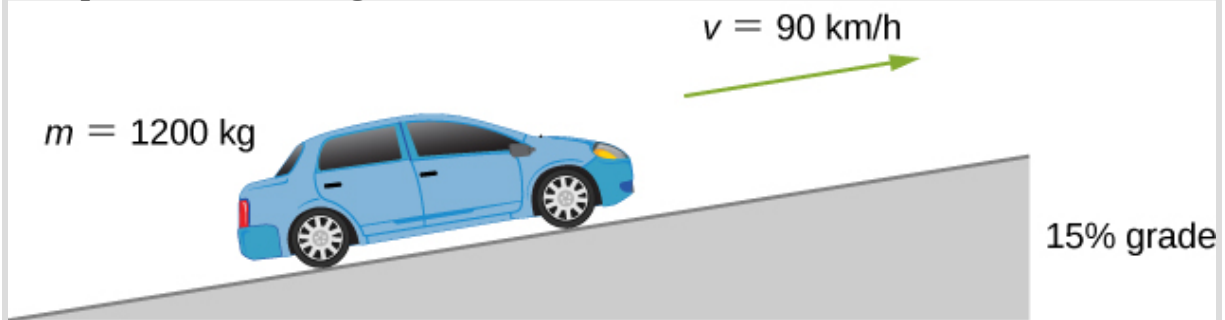
$$P = \frac{dW}{dt} = \frac{\vec{\mathbf{F}} \cdot d\vec{\mathbf{r}}}{dt} = \vec{\mathbf{F}} \cdot \left(\frac{d\vec{\mathbf{r}}}{dt} \right) = \vec{\mathbf{F}} \cdot \vec{\mathbf{v}},$$

where $\vec{\mathbf{v}}$ is the velocity of the body. The fact that the limits implied by the derivatives exist, for the motion of a real body, justifies the rearrangement of the infinitesimals.

Example:

Automotive Power Driving Uphill

How much power must an automobile engine expend to move a 1200-kg car up a 15% grade at 90 km/h ([link](#))? Assume that 25% of this power is dissipated overcoming air resistance and friction.



We want to calculate the power needed to move a car up a hill at constant speed.

Strategy

At constant velocity, there is no change in kinetic energy, so the net work done to move the car is zero. Therefore the power supplied by the engine to move the car equals the power expended against gravity and air resistance. By assumption, 75% of the power is supplied against gravity, which equals $m\vec{g} \cdot \vec{v} = mgv \sin \theta$, where θ is the angle of the incline. A 15% grade means $\tan \theta = 0.15$. This reasoning allows us to solve for the power required.

Solution

Carrying out the suggested steps, we find

Equation:

$$0.75 P = mgv \sin(\tan^{-1} 0.15),$$

or

Equation:

$$P = \frac{(1200 \times 9.8 \text{ N})(90 \text{ m}/3.6 \text{ s})\sin(8.53^\circ)}{0.75} = 58 \text{ kW},$$

or about 78 hp. (You should supply the steps used to convert units.)

Significance

This is a reasonable amount of power for the engine of a small to mid-size car to supply ($1 \text{ hp} = 0.746 \text{ kW}$). Note that this is only the power expended to move the car. Much of the engine's power goes elsewhere, for example, into waste heat. That's why cars need radiators. Any remaining power could be used for acceleration, or to operate the car's accessories.

Summary

- Power is the rate of doing work; that is, the derivative of work with respect to time.
- Alternatively, the work done, during a time interval, is the integral of the power supplied over the time interval.
- The power delivered by a force, acting on a moving particle, is the dot product of the force and the particle's velocity.

Key Equations

Work done by a force over an infinitesimal displacement	$dW = \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}} = \vec{\mathbf{F}} d\vec{\mathbf{r}} \cos \theta$
Work done by a force acting along a path from A to B	$W_{AB} = \int_{\text{path } AB} \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}}$
Work done by a constant force of kinetic friction	$W_{\text{fr}} = -f_k l_{AB} $
Work done going from A to B by Earth's gravity, near its surface	$W_{\text{grav}, AB} = -mg(y_B - y_A)$

Work done going from A to B by one-dimensional spring force	$W_{\text{spring},AB} = -\left(\frac{1}{2}k\right)(x_B^2 - x_A^2)$
Kinetic energy of a non-relativistic particle	$K = \frac{1}{2}mv^2 = \frac{p^2}{2m}$
Work-energy theorem	$W_{\text{net}} = K_B - K_A$
Power as rate of doing work	$P = \frac{dW}{dt}$
Power as the dot product of force and velocity	$P = \vec{\mathbf{F}} \cdot \vec{\mathbf{v}}$

Conceptual Questions

Exercise:

Problem:

Most electrical appliances are rated in watts. Does this rating depend on how long the appliance is on? (When off, it is a zero-watt device.) Explain in terms of the definition of power.

Solution:

Appliances are rated in terms of the energy consumed in a relatively small time interval. It does not matter how long the appliance is on, only the rate of change of energy per unit time.

Exercise:

Problem:

Explain, in terms of the definition of power, why energy consumption is sometimes listed in kilowatt-hours rather than joules. What is the relationship between these two energy units?

Exercise:**Problem:**

A spark of static electricity, such as that you might receive from a doorknob on a cold dry day, may carry a few hundred watts of power. Explain why you are not injured by such a spark.

Solution:

The spark occurs over a relatively short time span, thereby delivering a very low amount of energy to your body.

Exercise:**Problem:**

Does the work done in lifting an object depend on how fast it is lifted?
Does the power expended depend on how fast it is lifted?

Exercise:

Problem: Can the power expended by a force be negative?

Solution:

If the force is antiparallel or points in an opposite direction to the velocity, the power expended can be negative.

Exercise:**Problem:**

How can a 50-W light bulb use more energy than a 1000-W oven?

Problems**Exercise:**

Problem:

A person in good physical condition can put out 100 W of useful power for several hours at a stretch, perhaps by pedaling a mechanism that drives an electric generator. Neglecting any problems of generator efficiency and practical considerations such as resting time: (a) How many people would it take to run a 4.00-kW electric clothes dryer? (b) How many people would it take to replace a large electric power plant that generates 800 MW?

Solution:

a. 40; b. 8 million

Exercise:**Problem:**

What is the cost of operating a 3.00-W electric clock for a year if the cost of electricity is \$0.0900 per kW · h?

Exercise:**Problem:**

A large household air conditioner may consume 15.0 kW of power. What is the cost of operating this air conditioner 3.00 h per day for 30.0 d if the cost of electricity is \$0.110 per kW · h?

Solution:

\$149

Exercise:**Problem:**

(a) What is the average power consumption in watts of an appliance that uses 5.00 kW · h of energy per day? (b) How many joules of energy does this appliance consume in a year?

Exercise:**Problem:**

(a) What is the average useful power output of a person who does 6.00×10^6 J of useful work in 8.00 h? (b) Working at this rate, how long will it take this person to lift 2000 kg of bricks 1.50 m to a platform? (Work done to lift his body can be omitted because it is not considered useful output here.)

Solution:

a. 208 W; b. 141 s

Exercise:**Problem:**

A 500-kg dragster accelerates from rest to a final speed of 110 m/s in 400 m (about a quarter of a mile) and encounters an average frictional force of 1200 N. What is its average power output in watts and horsepower if this takes 7.30 s?

Exercise:**Problem:**

(a) How long will it take an 850-kg car with a useful power output of 40.0 hp (1 hp equals 746 W) to reach a speed of 15.0 m/s, neglecting friction? (b) How long will this acceleration take if the car also climbs a 3.00-m high hill in the process?

Solution:

a. 3.20 s; b. 4.04 s

Exercise:

Problem:

(a) Find the useful power output of an elevator motor that lifts a 2500-kg load a height of 35.0 m in 12.0 s, if it also increases the speed from rest to 4.00 m/s. Note that the total mass of the counterbalanced system is 10,000 kg—so that only 2500 kg is raised in height, but the full 10,000 kg is accelerated. (b) What does it cost, if electricity is \$0.0900 per kW · h ?

Exercise:**Problem:**

(a) How long would it take a 1.50×10^5 -kg airplane with engines that produce 100 MW of power to reach a speed of 250 m/s and an altitude of 12.0 km if air resistance were negligible? (b) If it actually takes 900 s, what is the power? (c) Given this power, what is the average force of air resistance if the airplane takes 1200 s? (*Hint: You must find the distance the plane travels in 1200 s assuming constant acceleration.*)

Solution:

a. 224 s; b. 24.8 MW; c. 49.7 kN

Exercise:**Problem:**

Calculate the power output needed for a 950-kg car to climb a 2.00° slope at a constant 30.0 m/s while encountering wind resistance and friction totaling 600 N.

Exercise:**Problem:**

A man of mass 80 kg runs up a flight of stairs 20 m high in 10 s. (a) how much power is used to lift the man? (b) If the man's body is 25% efficient, how much power does he expend?

Solution:

a. 1.57 kW; b. 6.28 kW

Exercise:**Problem:**

The man of the preceding problem consumes approximately 1.05×10^7 J (2500 food calories) of energy per day in maintaining a constant weight. What is the average power he produces over a day? Compare this with his power production when he runs up the stairs.

Exercise:**Problem:**

An electron in a television tube is accelerated uniformly from rest to a speed of 8.4×10^7 m/s over a distance of 2.5 cm. What is the power delivered to the electron at the instant that its displacement is 1.0 cm?

Solution:

$6.83 \mu\text{W}$

Exercise:**Problem:**

Coal is lifted out of a mine a vertical distance of 50 m by an engine that supplies 500 W to a conveyor belt. How much coal per minute can be brought to the surface? Ignore the effects of friction.

Exercise:**Problem:**

A girl pulls her 15-kg wagon along a flat sidewalk by applying a 10-N force at 37° to the horizontal. Assume that friction is negligible and that the wagon starts from rest. (a) How much work does the girl do on the wagon in the first 2.0 s. (b) How much instantaneous power does she exert at $t = 2.0$ s?

Solution:

a. 8.51 J; b. 8.51 W

Exercise:**Problem:**

A typical automobile engine has an efficiency of 25%. Suppose that the engine of a 1000-kg automobile has a maximum power output of 140 hp. What is the maximum grade that the automobile can climb at 50 km/h if the frictional retarding force on it is 300 N?

Exercise:**Problem:**

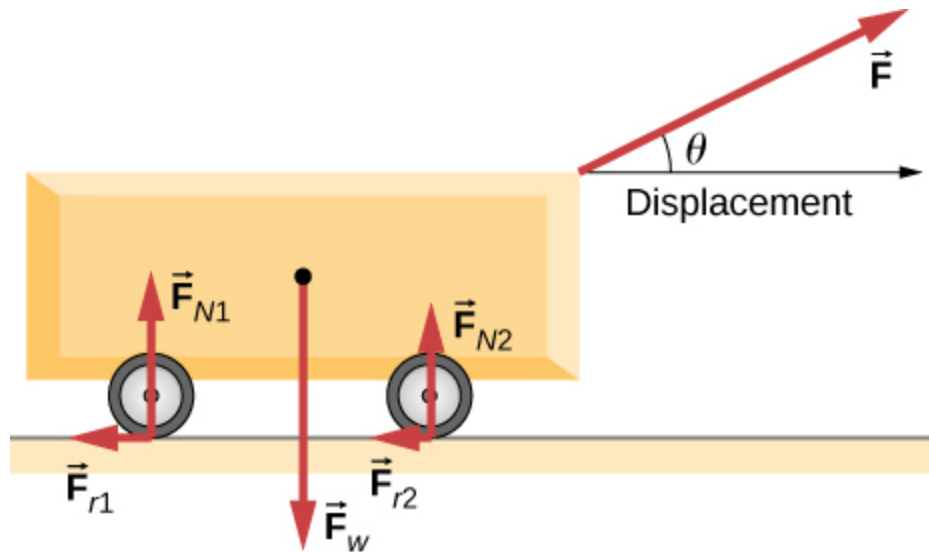
When jogging at 13 km/h on a level surface, a 70-kg man uses energy at a rate of approximately 850 W. Using the facts that the “human engine” is approximately 25% efficient, determine the rate at which this man uses energy when jogging up a 5.0° slope at this same speed. Assume that the frictional retarding force is the same in both cases.

Solution:

1.7 kW

Additional Problems**Exercise:****Problem:**

A cart is pulled a distance D on a flat, horizontal surface by a constant force F that acts at an angle θ with the horizontal direction. The other forces on the object during this time are gravity (F_w), normal forces (F_{N1}) and (F_{N2}), and rolling frictions F_{r1} and F_{r2} , as shown below. What is the work done by each force?



Exercise:

Problem:

Consider a particle on which several forces act, one of which is known to be constant in time: $\vec{F}_1 = (3 \text{ N})\hat{i} + (4 \text{ N})\hat{j}$. As a result, the particle moves along the x -axis from $x = 0$ to $x = 5 \text{ m}$ in some time interval. What is the work done by \vec{F}_1 ?

Solution:

$15 \text{ N} \cdot \text{m}$

Exercise:

Problem:

Consider a particle on which several forces act, one of which is known to be constant in time: $\vec{F}_1 = (3 \text{ N})\hat{i} + (4 \text{ N})\hat{j}$. As a result, the particle moves first along the x -axis from $x = 0$ to $x = 5 \text{ m}$ and then parallel to the y -axis from $y = 0$ to $y = 6 \text{ m}$. What is the work done by \vec{F}_1 ?

Exercise:

Problem:

Consider a particle on which several forces act, one of which is known to be constant in time: $\vec{F}_1 = (3 \text{ N})\hat{i} + (4 \text{ N})\hat{j}$. As a result, the particle moves along a straight path from a Cartesian coordinate of (0 m, 0 m) to (5 m, 6 m). What is the work done by \vec{F}_1 ?

Solution:

39 N · m

Exercise:**Problem:**

Consider a particle on which a force acts that depends on the position of the particle. This force is given by $\vec{F}_1 = (2y)\hat{i} + (3x)\hat{j}$. Find the work done by this force when the particle moves from the origin to a point 5 meters to the right on the x-axis.

Exercise:**Problem:**

A boy pulls a 5-kg cart with a 20-N force at an angle of 30° above the horizontal for a length of time. Over this time frame, the cart moves a distance of 12 m on the horizontal floor. (a) Find the work done on the cart by the boy. (b) What will be the work done by the boy if he pulled with the same force horizontally instead of at an angle of 30° above the horizontal over the same distance?

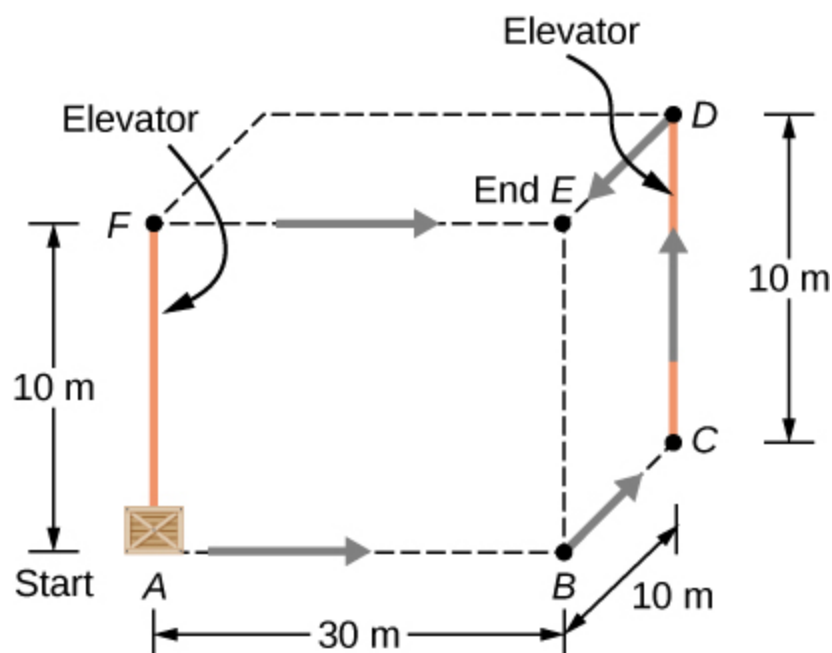
Solution:

a. 208 N · m; b. 240 N · m

Exercise:

Problem:

A crate of mass 200 kg is to be brought from a site on the ground floor to a third floor apartment. The workers know that they can either use the elevator first, then slide it along the third floor to the apartment, or first slide the crate to another location marked C below, and then take the elevator to the third floor and slide it on the third floor a shorter distance. The trouble is that the third floor is very rough compared to the ground floor. Given that the coefficient of kinetic friction between the crate and the ground floor is 0.100 and between the crate and the third floor surface is 0.300 , find the work needed by the workers for each path shown from A to E. Assume that the force the workers need to do is just enough to slide the crate at constant velocity (zero acceleration). *Note:* The work by the elevator against the force of gravity is not done by the workers.



Exercise:

Problem:

A hockey puck of mass 0.17 kg is shot across a rough floor with the roughness different at different places, which can be described by a position-dependent coefficient of kinetic friction. For a puck moving along the x -axis, the coefficient of kinetic friction is the following function of x , where x is in m: $\mu(x) = 0.1 + 0.05x$. Find the work done by the kinetic frictional force on the hockey puck when it has moved (a) from $x = 0$ to $x = 2$ m, and (b) from $x = 2$ m to $x = 4$ m.

Solution:

a. $-0.9 \text{ N} \cdot \text{m}$; b. $-0.83 \text{ N} \cdot \text{m}$

Exercise:**Problem:**

A horizontal force of 20 N is required to keep a 5.0 kg box traveling at a constant speed up a frictionless incline for a vertical height change of 3.0 m. (a) What is the work done by gravity during this change in height? (b) What is the work done by the normal force? (c) What is the work done by the horizontal force?

Exercise:**Problem:**

A 7.0-kg box slides along a horizontal frictionless floor at 1.7 m/s and collides with a relatively massless spring that compresses 23 cm before the box comes to a stop. (a) How much kinetic energy does the box have before it collides with the spring? (b) Calculate the work done by the spring. (c) Determine the spring constant of the spring.

Solution:

a. 10. J; b. 10. J; c. 380 N/m

Exercise:

Problem:

You are driving your car on a straight road with a coefficient of friction between the tires and the road of 0.55. A large piece of debris falls in front of your view and you immediately slam on the brakes, leaving a skid mark of 30.5 m (100-feet) long before coming to a stop. A policeman sees your car stopped on the road, looks at the skid mark, and gives you a ticket for traveling over the 13.4 m/s (30 mph) speed limit. Should you fight the speeding ticket in court?

Exercise:**Problem:**

A crate is being pushed across a rough floor surface. If no force is applied on the crate, the crate will slow down and come to a stop. If the crate of mass 50 kg moving at speed 8 m/s comes to rest in 10 seconds, what is the rate at which the frictional force on the crate takes energy away from the crate?

Solution:

160 J/s

Exercise:**Problem:**

Suppose a horizontal force of 20 N is required to maintain a speed of 8 m/s of a 50 kg crate. (a) What is the power of this force? (b) Note that the acceleration of the crate is zero despite the fact that 20 N force acts on the crate horizontally. What happens to the energy given to the crate as a result of the work done by this 20 N force?

Exercise:

Problem:

Grains from a hopper falls at a rate of 10 kg/s vertically onto a conveyor belt that is moving horizontally at a constant speed of 2 m/s.

(a) What force is needed to keep the conveyor belt moving at the constant velocity? (b) What is the minimum power of the motor driving the conveyor belt?

Solution:

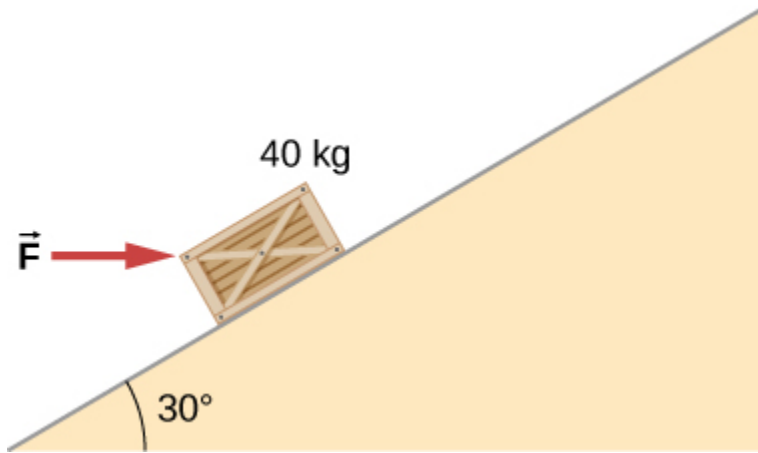
a. 10 N; b. 20 W

Exercise:**Problem:**

A cyclist in a race must climb a 5° hill at a speed of 8 m/s. If the mass of the bike and the biker together is 80 kg, what must be the power output of the biker to achieve the goal?

Challenge Problems**Exercise:****Problem:**

Shown below is a 40-kg crate that is pushed at constant velocity a distance 8.0 m along a 30° incline by the horizontal force \vec{F} . The coefficient of kinetic friction between the crate and the incline is $\mu_k = 0.40$. Calculate the work done by (a) the applied force, (b) the frictional force, (c) the gravitational force, and (d) the net force.



Solution:

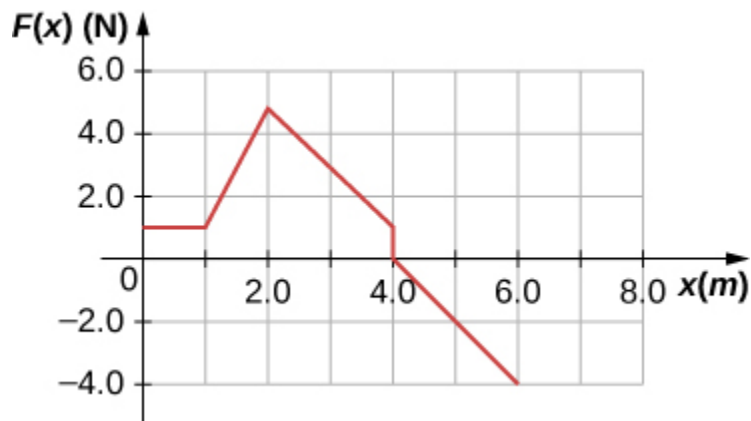
If crate goes up: a. 3.46 kJ; b. -1.89 kJ; c. -1.57 kJ; d. 0; If crate goes down: a. -0.39 kJ; b. -1.18 kJ; c. 1.57 kJ; d. 0

Exercise:**Problem:**

The surface of the preceding problem is modified so that the coefficient of kinetic friction is decreased. The same horizontal force is applied to the crate, and after being pushed 8.0 m, its speed is 5.0 m/s. How much work is now done by the force of friction? Assume that the crate starts at rest.

Exercise:**Problem:**

The force $F(x)$ varies with position, as shown below. Find the work done by this force on a particle as it moves from $x = 1.0$ m to $x = 5.0$ m.



Solution:

8.0 J

Exercise:

Problem:

Find the work done by the same force in [\[link\]](#), between the same points, $A = (0, 0)$ and $B = (2 \text{ m}, 2 \text{ m})$, over a circular arc of radius 2 m, centered at $(0, 2 \text{ m})$. Evaluate the path integral using Cartesian coordinates. (*Hint: You will probably need to consult a table of integrals.*)

Exercise:

Problem: Answer the preceding problem using polar coordinates.

Solution:

35.7 J

Exercise:

Problem:

Find the work done by the same force in [\[link\]](#), between the same points, $A = (0, 0)$ and $B = (2 \text{ m}, 2 \text{ m})$, over a circular arc of radius 2 m, centered at $(2 \text{ m}, 0)$. Evaluate the path integral using Cartesian coordinates. (*Hint:* You will probably need to consult a table of integrals.)

Exercise:

Problem: Answer the preceding problem using polar coordinates.

Solution:

24.3 J

Exercise:**Problem:**

Constant power P is delivered to a car of mass m by its engine. Show that if air resistance can be ignored, the distance covered in a time t by the car, starting from rest, is given by $s = (8P/9m)^{1/2}t^{3/2}$.

Exercise:**Problem:**

Suppose that the air resistance a car encounters is independent of its speed. When the car travels at 15 m/s, its engine delivers 20 hp to its wheels. (a) What is the power delivered to the wheels when the car travels at 30 m/s? (b) How much energy does the car use in covering 10 km at 15 m/s? At 30 m/s? Assume that the engine is 25% efficient. (c) Answer the same questions if the force of air resistance is proportional to the speed of the automobile. (d) What do these results, plus your experience with gasoline consumption, tell you about air resistance?

Solution:

a. 40 hp; b. 39.8 MJ, independent of speed; c. 80 hp, 79.6 MJ at 30 m/s; d. If air resistance is proportional to speed, the car gets about 22 mpg at 34 mph and half that at twice the speed, closer to actual driving experience.

Exercise:

Problem:

Consider a linear spring, as in [\[link\]](#)(a), with mass M uniformly distributed along its length. The left end of the spring is fixed, but the right end, at the equilibrium position $x = 0$, is moving with speed v in the x -direction. What is the total kinetic energy of the spring? (*Hint:* First express the kinetic energy of an infinitesimal element of the spring dm in terms of the total mass, equilibrium length, speed of the right-hand end, and position along the spring; then integrate.)

Glossary

average power

work done in a time interval divided by the time interval

power

(or instantaneous power) rate of doing work

Introduction

class="introduction"

Shown here
is part of a
Ball
Machine
sculpture by
George
Rhoads. A
ball in this
contraption
is lifted,
rolls, falls,
bounces,
and collides
with various
objects, but
throughout
its travels,
its kinetic
energy
changes in
definite,
predictable
amounts,
which
depend on
its position
and the
objects with
which it
interacts.
(credit:
modificatio
n of work

by Roland
Tanglao)



In George Rhoads' rolling ball sculpture, the principle of conservation of energy governs the changes in the ball's kinetic energy and relates them to changes and transfers for other types of energy associated with the ball's interactions. In this chapter, we introduce the important concept of potential energy. This will enable us to formulate the law of conservation of mechanical energy and to apply it to simple systems, making solving problems easier. In the final section on sources of energy, we will consider energy transfers and the general law of conservation of energy. Throughout this book, the law of conservation of energy will be applied in increasingly more detail, as you encounter more complex and varied systems, and other forms of energy.

Potential Energy of a System

By the end of this section, you will be able to:

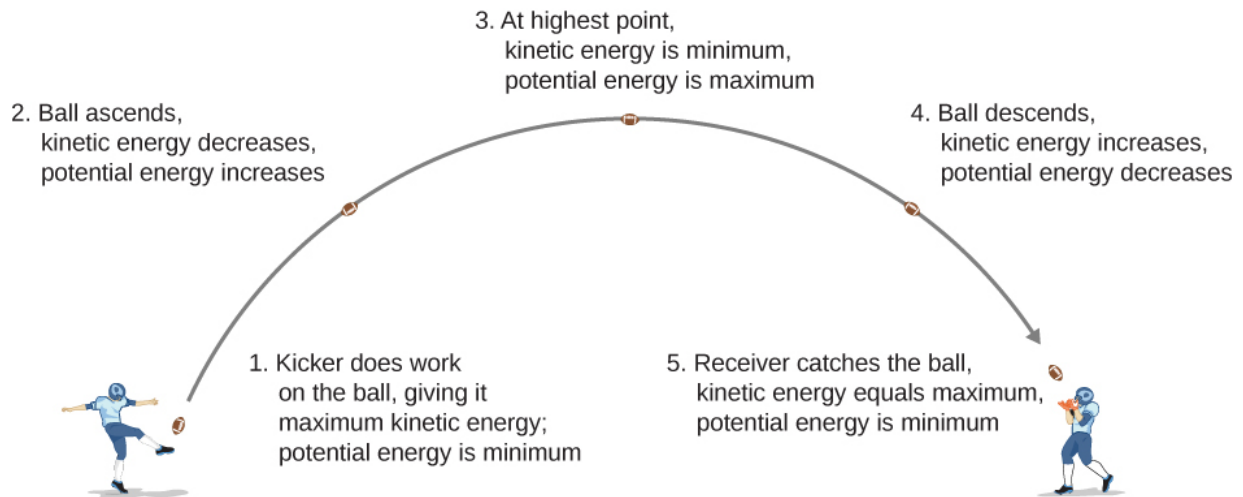
- Relate the difference of potential energy to work done on a particle for a system without friction or air drag
- Explain the meaning of the zero of the potential energy function for a system
- Calculate and apply the gravitational potential energy for an object near Earth's surface and the elastic potential energy of a mass-spring system

In [Work](#), we saw that the work done on an object by the constant gravitational force, near the surface of Earth, over any displacement is a function only of the difference in the positions of the end-points of the displacement. This property allows us to define a different kind of energy for the system than its kinetic energy, which is called **potential energy**. We consider various properties and types of potential energy in the following subsections.

Potential Energy Basics

In [Motion in Two and Three Dimensions](#), we analyzed the motion of a projectile, like kicking a football in [\[link\]](#). For this example, let's ignore friction and air resistance. As the football rises, the work done by the gravitational force on the football is negative, because the ball's displacement is positive vertically and the force due to gravity is negative vertically. We also noted that the ball slowed down until it reached its highest point in the motion, thereby decreasing the ball's kinetic energy. This loss in kinetic energy translates to a gain in gravitational potential energy of the football-Earth system.

As the football falls toward Earth, the work done on the football is now positive, because the displacement and the gravitational force both point vertically downward. The ball also speeds up, which indicates an increase in kinetic energy. Therefore, energy is converted from gravitational potential energy back into kinetic energy.



As a football starts its descent toward the wide receiver, gravitational potential energy is converted back into kinetic energy.

Based on this scenario, we can define the difference of potential energy from point *A* to point *B* as the negative of the work done:

Note:

Equation:

$$\Delta U_{AB} = U_B - U_A = -W_{AB}.$$

This formula explicitly states a **potential energy difference**, not just an absolute potential energy. Therefore, we need to define potential energy at a given position in such a way as to state standard values of potential energy on their own, rather than potential energy differences. We do this by rewriting the potential energy function in terms of an arbitrary constant,

Note:

Equation:

$$\Delta U = U(\vec{r}) - U(\vec{r}_0).$$

The choice of the potential energy at a starting location of \vec{r}_0 is made out of convenience in the given problem. Most importantly, whatever choice is made should be stated and kept consistent throughout the given problem. There are some well-accepted choices of initial potential energy. For example, the lowest height in a problem is usually defined as zero potential energy, or if an object is in space, the farthest point away from the system is often defined as zero potential energy. Then, the potential energy, with respect to zero at \vec{r}_0 , is just $U(\vec{r})$.

As long as there is no friction or air resistance, the change in kinetic energy of the football equals negative of the change in gravitational potential energy of the football. This can be generalized to any potential energy:

Equation:

$$\Delta K_{AB} = -\Delta U_{AB}.$$

Let's look at a specific example, choosing zero potential energy for gravitational potential energy at convenient points.

Example:**Basic Properties of Potential Energy**

A particle moves along the x -axis under the action of a force given by $F = -ax^2$, where $a = 3 \text{ N/m}^2$. (a) What is the difference in its potential energy as it moves from $x_A = 1 \text{ m}$ to $x_B = 2 \text{ m}$? (b) What is the particle's potential energy at $x = 1 \text{ m}$ with respect to a given 0.5 J of potential energy at $x = 0$?

Strategy

(a) The difference in potential energy is the negative of the work done, as defined by [\[link\]](#). The work is defined in the previous chapter as the dot

product of the force with the distance. Since the particle is moving forward in the x -direction, the dot product simplifies to a multiplication ($\hat{\mathbf{i}} \cdot \hat{\mathbf{i}} = 1$). To find the total work done, we need to integrate the function between the given limits. After integration, we can state the work or the change in potential energy. (b) The potential energy function, with respect to zero at $x = 0$, is the indefinite integral encountered in part (a), with the constant of integration determined from [\[link\]](#). Then, we substitute the x -value into the function of potential energy to calculate the potential energy at $x = 1 \text{ m}$.

Solution

- a. The work done by the given force as the particle moves from coordinate x to $x + dx$ in one dimension is

Equation:

$$dW = \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}} = Fdx = -ax^2dx.$$

Substituting this expression into [\[link\]](#), we obtain

Equation:

$$\Delta U = -W = \int_{x_1}^{x_2} ax^2dx = \frac{1}{3}(3 \text{ N/m}^2)x^3 \Big|_{1 \text{ m}}^{2 \text{ m}} = 7 \text{ J}.$$

- b. The indefinite integral for the potential energy function in part (a) is

Equation:

$$U(x) = \frac{1}{3}ax^3 + \text{const.},$$

and we want the constant to be determined by

Equation:

$$U(0) = 0.5 \text{ J}.$$

Thus, the potential energy with respect to zero at $x = 0$ is just

Equation:

$$U(x) = \frac{1}{3}ax^3 + 0.5 \text{ J}.$$

Therefore, the potential energy at $x = 1 \text{ m}$ is

Equation:

$$U(1 \text{ m}) = \frac{1}{3} (3 \text{ N/m}^2) (1 \text{ m})^3 + 0.5 \text{ J} = 1.5 \text{ J}.$$

Significance

In this one-dimensional example, any function we can integrate, independent of path, is conservative. Notice how we applied the definition of potential energy difference to determine the potential energy function with respect to zero at a chosen point. Also notice that the potential energy, as determined in part (b), at $x = 1 \text{ m}$ is $U(1 \text{ m}) = 1 \text{ J}$ and at $x = 2 \text{ m}$ is $U(2 \text{ m}) = 8 \text{ J}$; their difference is the result in part (a).

Note:

Exercise:

Problem:

Check Your Understanding In [\[link\]](#), what are the potential energies of the particle at $x = 1 \text{ m}$ and $x = 2 \text{ m}$ with respect to zero at $x = 1.5 \text{ m}$? Verify that the difference of potential energy is still 7 J.

Solution:

$$(4.63 \text{ J}) - (-2.38 \text{ J}) = 7.00 \text{ J}$$

Systems of Several Particles

In general, a system of interest could consist of several particles. The difference in the potential energy of the system is the negative of the work done by gravitational or elastic forces, which, as we will see in the next section, are conservative forces. The potential energy difference depends only on the initial and final positions of the particles, and on some parameters that

characterize the interaction (like mass for gravity or the spring constant for a Hooke's law force).

It is important to remember that potential energy is a property of the interactions between objects in a chosen system, and not just a property of each object. This is especially true for electric forces, although in the examples of potential energy we consider below, parts of the system are either so big (like Earth, compared to an object on its surface) or so small (like a massless spring), that the changes those parts undergo are negligible when included in the system.

Types of Potential Energy

For each type of interaction present in a system, you can label a corresponding type of potential energy. The total potential energy of the system is the sum of the potential energies of all the types. (This follows from the additive property of the dot product in the expression for the work done.) Let's look at some specific examples of types of potential energy discussed in [Work](#). First, we consider each of these forces when acting separately, and then when both act together.

Gravitational potential energy near Earth's surface

The system of interest consists of our planet, Earth, and one or more particles near its surface (or bodies small enough to be considered as particles, compared to Earth). The gravitational force on each particle (or body) is just its weight mg near the surface of Earth, acting vertically down. According to Newton's third law, each particle exerts a force on Earth of equal magnitude but in the opposite direction. Newton's second law tells us that the magnitude of the acceleration produced by each of these forces on Earth is mg divided by Earth's mass. Since the ratio of the mass of any ordinary object to the mass of Earth is vanishingly small, the motion of Earth can be completely neglected. Therefore, we consider this system to be a group of single-particle systems, subject to the uniform gravitational force of Earth.

In [Work](#), the work done on a body by Earth's uniform gravitational force, near its surface, depended on the mass of the body, the acceleration due to gravity,

and the difference in height the body traversed, as given by [\[link\]](#). By definition, this work is the negative of the difference in the gravitational potential energy, so that difference is

Equation:

$$\Delta U_{\text{grav}} = -W_{\text{grav},AB} = mg(y_B - y_A).$$

You can see from this that the gravitational potential energy function, near Earth's surface, is

Note:

Equation:

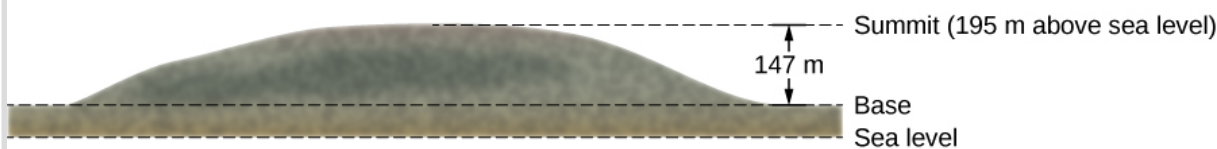
$$U(y) = mgy + \text{const.}$$

You can choose the value of the constant, as described in the discussion of [\[link\]](#); however, for solving most problems, the most convenient constant to choose is zero for when $y = 0$, which is the lowest vertical position in the problem.

Example:

Gravitational Potential Energy of a Hiker

The summit of Great Blue Hill in Milton, MA, is 147 m above its base and has an elevation above sea level of 195 m ([\[link\]](#)). (Its Native American name, *Massachusett*, was adopted by settlers for naming the Bay Colony and state near its location.) A 75-kg hiker ascends from the base to the summit. What is the gravitational potential energy of the hiker-Earth system with respect to zero gravitational potential energy at base height, when the hiker is (a) at the base of the hill, (b) at the summit, and (c) at sea level, afterward?



Sketch of the profile of Great Blue Hill, Milton, MA. The altitudes of the three levels are indicated.

Strategy

First, we need to pick an origin for the y -axis and then determine the value of the constant that makes the potential energy zero at the height of the base. Then, we can determine the potential energies from [\[link\]](#), based on the relationship between the zero potential energy height and the height at which the hiker is located.

Solution

- a. Let's choose the origin for the y -axis at base height, where we also want the zero of potential energy to be. This choice makes the constant equal to zero and

Equation:

$$U(\text{base}) = U(0) = 0.$$

- b. At the summit, $y = 147$ m, so

Equation:

$$U(\text{summit}) = U(147 \text{ m}) = mgh = (75 \times 9.8 \text{ N})(147 \text{ m}) = 108 \text{ kJ}.$$

- c. At sea level, $y = (147 - 195)\text{m} = -48$ m, so

Equation:

$$U(\text{sea-level}) = (75 \times 9.8 \text{ N})(-48 \text{ m}) = -35.3 \text{ kJ}.$$

Significance

Besides illustrating the use of [\[link\]](#) and [\[link\]](#), the values of gravitational potential energy we found are reasonable. The gravitational potential energy is higher at the summit than at the base, and lower at sea level than at the base. Gravity does work on you on your way up, too! It does negative work and not quite as much (in magnitude), as your muscles do. But it certainly does work. Similarly, your muscles do work on your way down, as negative work. The numerical values of the potential energies depend on the choice of zero of potential energy, but the physically meaningful differences of potential energy do not. [Note that since [\[link\]](#) is a difference, the numerical values do not depend on the origin of coordinates.]

Note:

Exercise:

Problem:

Check Your Understanding What are the values of the gravitational potential energy of the hiker at the base, summit, and sea level, with respect to a sea-level zero of potential energy?

Solution:

35.3 kJ, 143 kJ, 0

Elastic potential energy

In [Work](#), we saw that the work done by a perfectly elastic spring, in one dimension, depends only on the spring constant and the squares of the displacements from the unstretched position, as given in [\[link\]](#). This work involves only the properties of a Hooke's law interaction and not the properties of real springs and whatever objects are attached to them. Therefore, we can define the difference of elastic potential energy for a spring force as the negative of the work done by the spring force in this equation, before we consider systems that embody this type of force. Thus,

Equation:

$$\Delta U = -W_{AB} = \frac{1}{2}k(x_B^2 - x_A^2),$$

where the object travels from point A to point B . The potential energy function corresponding to this difference is

Note:

Equation:

$$U(x) = \frac{1}{2}kx^2 + \text{const.}$$

If the spring force is the only force acting, it is simplest to take the zero of potential energy at $x = 0$, when the spring is at its unstretched length. Then, the constant is [\[link\]](#) is zero. (Other choices may be more convenient if other forces are acting.)

Example:

Spring Potential Energy

A system contains a perfectly elastic spring, with an unstretched length of 20 cm and a spring constant of 4 N/cm. (a) How much elastic potential energy does the spring contribute when its length is 23 cm? (b) How much more potential energy does it contribute if its length increases to 26 cm?

Strategy

When the spring is at its unstretched length, it contributes nothing to the potential energy of the system, so we can use [\[link\]](#) with the constant equal to zero. The value of x is the length minus the unstretched length. When the spring is expanded, the spring's displacement or difference between its relaxed length and stretched length should be used for the x -value in calculating the potential energy of the spring.

Solution

- a. The displacement of the spring is $x = 23 \text{ cm} - 20 \text{ cm} = 3 \text{ cm}$, so the contributed potential energy is
- $$U = \frac{1}{2} kx^2 = \frac{1}{2} (4 \text{ N/cm})(3 \text{ cm})^2 = 0.18 \text{ J}.$$
- b. When the spring's displacement is $x = 26 \text{ cm} - 20 \text{ cm} = 6 \text{ cm}$, the potential energy is $U = \frac{1}{2} kx^2 = \frac{1}{2} (4 \text{ N/cm})(6 \text{ cm})^2 = 0.72 \text{ J}$, which is a 0.54-J increase over the amount in part (a).

Significance

Calculating the elastic potential energy and potential energy differences from [\[link\]](#) involves solving for the potential energies based on the given lengths of the spring. Since U depends on x^2 , the potential energy for a compression (negative x) is the same as for an extension of equal magnitude.

Note:

Exercise:

Problem:

Check Your Understanding When the length of the spring in [\[link\]](#) changes from an initial value of 22.0 cm to a final value, the elastic potential energy it contributes changes by -0.0800 J . Find the final length.

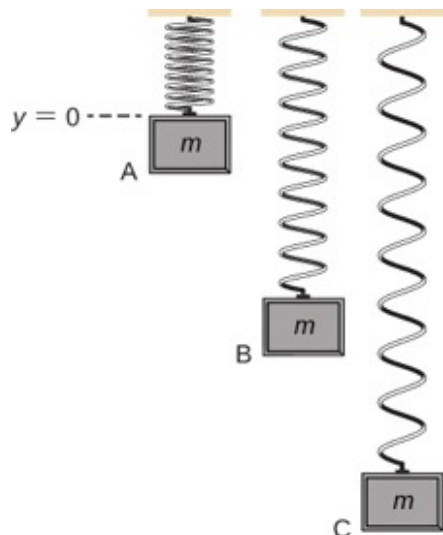
Solution:

22.8 cm. Using 0.02 m for the initial displacement of the spring (see above), we calculate the final displacement of the spring to be 0.028 m; therefore the length of the spring is the unstretched length plus the displacement, or 22.8 cm.

Gravitational and elastic potential energy

A simple system embodying both gravitational and elastic types of potential energy is a one-dimensional, vertical mass-spring system. This consists of a

massive particle (or block), hung from one end of a perfectly elastic, massless spring, the other end of which is fixed, as illustrated in [\[link\]](#).



A vertical mass-spring system, with the positive y -axis pointing upward. The mass is initially at an unstretched spring length, point A. Then it is released, expanding past point B to point C, where it comes to a stop.

First, let's consider the potential energy of the system. We need to define the constant in the potential energy function of [\[link\]](#). Often, the ground is a suitable choice for when the gravitational potential energy is zero; however, in this case, the highest point or when $y = 0$ is a convenient location for zero gravitational potential energy. Note that this choice is arbitrary, and the problem can be solved correctly even if another choice is picked.

We must also define the elastic potential energy of the system and the corresponding constant, as detailed in [\[link\]](#). This is where the spring is unstretched, or at the $y = 0$ position.

If we consider that the total energy of the system is conserved, then the energy at point A equals point C. The block is placed just on the spring so its initial kinetic energy is zero. By the setup of the problem discussed previously, both the gravitational potential energy and elastic potential energy are equal to zero. Therefore, the initial energy of the system is zero. When the block arrives at point C, its kinetic energy is zero. However, it now has both gravitational potential energy and elastic potential energy. Therefore, we can solve for the distance y that the block travels before coming to a stop:

Equation:

$$\begin{aligned}K_A + U_A &= K_C + U_C \\0 &= 0 + mgy_C + \frac{1}{2}k(y_C)^2 \\y_C &= \frac{-2mg}{k}\end{aligned}$$



A bungee jumper transforms gravitational potential energy at the start of the jump into elastic potential energy at the bottom of the jump.

Example:**Potential Energy of a Vertical Mass-Spring System**

A block weighing 1.2 N is hung from a spring with a spring constant of 6.0 N/m, as shown in [\[link\]](#). (a) What is the maximum expansion of the spring, as seen at point C? (b) What is the total potential energy at point B, halfway between A and C? (c) What is the speed of the block at point B?

Strategy

In part (a) we calculate the distance y_C as discussed in the previous text. Then in part (b), we use half of the y value to calculate the potential energy at point B using equations [\[link\]](#) and [\[link\]](#). This energy must be equal to the kinetic

energy, [link](#)], at point B since the initial energy of the system is zero. By calculating the kinetic energy at point B, we can now calculate the speed of the block at point B.

Solution

- a. Since the total energy of the system is zero at point A as discussed previously, the maximum expansion of the spring is calculated to be:

Equation:

$$y_C = \frac{-2mg}{k}$$
$$y_C = \frac{-2(1.2\text{N})}{(6.0\text{N/m})} = -0.40 \text{ m}$$

- b. The position of y_B is half of the position at y_C or -0.20 m . The total potential energy at point B would therefore be:

Equation:

$$U_B = mgy_B + \frac{1}{2}k(y_C)^2$$
$$U_B = (1.2 \text{ N})(-0.20 \text{ m}) + \frac{1}{2}(6 \text{ N/m})(-0.20 \text{ m})^2$$
$$U_B = -0.12 \text{ J}$$

- c. The mass of the block is the weight divided by gravity.

Equation:

$$m = \frac{F_w}{g} = \frac{1.2 \text{ N}}{9.8\text{m/s}^2} = 0.12\text{kg}$$

The kinetic energy at point B therefore is 0.12 J because the total energy is zero. Therefore, the speed of the block at point B is equal to

Equation:

$$K = \frac{1}{2}mv^2$$
$$v = \sqrt{\frac{2K}{m}} = \sqrt{\frac{2(0.12\text{J})}{(0.12\text{kg})}} = 1.4 \text{ m/s}$$

Significance

Even though the potential energy due to gravity is relative to a chosen zero location, the solutions to this problem would be the same if the zero energy points were chosen at different locations.

Note:

Exercise:

Problem:

Check Your Understanding Suppose the mass in [\[link\]](#) is doubled while keeping the all other conditions the same. Would the maximum expansion of the spring increase, decrease, or remain the same? Would the speed at point B be larger, smaller, or the same compared to the original mass?

Solution:

It increases because you had to exert a downward force, doing positive work, to pull the mass down, and that's equal to the change in the total potential energy.

Note:

View this [simulation](#) to learn about conservation of energy with a skater! Build tracks, ramps and jumps for the skater and view the kinetic energy, potential energy and friction as he moves. You can also take the skater to different planets or even space!

A sample chart of a variety of energies is shown in [\[link\]](#) to give you an idea about typical energy values associated with certain events. Some of these are calculated using kinetic energy, whereas others are calculated by using quantities found in a form of potential energy that may not have been discussed at this point.

Object/phenomenon	Energy in joules
Big Bang	10^{68}
Annual world energy use	4.0×10^{20}
Large fusion bomb (9 megaton)	3.8×10^{16}
Hiroshima-size fission bomb (10 kiloton)	4.2×10^{13}
1 barrel crude oil	5.9×10^9
1 metric ton TNT	4.2×10^9
1 gallon of gasoline	1.2×10^8
Daily adult food intake (recommended)	1.2×10^7
1000-kg car at 90 km/h	3.1×10^5
Tennis ball at 100 km/h	22
Mosquito (10^{-2} g at 0.5 m/s)	1.3×10^{-6}
Single electron in a TV tube beam	4.0×10^{-15}
Energy to break one DNA strand	10^{-19}

Energy of Various Objects and Phenomena

Summary

- For a single-particle system, the difference of potential energy is the opposite of the work done by the forces acting on the particle as it moves from one position to another.
- Since only differences of potential energy are physically meaningful, the zero of the potential energy function can be chosen at a convenient

location.

- The potential energies for Earth's constant gravity, near its surface, and for a Hooke's law force are linear and quadratic functions of position, respectively.

Conceptual Questions

Exercise:

Problem:

The kinetic energy of a system must always be positive or zero. Explain whether this is true for the potential energy of a system.

Solution:

The potential energy of a system can be negative because its value is relative to a defined point.

Exercise:

Problem:

The force exerted by a diving board is conservative, provided the internal friction is negligible. Assuming friction is negligible, describe changes in the potential energy of a diving board as a swimmer drives from it, starting just before the swimmer steps on the board until just after his feet leave it.

Exercise:

Problem:

Describe the gravitational potential energy transfers and transformations for a javelin, starting from the point at which an athlete picks up the javelin and ending when the javelin is stuck into the ground after being thrown.

Solution:

If the reference point of the ground is zero gravitational potential energy, the javelin first increases its gravitational potential energy, followed by a decrease in its gravitational potential energy as it is thrown until it hits the ground. The overall change in gravitational potential energy of the javelin is zero unless the center of mass of the javelin is lower than from where it is initially thrown, and therefore would have slightly less gravitational potential energy.

Exercise:

Problem:

A couple of soccer balls of equal mass are kicked off the ground at the same speed but at different angles. Soccer ball A is kicked off at an angle slightly above the horizontal, whereas ball B is kicked slightly below the vertical. How do each of the following compare for ball A and ball B? (a) The initial kinetic energy and (b) the change in gravitational potential energy from the ground to the highest point? If the energy in part (a) differs from part (b), explain why there is a difference between the two energies.

Exercise:

Problem:

What is the dominant factor that affects the speed of an object that started from rest down a frictionless incline if the only work done on the object is from gravitational forces?

Solution:

the vertical height from the ground to the object

Exercise:

Problem:

Two people observe a leaf falling from a tree. One person is standing on a ladder and the other is on the ground. If each person were to compare the energy of the leaf observed, would each person find the following to be the same or different for the leaf, from the point where it falls off the tree to when it hits the ground: (a) the kinetic energy of the leaf; (b) the change in gravitational potential energy; (c) the final gravitational potential energy?

Problems**Exercise:****Problem:**

Using values from [\[link\]](#), how many DNA molecules could be broken by the energy carried by a single electron in the beam of an old-fashioned TV tube? (These electrons were not dangerous in themselves, but they did create dangerous X-rays. Later-model tube TVs had shielding that absorbed X-rays before they escaped and exposed viewers.)

Solution:

40,000

Exercise:**Problem:**

If the energy in fusion bombs were used to supply the energy needs of the world, how many of the 9-megaton variety would be needed for a year's supply of energy (using data from [\[link\]](#))?

Exercise:

Problem:

A camera weighing 10 N falls from a small drone hovering 20 m overhead and enters free fall. What is the gravitational potential energy change of the camera from the drone to the ground if you take a reference point of (a) the ground being zero gravitational potential energy? (b) The drone being zero gravitational potential energy? What is the gravitational potential energy of the camera (c) before it falls from the drone and (d) after the camera lands on the ground if the reference point of zero gravitational potential energy is taken to be a second person looking out of a building 30 m from the ground?

Solution:

a. -200 J ; b. -200 J ; c. -100 J ; d. -300 J

Exercise:**Problem:**

Someone drops a 50 — g pebble off of a docked cruise ship, 70.0 m from the water line. A person on a dock 3.0 m from the water line holds out a net to catch the pebble. (a) How much work is done on the pebble by gravity during the drop? (b) What is the change in the gravitational potential energy during the drop? If the gravitational potential energy is zero at the water line, what is the gravitational potential energy (c) when the pebble is dropped? (d) When it reaches the net? What if the gravitational potential energy was 30.0 Joules at water level? (e) Find the answers to the same questions in (c) and (d).

Exercise:

Problem:

A cat's crinkle ball toy of mass 15 g is thrown straight up with an initial speed of 3 m/s. Assume in this problem that air drag is negligible. (a) What is the kinetic energy of the ball as it leaves the hand? (b) How much work is done by the gravitational force during the ball's rise to its peak? (c) What is the change in the gravitational potential energy of the ball during the rise to its peak? (d) If the gravitational potential energy is taken to be zero at the point where it leaves your hand, what is the gravitational potential energy when it reaches the maximum height? (e) What if the gravitational potential energy is taken to be zero at the maximum height the ball reaches, what would the gravitational potential energy be when it leaves the hand? (f) What is the maximum height the ball reaches?

Solution:

a. 0.068 J; b. -0.068 J; c. 0.068 J; d. 0.068 J; e. -0.068 J; f. 46 cm

Glossary

potential energy

function of position, energy possessed by an object relative to the system considered

potential energy difference

negative of the work done acting between two points in space

Conservative and Non-Conservative Forces

By the end of this section, you will be able to:

- Characterize a conservative force in several different ways
- Specify mathematical conditions that must be satisfied by a conservative force and its components
- Relate the conservative force between particles of a system to the potential energy of the system
- Calculate the components of a conservative force in various cases

In [Potential Energy and Conservation of Energy](#), any transition between kinetic and potential energy conserved the total energy of the system. This was path independent, meaning that we can start and stop at any two points in the problem, and the total energy of the system—kinetic plus potential—at these points are equal to each other. This is characteristic of a **conservative force**. We dealt with conservative forces in the preceding section, such as the gravitational force and spring force. When comparing the motion of the football in [\[link\]](#), the total energy of the system never changes, even though the gravitational potential energy of the football increases, as the ball rises relative to ground and falls back to the initial gravitational potential energy when the football player catches the ball. **Non-conservative forces** are dissipative forces such as friction or air resistance. These forces take energy away from the system as the system progresses, energy that you can't get back. These forces are path dependent; therefore it matters where the object starts and stops.

Note:

Conservative Force

The work done by a conservative force is independent of the path; in other words, the work done by a conservative force is the same for any path connecting two points:

Equation:

$$W_{AB,\text{path-1}} = \int_{AB,\text{path-1}} \vec{\mathbf{F}}_{\text{cons}} \cdot d\vec{\mathbf{r}} = W_{AB,\text{path-2}} = \int_{AB,\text{path-2}} \vec{\mathbf{F}}_{\text{cons}} \cdot d\vec{\mathbf{r}}.$$

The work done by a non-conservative force depends on the path taken. Equivalently, a force is conservative if the work it does around any closed path is zero:

Equation:

$$W_{\text{closed path}} = \oint \vec{\mathbf{F}}_{\text{cons}} \cdot d\vec{\mathbf{r}} = 0.$$

[In [\[link\]](#), we use the notation of a circle in the middle of the integral sign for a line integral over a closed path, a notation found in most physics and engineering texts.] [\[link\]](#) and [\[link\]](#) are equivalent because any closed path is the sum of two paths: the first going from A to B , and the second going from B to A . The work done going along a path from B to A is the negative of the work done going along the same path from A to B , where A and B are any two points on the closed path:

Equation:

$$\begin{aligned} 0 = \int \vec{\mathbf{F}}_{\text{cons}} \cdot d\vec{\mathbf{r}} &= \int_{AB,\text{path-1}} \vec{\mathbf{F}}_{\text{cons}} \cdot d\vec{\mathbf{r}} + \int_{BA,\text{path-2}} \vec{\mathbf{F}}_{\text{cons}} \cdot d\vec{\mathbf{r}} \\ &= \int_{AB,\text{path-1}} \vec{\mathbf{F}}_{\text{cons}} \cdot d\vec{\mathbf{r}} - \int_{AB,\text{path-2}} \vec{\mathbf{F}}_{\text{cons}} \cdot d\vec{\mathbf{r}} = 0. \end{aligned}$$

You might ask how we go about proving whether or not a force is conservative, since the definitions involve any and all paths from A to B , or any and all closed paths, but to do the integral for the work, you have to choose a particular path. One answer is that the work done is independent of path if the infinitesimal work $\vec{\mathbf{F}} \cdot d\vec{\mathbf{r}}$ is an **exact differential**, the way the

infinitesimal net work was equal to the exact differential of the kinetic energy, $dW_{\text{net}} = m\vec{\mathbf{v}} \cdot d\vec{\mathbf{v}} = d\frac{1}{2}mv^2$,

when we derived the work-energy theorem in [Work-Energy Theorem](#). There are mathematical conditions that you can use to test whether the infinitesimal work done by a force is an exact differential, and the force is conservative. These conditions only involve differentiation and are thus relatively easy to apply. In two dimensions, the condition for $\vec{\mathbf{F}} \cdot d\vec{\mathbf{r}} = F_x dx + F_y dy$ to be an exact differential is

Note:

Equation:

$$\frac{dF_x}{dy} = \frac{dF_y}{dx}.$$

You may recall that the work done by the force in [\[link\]](#) depended on the path. For that force,

Equation:

$$F_x = (5 \text{ N/m})y \text{ and } F_y = (10 \text{ N/m})x.$$

Therefore,

Equation:

$$(dF_x/dy) = 5 \text{ N/m} \neq (dF_y/dx) = 10 \text{ N/m},$$

which indicates it is a non-conservative force. Can you see what you could change to make it a conservative force?



A grinding wheel applies a non-conservative force, because the work done depends on how many rotations the wheel makes, so it is path-dependent. (credit: modification of work by Grantez Stephens, U.S. Navy)

Example:

Conservative or Not?

Which of the following two-dimensional forces are conservative and which are not? Assume a and b are constants with appropriate units:

(a) $axy^3\hat{\mathbf{i}} + ayx^3\hat{\mathbf{j}}$, (b) $a \left[(y^2/x)\hat{\mathbf{i}} + 2y \ln(x/b)\hat{\mathbf{j}} \right]$, (c) $\frac{ax\hat{\mathbf{i}} + ay\hat{\mathbf{j}}}{x^2 + y^2}$

Strategy

Apply the condition stated in [\[link\]](#), namely, using the derivatives of the components of each force indicated. If the derivative of the y -component of the force with respect to x is equal to the derivative of the x -component of the force with respect to y , the force is a conservative force, which means the path taken for potential energy or work calculations always yields the same results.

Solution

- a. $\frac{dF_x}{dy} = \frac{d(axy^3)}{dy} = 3axy^2$ and $\frac{dF_y}{dx} = \frac{d(ayx^3)}{dx} = 3ayx^2$, so this force is non-conservative.
- b. $\frac{dF_x}{dy} = \frac{d(ay^2/x)}{dy} = \frac{2ay}{x}$ and $\frac{dF_y}{dx} = \frac{d(2ay \ln(x/b))}{dx} = \frac{2ay}{x}$, so this force is conservative.
- c. $\frac{dF_x}{dy} = \frac{d(ax/(x^2+y^2))}{dy} = -\frac{ax(2y)}{(x^2+y^2)^2} = \frac{dF_y}{dx} = \frac{d(ay/(x^2+y^2))}{dx}$, again conservative.

Significance

The conditions in [\[link\]](#) are derivatives as functions of a single variable; in three dimensions, similar conditions exist that involve more derivatives.

Note:

Exercise:

Problem:

Check Your Understanding A two-dimensional, conservative force is zero on the x - and y -axes, and satisfies the condition

$(dF_x/dy) = (dF_y/dx) = (4 \text{ N/m}^3)xy$. What is the magnitude of the force at the point $x = y = 1 \text{ m}$?

Solution:

2.83 N

Before leaving this section, we note that non-conservative forces do not have potential energy associated with them because the energy is lost to the system and can't be turned into useful work later. So there is always a conservative force associated with every potential energy. We have seen

that potential energy is defined in relation to the work done by conservative forces. That relation, [\[link\]](#), involved an integral for the work; starting with the force and displacement, you integrated to get the work and the change in potential energy. However, integration is the inverse operation of differentiation; you could equally well have started with the potential energy and taken its derivative, with respect to displacement, to get the force. The infinitesimal increment of potential energy is the dot product of the force and the infinitesimal displacement,

Equation:

$$dU = -\vec{\mathbf{F}} \cdot d\vec{\mathbf{l}} = -F_l dl.$$

Here, we chose to represent the displacement in an arbitrary direction by $d\vec{\mathbf{l}}$, so as not to be restricted to any particular coordinate direction. We also expressed the dot product in terms of the magnitude of the infinitesimal displacement and the component of the force in its direction. Both these quantities are scalars, so you can divide by dl to get

Note:

Equation:

$$F_l = -\frac{dU}{dl}.$$

This equation gives the relation between force and the potential energy associated with it. In words, the component of a conservative force, in a particular direction, equals the negative of the derivative of the corresponding potential energy, with respect to a displacement in that direction. For one-dimensional motion, say along the x -axis, [\[link\]](#) give the entire vector force, $\mathbf{F} = F_x \hat{\mathbf{i}} = -\frac{\partial U}{\partial x} \hat{\mathbf{i}}$.

In two dimensions,

Equation:

$$\mathbf{F} = F_x \hat{\mathbf{i}} + F_y \hat{\mathbf{j}} = - \left(\frac{\partial U}{\partial x} \right) \hat{\mathbf{i}} - \left(\frac{\partial U}{\partial y} \right) \hat{\mathbf{j}}.$$

From this equation, you can see why [\[link\]](#) is the condition for the work to be an exact differential, in terms of the derivatives of the components of the force. In general, a partial derivative notation is used. If a function has many variables in it, the derivative is taken only of the variable the partial derivative specifies. The other variables are held constant. In three dimensions, you add another term for the z-component, and the result is that the force is the negative of the gradient of the potential energy. However, we won't be looking at three-dimensional examples just yet.

Example:**Force due to a Quartic Potential Energy**

The potential energy for a particle undergoing one-dimensional motion along the x-axis is

Equation:

$$U(x) = \frac{1}{4}cx^4,$$

where $c = 8 \text{ N/m}^3$. Its total energy at $x = 0$ is 2 J, and it is not subject to any non-conservative forces. Find (a) the positions where its kinetic energy is zero and (b) the forces at those positions.

Strategy

(a) We can find the positions where $K = 0$, so the potential energy equals the total energy of the given system. (b) Using [\[link\]](#), we can find the force evaluated at the positions found from the previous part, since the mechanical energy is conserved.

Solution

- a. The total energy of the system of 2 J equals the quartic elastic energy as given in the problem,

Equation:

$$2 \text{ J} = \frac{1}{4} (8 \text{ N/m}^3) x_f^4.$$

Solving for x_f results in $x_f = \pm 1 \text{ m}$.

b. From [\[link\]](#),

Equation:

$$F_x = -dU/dx = -cx^3.$$

Thus, evaluating the force at $\pm 1 \text{ m}$, we get

Equation:

$$\vec{F} = -(8 \text{ N/m}^3)(\pm 1 \text{ m})^3 \hat{i} = \pm 8 \text{ N} \hat{i}.$$

At both positions, the magnitude of the forces is 8 N and the directions are toward the origin, since this is the potential energy for a restoring force.

Significance

Finding the force from the potential energy is mathematically easier than finding the potential energy from the force, because differentiating a function is generally easier than integrating one.

Note:

Exercise:

Problem:

Check Your Understanding Find the forces on the particle in [\[link\]](#) when its kinetic energy is 1.0 J at $x = 0$.

Solution:

$F = 4.8 \text{ N}$, directed toward the origin

Summary

- A conservative force is one for which the work done is independent of path. Equivalently, a force is conservative if the work done over any closed path is zero.
- A non-conservative force is one for which the work done depends on the path.
- For a conservative force, the infinitesimal work is an exact differential. This implies conditions on the derivatives of the force's components.
- The component of a conservative force, in a particular direction, equals the negative of the derivative of the potential energy for that force, with respect to a displacement in that direction.

Conceptual Questions

Exercise:

Problem: What is the physical meaning of a non-conservative force?

Solution:

A force that takes energy away from the system that can't be recovered if we were to reverse the action.

Exercise:

Problem:

A bottle rocket is shot straight up in the air with a speed 30 m/s . If the air resistance is ignored, the bottle would go up to a height of approximately 46 m . However, the rocket goes up to only 35 m before returning to the ground. What happened? Explain, giving only a qualitative response.

Exercise:

Problem:

An external force acts on a particle during a trip from one point to another and back to that same point. This particle is only effected by conservative forces. Does this particle's kinetic energy and potential energy change as a result of this trip?

Solution:

The change in kinetic energy is the net work. Since conservative forces are path independent, when you are back to the same point the kinetic and potential energies are exactly the same as the beginning. During the trip the total energy is conserved, but both the potential and kinetic energy change.

Problems**Exercise:****Problem:**

A force $F(x) = (3.0/x)$ N acts on a particle as it moves along the positive x -axis. (a) How much work does the force do on the particle as it moves from $x = 2.0$ m to $x = 5.0$ m? (b) Picking a convenient reference point of the potential energy to be zero at $x = \infty$, find the potential energy for this force.

Exercise:**Problem:**

A force $F(x) = (-5.0x^2 + 7.0x)$ N acts on a particle. How much work does the force do on the particle as it moves from $x = 2.0$ m to $x = 5.0$ m?

Solution:

-120 J

Exercise:**Problem:**

Find the force corresponding to the potential energy

$$U(x) = -a/x + b/x^2.$$

Exercise:**Problem:**

The potential energy function for either one of the two atoms in a diatomic molecule is often approximated by $U(x) = a/x^{12} - b/x^6$ where x is the distance between the atoms. (a) At what distance of separation does the potential energy have a local minimum (not at $x = \infty$)? (b) What is the force on an atom at this separation? (c) How does the force vary with the separation distance?

Solution:

$$\text{a. } \left(\frac{2a}{b}\right)^{1/6}; \text{ b. } 0; \text{ c. } \sim x^6$$

Exercise:**Problem:**

A particle of mass 2.0 kg moves under the influence of the force $F(x) = (3/\sqrt{x})$ N. If its speed at $x = 2.0$ m is $v = 6.0$ m/s, what is its speed at $x = 7.0$ m?

Exercise:**Problem:**

A particle of mass 2.0 kg moves under the influence of the force $F(x) = (-5x^2 + 7x)$ N. If its speed at $x = -4.0$ m is $v = 20.0$ m/s, what is its speed at $x = 4.0$ m?

Solution:

$$14 \text{ m/s}$$

Exercise:

Problem:

A crate on rollers is being pushed without frictional loss of energy across the floor of a freight car (see the following figure). The car is moving to the right with a constant speed v_0 . If the crate starts at rest relative to the freight car, then from the work-energy theorem, $Fd = mv^2/2$, where d , the distance the crate moves, and v , the speed of the crate, are both measured relative to the freight car. (a) To an observer at rest beside the tracks, what distance d' is the crate pushed when it moves the distance d in the car? (b) What are the crate's initial and final speeds v_0' and v' as measured by the observer beside the tracks? (c) Show that $Fd' = m(v')^2/2 - m(v_0')^2/2$ and, consequently, that work is equal to the change in kinetic energy in both reference systems.



Glossary

conservative force

force that does work independent of path

exact differential

is the total differential of a function and requires the use of partial derivatives if the function involves more than one dimension

non-conservative force

force that does work that depends on path

Conservation of Energy

By the end of this section, you will be able to:

- Formulate the principle of conservation of mechanical energy, with or without the presence of non-conservative forces
- Use the conservation of mechanical energy to calculate various properties of simple systems

In this section, we elaborate and extend the result we derived in [Potential Energy of a System](#), where we re-wrote the work-energy theorem in terms of the change in the kinetic and potential energies of a particle. This will lead us to a discussion of the important principle of the conservation of mechanical energy. As you continue to examine other topics in physics, in later chapters of this book, you will see how this conservation law is generalized to encompass other types of energy and energy transfers. The last section of this chapter provides a preview.

The terms ‘conserved quantity’ and ‘conservation law’ have specific, scientific meanings in physics, which are different from the everyday meanings associated with the use of these words. (The same comment is also true about the scientific and everyday uses of the word ‘work.’) In everyday usage, you could conserve water by not using it, or by using less of it, or by re-using it. Water is composed of molecules consisting of two atoms of hydrogen and one of oxygen. Bring these atoms together to form a molecule and you create water; dissociate the atoms in such a molecule and you destroy water. However, in scientific usage, a **conserved quantity** for a system stays constant, changes by a definite amount that is transferred to other systems, and/or is converted into other forms of that quantity. A conserved quantity, in the scientific sense, can be transformed, but not strictly created or destroyed. Thus, there is no physical law of conservation of water.

Systems with a Single Particle or Object

We first consider a system with a single particle or object. Returning to our development of [\[link\]](#), recall that we first separated all the forces acting on a particle into conservative and non-conservative types, and wrote the work done by each type of force as a separate term in the work-energy theorem. We then replaced the work done by the conservative forces by the change in the potential energy of the particle, combining it with the change in the particle’s kinetic energy to get [\[link\]](#). Now, we write this equation without the middle step and define the sum of the kinetic and potential energies, $K + U = E$; to be the **mechanical energy** of the particle.

Note:

Conservation of Energy

The mechanical energy E of a particle stays constant unless forces outside the system or non-conservative forces do work on it, in which case, the change in the mechanical energy is equal to the work done by the non-conservative forces:

Equation:

$$W_{\text{nc},AB} = \Delta(K + U)_{AB} = \Delta E_{AB}.$$

This statement expresses the concept of **energy conservation** for a classical particle as long as there is no non-conservative work. Recall that a classical particle is just a point mass, is nonrelativistic, and obeys Newton's laws of motion. In [Relativity](#), we will see that conservation of energy still applies to a non-classical particle, but for that to happen, we have to make a slight adjustment to the definition of energy.

It is sometimes convenient to separate the case where the work done by non-conservative forces is zero, either because no such forces are assumed present, or, like the normal force, they do zero work when the motion is parallel to the surface. Then

Note:

Equation:

$$0 = W_{\text{nc},AB} = \Delta(K + U)_{AB} = \Delta E_{AB}.$$

In this case, the conservation of mechanical energy can be expressed as follows: The mechanical energy of a particle does not change if all the non-conservative forces that may act on it do no work. Understanding the concept of energy conservation is the important thing, not the particular equation you use to express it.

Note:

Conservation of Energy

1. Identify the body or bodies to be studied (the system). Often, in applications of the principle of mechanical energy conservation, we study more than one body at the same time.
2. Identify all forces acting on the body or bodies.
3. Determine whether each force that does work is conservative. If a non-conservative force (e.g., friction) is doing work, then mechanical energy is not conserved. The

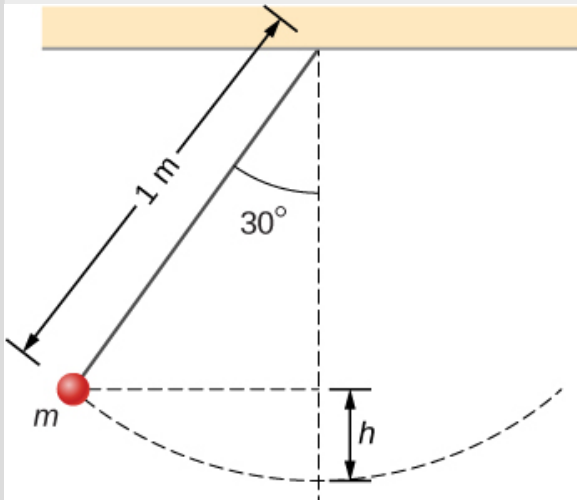
system must then be analyzed with non-conservative work, [\[link\]](#).

4. For every force that does work, choose a reference point and determine the potential energy function for the force. The reference points for the various potential energies do not have to be at the same location.
5. Apply the principle of mechanical energy conservation by setting the sum of the kinetic energies and potential energies equal at every point of interest.

Example:

Simple Pendulum

A particle of mass m is hung from the ceiling by a massless string of length 1.0 m , as shown in [\[link\]](#). The particle is released from rest, when the angle between the string and the downward vertical direction is 30° . What is its speed when it reaches the lowest point of its arc?



A particle hung from a string constitutes a simple pendulum. It is shown when released from rest, along with some distances used in analyzing the motion.

Strategy

Using our problem-solving strategy, the first step is to define that we are interested in the particle-Earth system. Second, only the gravitational force is acting on the particle, which is conservative (step 3). We neglect air resistance in the problem, and no work is done by the string tension, which is perpendicular to the arc of the motion. Therefore, the mechanical energy of the system is conserved, as represented by [\[link\]](#), $0 = \Delta (K + U)$. Because the particle starts from rest, the increase in the kinetic energy is just the kinetic

energy at the lowest point. This increase in kinetic energy equals the decrease in the gravitational potential energy, which we can calculate from the geometry. In step 4, we choose a reference point for zero gravitational potential energy to be at the lowest vertical point the particle achieves, which is mid-swing. Lastly, in step 5, we set the sum of energies at the highest point (initial) of the swing to the lowest point (final) of the swing to ultimately solve for the final speed.

Solution

We are neglecting non-conservative forces, so we write the energy conservation formula relating the particle at the highest point (initial) and the lowest point in the swing (final) as

Equation:

$$K_i + U_i = K_f + U_f.$$

Since the particle is released from rest, the initial kinetic energy is zero. At the lowest point, we define the gravitational potential energy to be zero. Therefore our conservation of energy formula reduces to

Equation:

$$\begin{aligned} 0 + mgh &= \frac{1}{2}mv^2 + 0 \\ v &= \sqrt{2gh}. \end{aligned}$$

The vertical height of the particle is not given directly in the problem. This can be solved for by using trigonometry and two givens: the length of the pendulum and the angle through which the particle is vertically pulled up. Looking at the diagram, the vertical dashed line is the length of the pendulum string. The vertical height is labeled h . The other partial length of the vertical string can be calculated with trigonometry. That piece is solved for by

Equation:

$$\cos \theta = x/L, x = L \cos \theta.$$

Therefore, by looking at the two parts of the string, we can solve for the height h ,

Equation:

$$\begin{aligned} x + h &= L \\ L \cos \theta + h &= L \\ h &= L - L \cos \theta = L(1 - \cos \theta). \end{aligned}$$

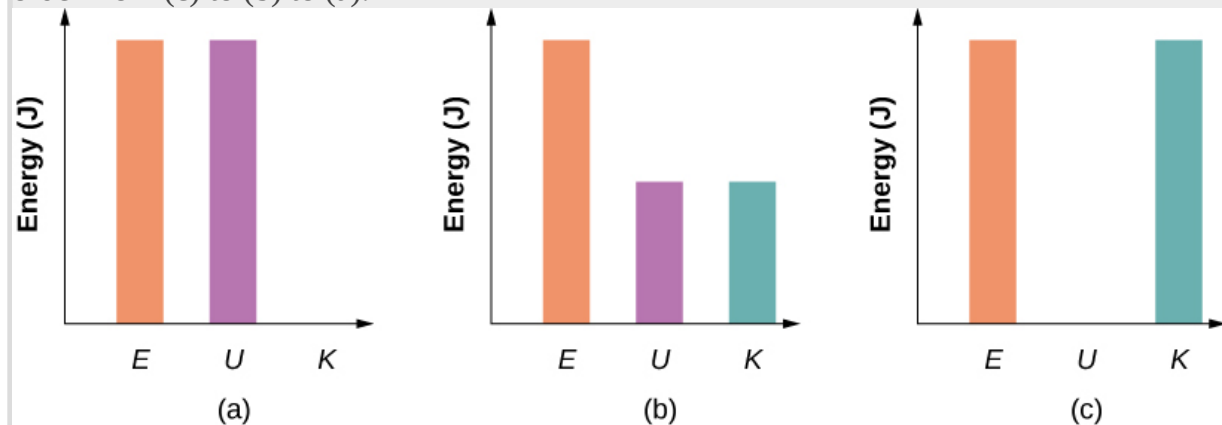
We substitute this height into the previous expression solved for speed to calculate our result:

Equation:

$$v = \sqrt{2gL(1 - \cos \theta)} = \sqrt{2(9.8 \text{ m/s}^2)(1 \text{ m})(1 - \cos 30^\circ)} = 1.62 \text{ m/s}.$$

Significance

We found the speed directly from the conservation of mechanical energy, without having to solve the differential equation for the motion of a pendulum (see [Oscillations](#)). We can approach this problem in terms of bar graphs of total energy. Initially, the particle has all potential energy, being at the highest point, and no kinetic energy. When the particle crosses the lowest point at the bottom of the swing, the energy moves from the potential energy column to the kinetic energy column. Therefore, we can imagine a progression of this transfer as the particle moves between its highest point, lowest point of the swing, and back to the highest point ([link](#)). As the particle travels from the lowest point in the swing to the highest point on the far right hand side of the diagram, the energy bars go in reverse order from (c) to (b) to (a).



Bar graphs representing the total energy (E), potential energy (U), and kinetic energy (K) of the particle in different positions. (a) The total energy of the system equals the potential energy and the kinetic energy is zero, which is found at the highest point the particle reaches. (b) The particle is midway between the highest and lowest point, so the kinetic energy plus potential energy bar graphs equal the total energy. (c) The particle is at the lowest point of the swing, so the kinetic energy bar graph is the highest and equal to the total energy of the system.

Note:

Exercise:

Problem:

Check Your Understanding How high above the bottom of its arc is the particle in the simple pendulum above, when its speed is 0.81 m/s ?

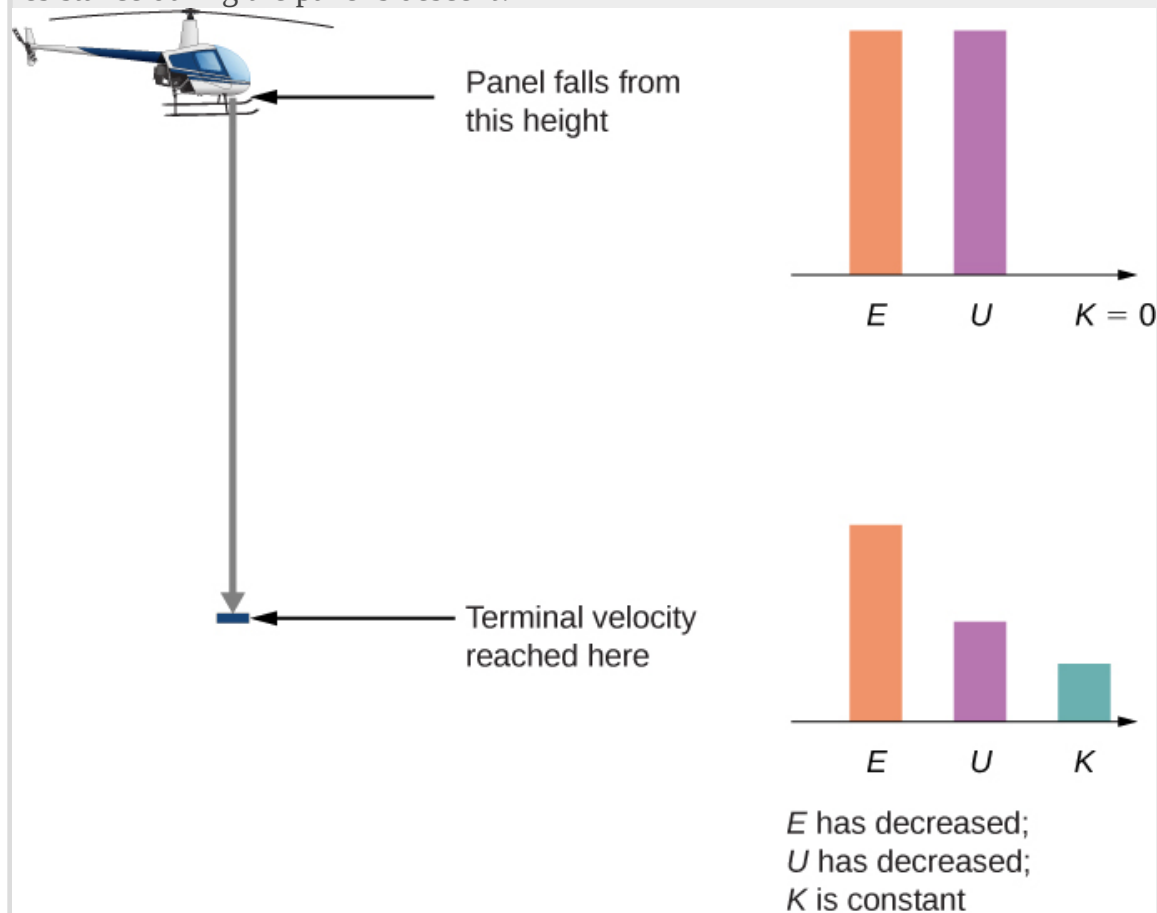
Solution:

0.033 m

Example:

Air Resistance on a Falling Object

A helicopter is hovering at an altitude of 1 km when a panel from its underside breaks loose and plummets to the ground ([link](#)). The mass of the panel is 15 kg, and it hits the ground with a speed of 45 m/s. How much mechanical energy was dissipated by air resistance during the panel's descent?



A helicopter loses a panel that falls until it reaches terminal velocity of 45 m/s.
How much did air resistance contribute to the dissipation of energy in this problem?

Strategy

Step 1: Here only one body is being investigated.

Step 2: Gravitational force is acting on the panel, as well as air resistance, which is stated in the problem.

Step 3: Gravitational force is conservative; however, the non-conservative force of air resistance does negative work on the falling panel, so we can use the conservation of mechanical energy, in the form expressed by [\[link\]](#), to find the energy dissipated. This energy is the magnitude of the work:

Equation:

$$\Delta E_{\text{diss}} = |W_{\text{nc,if}}| = |\Delta(K + U)_{\text{if}}|.$$

Step 4: The initial kinetic energy, at $y_i = 1 \text{ km}$, is zero. We set the gravitational potential energy to zero at ground level out of convenience.

Step 5: The non-conservative work is set equal to the energies to solve for the work dissipated by air resistance.

Solution

The mechanical energy dissipated by air resistance is the algebraic sum of the gain in the kinetic energy and loss in potential energy. Therefore the calculation of this energy is

Equation:

$$\begin{aligned}\Delta E_{\text{diss}} &= |K_f - K_i + U_f - U_i| \\ &= \left| \frac{1}{2}(15 \text{ kg})(45 \text{ m/s})^2 - 0 + 0 - (15 \text{ kg}) \left(9.8 \text{ m/s}^2\right) (1000 \text{ m}) \right| = 130 \text{ kJ}.\end{aligned}$$

Significance

Most of the initial mechanical energy of the panel (U_i), 147 kJ, was lost to air resistance.

Notice that we were able to calculate the energy dissipated without knowing what the force of air resistance was, only that it was dissipative.

Note:

Exercise:

Problem:

Check Your Understanding You probably recall that, neglecting air resistance, if you throw a projectile straight up, the time it takes to reach its maximum height equals the time it takes to fall from the maximum height back to the starting height. Suppose you cannot neglect air resistance, as in [\[link\]](#). Is the time the projectile takes to go up (a) greater than, (b) less than, or (c) equal to the time it takes to come back down? Explain.

Solution:

b. At any given height, the gravitational potential energy is the same going up or down, but the kinetic energy is less going down than going up, since air resistance is dissipative and does negative work. Therefore, at any height, the speed going down is less than the speed going up, so it must take a longer time to go down than to go up.

In these examples, we were able to use conservation of energy to calculate the speed of a particle just at particular points in its motion. But the method of analyzing particle motion, starting from energy conservation, is more powerful than that. More advanced treatments of the theory of mechanics allow you to calculate the full time dependence of a particle's motion, for a given potential energy. In fact, it is often the case that a better model for particle motion is provided by the form of its kinetic and potential energies, rather than an equation for force acting on it. (This is especially true for the quantum mechanical description of particles like electrons or atoms.)

We can illustrate some of the simplest features of this energy-based approach by considering a particle in one-dimensional motion, with potential energy $U(x)$ and no non-conservative interactions present. [\[link\]](#) and the definition of velocity require

Equation:

$$K = \frac{1}{2}mv^2 = E - U(x)$$

$$v = \frac{dx}{dt} = \sqrt{\frac{2(E - U(x))}{m}}.$$

Separate the variables x and t and integrate, from an initial time $t = 0$ to an arbitrary time, to get

Equation:

$$t = \int_0^t dt = \int_{x_0}^x \frac{dx}{\sqrt{2[E - U(x)]/m}}.$$

If you can do the integral in [\[link\]](#), then you can solve for x as a function of t .

Example:

Constant Acceleration

Use the potential energy $U(x) = -E(x/x_0)$, for $E > 0$, in [\[link\]](#) to find the position x of a particle as a function of time t .

Strategy

Since we know how the potential energy changes as a function of x , we can substitute for $U(x)$ in [\[link\]](#), integrate, and then solve for x . This results in an expression of x as a function of time with constants of energy E , mass m , and the initial position x_0 .

Solution

Following the first two suggested steps in the above strategy,

Equation:

$$t = \int_{x_0}^x \frac{dx}{\sqrt{(2E/mx_0)(x_0 - x)}} = \frac{1}{\sqrt{(2E/mx_0)}} \left| -2\sqrt{(x_0 - x)} \right|_{x_0}^x = -\frac{2\sqrt{(x_0 - x)}}{\sqrt{(2E/mx_0)}}.$$

Solving for the position, we obtain $x(t) = x_0 - \frac{1}{2}(E/mx_0)t^2$.

Significance

The position as a function of time, for this potential, represents one-dimensional motion with constant acceleration, $a = (E/mx_0)$, starting at rest from position x_0 . This is not so surprising, since this is a potential energy for a constant force, $F = -dU/dx = E/x_0$, and $a = F/m$.

Note:**Exercise:****Problem:**

Check Your Understanding What potential energy $U(x)$ can you substitute in [\[link\]](#) that will result in motion with constant velocity of 2 m/s for a particle of mass 1 kg and mechanical energy 1 J?

Solution:

constant $U(x) = -1 \text{ J}$

We will look at another more physically appropriate example of the use of [\[link\]](#) after we have explored some further implications that can be drawn from the functional form of a particle's potential energy.

Systems with Several Particles or Objects

Systems generally consist of more than one particle or object. However, the conservation of mechanical energy, in one of the forms in [\[link\]](#) or [\[link\]](#), is a fundamental law of physics and applies to any system. You just have to include the kinetic and potential energies of all the particles, and the work done by all the non-conservative forces acting on them. Until you learn more about the dynamics of systems composed of many particles, in [Linear Momentum and Collisions](#), [Fixed-Axis Rotation](#), and [Angular Momentum](#), it is better to postpone discussing the application of energy conservation to then.

Summary

- A conserved quantity is a physical property that stays constant regardless of the path taken.
- A form of the work-energy theorem says that the change in the mechanical energy of a particle equals the work done on it by non-conservative forces.
- If non-conservative forces do no work and there are no external forces, the mechanical energy of a particle stays constant. This is a statement of the conservation of mechanical energy and there is no change in the total mechanical energy.
- For one-dimensional particle motion, in which the mechanical energy is constant and the potential energy is known, the particle's position, as a function of time, can be found by evaluating an integral that is derived from the conservation of mechanical energy.

Conceptual Questions

Exercise:

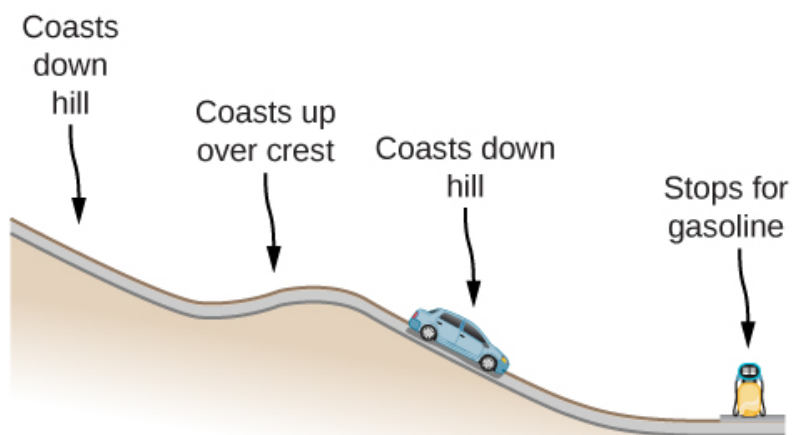
Problem:

When a body slides down an inclined plane, does the work of friction depend on the body's initial speed? Answer the same question for a body sliding down a curved surface.

Exercise:

Problem:

Consider the following scenario. A car for which friction is *not* negligible accelerates from rest down a hill, running out of gasoline after a short distance (see below). The driver lets the car coast farther down the hill, then up and over a small crest. He then coasts down that hill into a gas station, where he brakes to a stop and fills the tank with gasoline. Identify the forms of energy the car has, and how they are changed and transferred in this series of events.



Solution:

The car experiences a change in gravitational potential energy as it goes down the hills because the vertical distance is decreasing. Some of this change of gravitational potential energy will be taken away by work done by friction. The rest of the energy results in a kinetic energy increase, making the car go faster. Lastly, the car brakes and will lose its kinetic energy to the work done by braking to a stop.

Exercise:**Problem:**

A dropped ball bounces to one-half its original height. Discuss the energy transformations that take place.

Exercise:**Problem:**

“ $E = K + U$ constant is a special case of the work-energy theorem.” Discuss this statement.

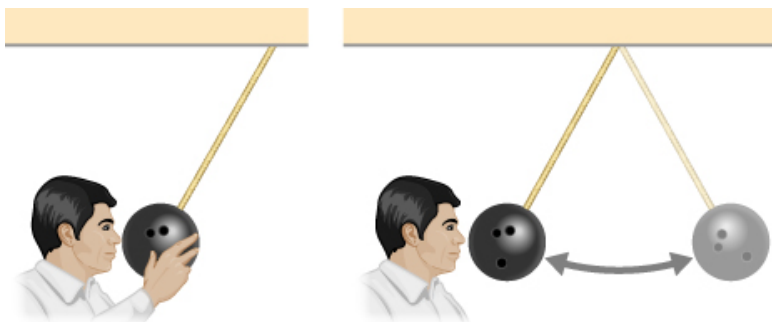
Solution:

It states that total energy of the system E is conserved as long as there are no non-conservative forces acting on the object.

Exercise:**Problem:**

In a common physics demonstration, a bowling ball is suspended from the ceiling by a rope.

The professor pulls the ball away from its equilibrium position and holds it adjacent to his nose, as shown below. He releases the ball so that it swings directly away from him. Does he get struck by the ball on its return swing? What is he trying to show in this demonstration?



Exercise:**Problem:**

A child jumps up and down on a bed, reaching a higher height after each bounce. Explain how the child can increase his maximum gravitational potential energy with each bounce.

Solution:

He puts energy into the system through his legs compressing and expanding.

Exercise:

Problem: Can a non-conservative force increase the mechanical energy of the system?

Exercise:**Problem:**

Neglecting air resistance, how much would I have to raise the vertical height if I wanted to double the impact speed of a falling object?

Solution:

Four times the original height would double the impact speed.

Exercise:**Problem:**

A box is dropped onto a spring at its equilibrium position. The spring compresses with the box attached and comes to rest. Since the spring is in the vertical position, does the change in the gravitational potential energy of the box while the spring is compressing need to be considered in this problem?

Problems**Exercise:****Problem:**

A boy throws a ball of mass 0.25 kg straight upward with an initial speed of 20 m/s. When the ball returns to the boy, its speed is 17 m/s. How much work does air resistance do on the ball during its flight?

Solution:

14 J

Exercise:

Problem:

A mouse of mass 200 g falls 100 m down a vertical mine shaft and lands at the bottom with a speed of 8.0 m/s. During its fall, how much work is done on the mouse by air resistance?

Exercise:

Problem:

Using energy considerations and assuming negligible air resistance, show that a rock thrown from a bridge 20.0 m above water with an initial speed of 15.0 m/s strikes the water with a speed of 24.8 m/s independent of the direction thrown. (*Hint:* show that $K_i + U_i = K_f + U_f$)

Solution:

proof

Exercise:

Problem:

A 1.0-kg ball at the end of a 2.0-m string swings in a vertical plane. At its lowest point the ball is moving with a speed of 10 m/s. (a) What is its speed at the top of its path? (b) What is the tension in the string when the ball is at the bottom and at the top of its path?

Exercise:

Problem:

Ignoring details associated with friction, extra forces exerted by arm and leg muscles, and other factors, we can consider a pole vault as the conversion of an athlete's running kinetic energy to gravitational potential energy. If an athlete is to lift his body 4.8 m during a vault, what speed must he have when he plants his pole?

Solution:

9.7 m/s

Exercise:

Problem:

Tarzan grabs a vine hanging vertically from a tall tree when he is running at 9.0 m/s. (a) How high can he swing upward? (b) Does the length of the vine affect this height?

Exercise:**Problem:**

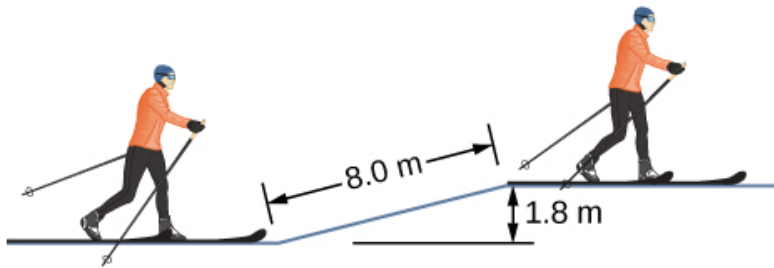
Assume that the force of a bow on an arrow behaves like the spring force. In aiming the arrow, an archer pulls the bow back 50 cm and holds it in position with a force of 150 N. If the mass of the arrow is 50 g and the “spring” is massless, what is the speed of the arrow immediately after it leaves the bow?

Solution:

39 m/s

Exercise:**Problem:**

A 100 – kg man is skiing across level ground at a speed of 8.0 m/s when he comes to the small slope 1.8 m higher than ground level shown in the following figure. (a) If the skier coasts up the hill, what is his speed when he reaches the top plateau? Assume friction between the snow and skis is negligible. (b) What is his speed when he reaches the upper level if an 80 – N frictional force acts on the skis?

**Exercise:****Problem:**

A sled of mass 70 kg starts from rest and slides down a 10° incline 80 m long. It then travels for 20 m horizontally before starting back up an 8° incline. It travels 80 m along this incline before coming to rest. What is the magnitude of the net work done on the sled by friction?

Solution:

1900 J

Exercise:

Problem:

A girl on a skateboard (total mass of 40 kg) is moving at a speed of 10 m/s at the bottom of a long ramp. The ramp is inclined at 20° with respect to the horizontal. If she travels 14.2 m upward along the ramp before stopping, what is the net frictional force on her?

Exercise:**Problem:**

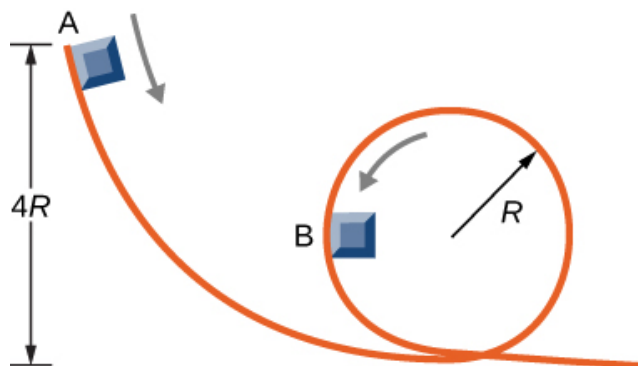
A baseball of mass 0.25 kg is hit at home plate with a speed of 40 m/s. When it lands in a seat in the left-field bleachers a horizontal distance 120 m from home plate, it is moving at 30 m/s. If the ball lands 20 m above the spot where it was hit, how much work is done on it by air resistance?

Solution:

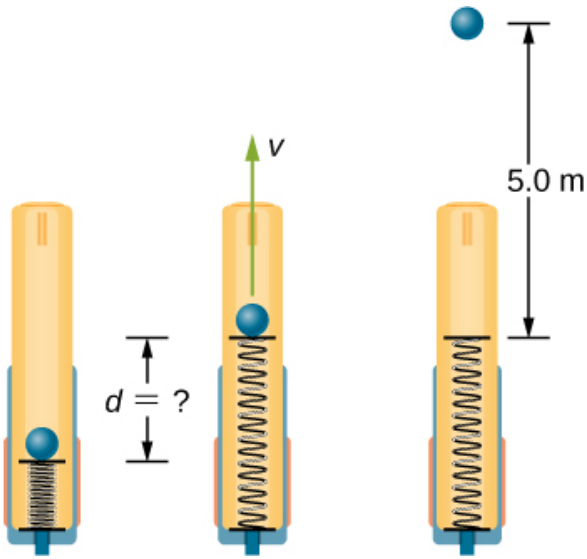
-39 J

Exercise:**Problem:**

A small block of mass m slides without friction around the loop-the-loop apparatus shown below. (a) If the block starts from rest at A, what is its speed at B? (b) What is the force of the track on the block at B?

**Exercise:****Problem:**

The massless spring of a spring gun has a force constant $k = 12 \text{ N/cm}$. When the gun is aimed vertically, a 15-g projectile is shot to a height of 5.0 m above the end of the expanded spring. (See below.) How much was the spring compressed initially?



Solution:

3.5 cm

Exercise:

Problem:

A small ball is tied to a string and set rotating with negligible friction in a vertical circle. If the ball moves over the top of the circle at its slowest possible speed (so that the tension in the string is negligible), what is the tension in the string at the bottom of the circle, assuming there is no additional energy added to the ball during rotation?

Glossary

conserved quantity

one that cannot be created or destroyed, but may be transformed between different forms of itself

energy conservation

total energy of an isolated system is constant

mechanical energy

sum of the kinetic and potential energies

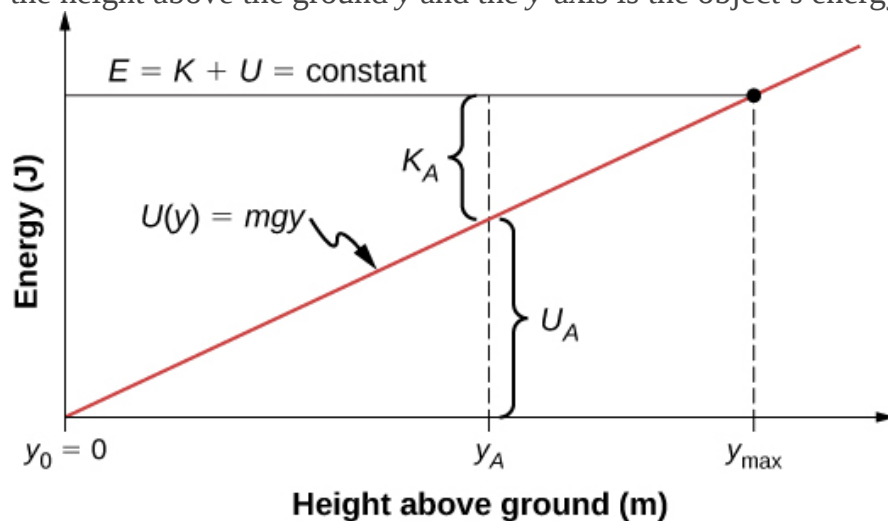
Potential Energy Diagrams and Stability

By the end of this section, you will be able to:

- Create and interpret graphs of potential energy
- Explain the connection between stability and potential energy

Often, you can get a good deal of useful information about the dynamical behavior of a mechanical system just by interpreting a graph of its potential energy as a function of position, called a **potential energy diagram**. This is most easily accomplished for a one-dimensional system, whose potential energy can be plotted in one two-dimensional graph—for example, $U(x)$ versus x —on a piece of paper or a computer program. For systems whose motion is in more than one dimension, the motion needs to be studied in three-dimensional space. We will simplify our procedure for one-dimensional motion only.

First, let's look at an object, freely falling vertically, near the surface of Earth, in the absence of air resistance. The mechanical energy of the object is conserved, $E = K + U$, and the potential energy, with respect to zero at ground level, is $U(y) = mgy$, which is a straight line through the origin with slope mg . In the graph shown in [\[link\]](#), the x -axis is the height above the ground y and the y -axis is the object's energy.



The potential energy graph for an object in vertical free fall, with various quantities indicated.

The line at energy E represents the constant mechanical energy of the object, whereas the kinetic and potential energies, K_A and U_A , are indicated at a particular height y_A . You can see how the total energy is divided between kinetic and potential energy as the object's height changes. Since kinetic energy can never be negative, there is a maximum

potential energy and a maximum height, which an object with the given total energy cannot exceed:

Equation:

$$K = E - U \geq 0,$$
$$U \leq E.$$

If we use the gravitational potential energy reference point of zero at y_0 , we can rewrite the gravitational potential energy U as mgy . Solving for y results in

Equation:

$$y \leq E/mg = y_{\max}.$$

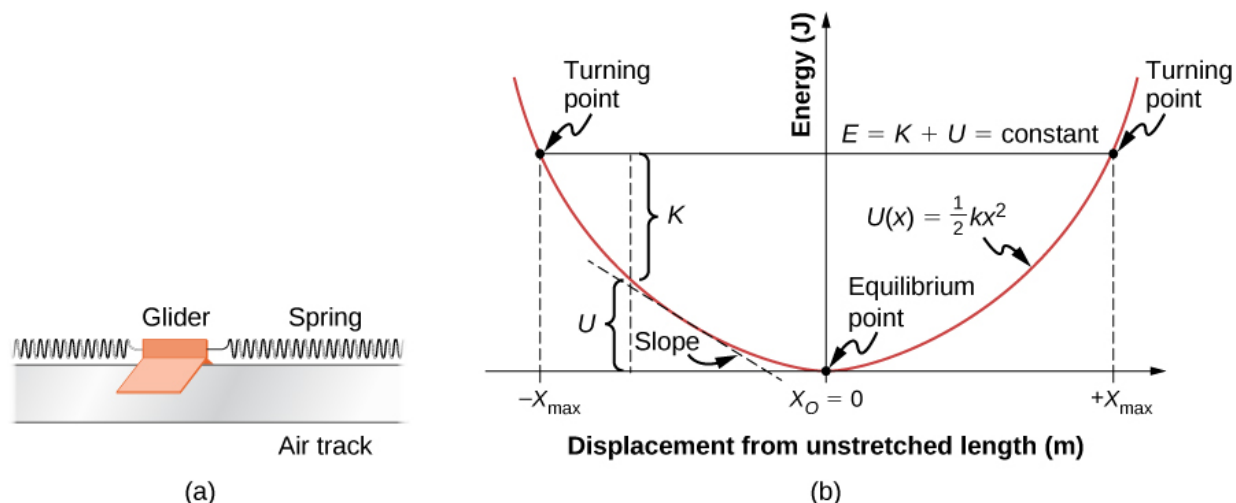
We note in this expression that the quantity of the total energy divided by the weight (mg) is located at the maximum height of the particle, or y_{\max} . At the maximum height, the kinetic energy and the speed are zero, so if the object were initially traveling upward, its velocity would go through zero there, and y_{\max} would be a turning point in the motion. At ground level, $y_0 = 0$, the potential energy is zero, and the kinetic energy and the speed are maximum:

Equation:

$$U_0 = 0 = E - K_0,$$
$$E = K_0 = \frac{1}{2}mv_0^2,$$
$$v_0 = \pm\sqrt{2E/m}.$$

The maximum speed $\pm v_0$ gives the initial velocity necessary to reach y_{\max} , the maximum height, and $-v_0$ represents the final velocity, after falling from y_{\max} . You can read all this information, and more, from the potential energy diagram we have shown.

Consider a mass-spring system on a frictionless, stationary, horizontal surface, so that gravity and the normal contact force do no work and can be ignored ([\[link\]](#)). This is like a one-dimensional system, whose mechanical energy E is a constant and whose potential energy, with respect to zero energy at zero displacement from the spring's unstretched length, $x = 0$, is $U(x) = \frac{1}{2}kx^2$.



(a) A glider between springs on an air track is an example of a horizontal mass-spring system. (b) The potential energy diagram for this system, with various quantities indicated.

You can read off the same type of information from the potential energy diagram in this case, as in the case for the body in vertical free fall, but since the spring potential energy describes a variable force, you can learn more from this graph. As for the object in vertical free fall, you can deduce the physically allowable range of motion and the maximum values of distance and speed, from the limits on the kinetic energy, $0 \leq K \leq E$. Therefore, $K = 0$ and $U = E$ at a **turning point**, of which there are two for the elastic spring potential energy,

Equation:

$$x_{\max} = \pm \sqrt{2E/k}.$$

The glider's motion is confined to the region between the turning points, $-x_{\max} \leq x \leq x_{\max}$. This is true for any (positive) value of E because the potential energy is unbounded with respect to x . For this reason, as well as the shape of the potential energy curve, $U(x)$ is called an infinite potential well. At the bottom of the potential well, $x = 0$, $U = 0$ and the kinetic energy is a maximum, $K = E$, so $v_{\max} = \pm \sqrt{2E/m}$.

However, from the slope of this potential energy curve, you can also deduce information about the force on the glider and its acceleration. We saw earlier that the negative of the slope of the potential energy is the spring force, which in this case is also the net force, and thus is proportional to the acceleration. When $x = 0$, the slope, the force, and the

acceleration are all zero, so this is an **equilibrium point**. The negative of the slope, on either side of the equilibrium point, gives a force pointing back to the equilibrium point, $F = \pm kx$, so the equilibrium is termed stable and the force is called a restoring force. This implies that $U(x)$ has a relative minimum there. If the force on either side of an equilibrium point has a direction opposite from that direction of position change, the equilibrium is termed unstable, and this implies that $U(x)$ has a relative maximum there.

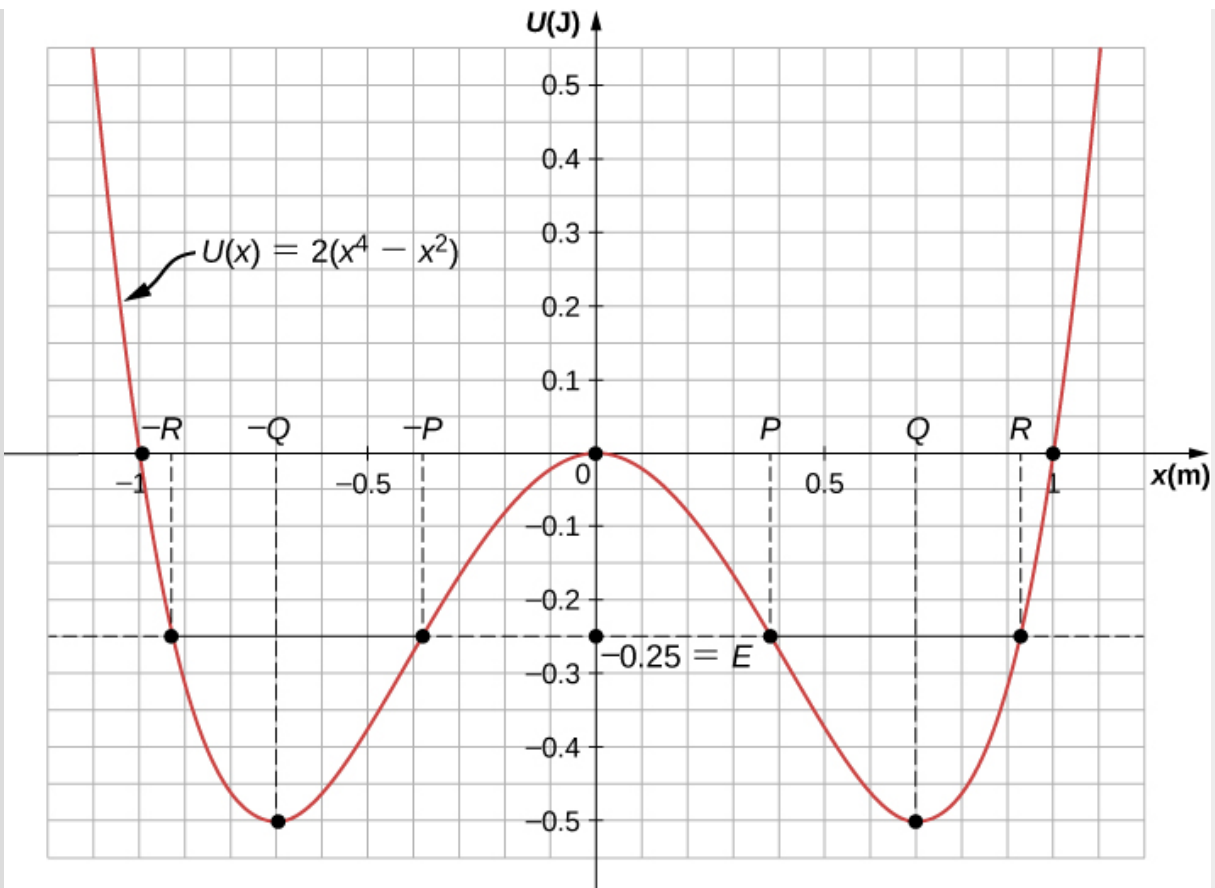
Example:

Quartic and Quadratic Potential Energy Diagram

The potential energy for a particle undergoing one-dimensional motion along the x -axis is $U(x) = 2(x^4 - x^2)$, where U is in joules and x is in meters. The particle is not subject to any non-conservative forces and its mechanical energy is constant at $E = -0.25$ J. (a) Is the motion of the particle confined to any regions on the x -axis, and if so, what are they? (b) Are there any equilibrium points, and if so, where are they and are they stable or unstable?

Strategy

First, we need to graph the potential energy as a function of x . The function is zero at the origin, becomes negative as x increases in the positive or negative directions (x^2 is larger than x^4 for $x < 1$), and then becomes positive at sufficiently large $|x|$. Your graph should look like a double potential well, with the zeros determined by solving the equation $U(x) = 0$, and the extremes determined by examining the first and second derivatives of $U(x)$, as shown in [\[link\]](#).



The potential energy graph for a one-dimensional, quartic and quadratic potential energy, with various quantities indicated.

You can find the values of (a) the allowed regions along the x -axis, for the given value of the mechanical energy, from the condition that the kinetic energy can't be negative, and (b) the equilibrium points and their stability from the properties of the force (stable for a relative minimum and unstable for a relative maximum of potential energy).

You can just eyeball the graph to reach qualitative answers to the questions in this example. That, after all, is the value of potential energy diagrams. You can see that there are two allowed regions for the motion ($E > U$) and three equilibrium points (slope $dU/dx = 0$), of which the central one is unstable ($d^2U/dx^2 < 0$), and the other two are stable ($d^2U/dx^2 > 0$).

Solution

- a. To find the allowed regions for x , we use the condition

Equation:

$$K = E - U = -\frac{1}{4} - 2(x^4 - x^2) \geq 0.$$

If we complete the square in x^2 , this condition simplifies to $2\left(x^2 - \frac{1}{2}\right)^2 \leq \frac{1}{4}$, which we can solve to obtain

Equation:

$$\frac{1}{2} - \sqrt{\frac{1}{8}} \leq x^2 \leq \frac{1}{2} + \sqrt{\frac{1}{8}}.$$

This represents two allowed regions, $x_p \leq x \leq x_R$ and $-x_R \leq x \leq -x_p$, where $x_p = 0.38$ and $x_R = 0.92$ (in meters).

b. To find the equilibrium points, we solve the equation

Equation:

$$dU/dx = 8x^3 - 4x = 0$$

and find $x = 0$ and $x = \pm x_Q$, where $x_Q = 1/\sqrt{2} = 0.707$ (meters). The second derivative

Equation:

$$d^2U/dx^2 = 24x^2 - 4$$

is negative at $x = 0$, so that position is a relative maximum and the equilibrium there is unstable. The second derivative is positive at $x = \pm x_Q$, so these positions are relative minima and represent stable equilibria.

Significance

The particle in this example can oscillate in the allowed region about either of the two stable equilibrium points we found, but it does not have enough energy to escape from whichever potential well it happens to initially be in. The conservation of mechanical energy and the relations between kinetic energy and speed, and potential energy and force, enable you to deduce much information about the qualitative behavior of the motion of a particle, as well as some quantitative information, from a graph of its potential energy.

Note:

Exercise:

Problem:

Check Your Understanding Repeat [\[link\]](#) when the particle's mechanical energy is $+0.25$ J.

Solution:

a. yes, motion confined to $-1.055 \text{ m} \leq x \leq 1.055 \text{ m}$; b. same equilibrium points and types as in example

Before ending this section, let's practice applying the method based on the potential energy of a particle to find its position as a function of time, for the one-dimensional, mass-spring system considered earlier in this section.

Example:

Sinusoidal Oscillations

Find $x(t)$ for a particle moving with a constant mechanical energy $E > 0$ and a potential energy $U(x) = \frac{1}{2}kx^2$, when the particle starts from rest at time $t = 0$.

Strategy

We follow the same steps as we did in [\[link\]](#). Substitute the potential energy U into [\[link\]](#) and factor out the constants, like m or k . Integrate the function and solve the resulting expression for position, which is now a function of time.

Solution

Substitute the potential energy in [\[link\]](#) and integrate using an integral solver found on a web search:

Equation:

$$t = \int_{x_0}^x \frac{dx}{\sqrt{(k/m) [(2E/k) - x^2]}} = \sqrt{\frac{m}{k}} \left[\sin^{-1} \left(\frac{x}{\sqrt{2E/k}} \right) - \sin^{-1} \left(\frac{x_0}{\sqrt{2E/k}} \right) \right].$$

From the initial conditions at $t = 0$, the initial kinetic energy is zero and the initial potential energy is $\frac{1}{2}kx_0^2 = E$, from which you can see that $x_0/\sqrt{(2E/k)} = \pm 1$ and $\sin^{-1}(\pm 1) = \pm 90^\circ$. Now you can solve for x :

Equation:

$$x(t) = \sqrt{(2E/k)} \sin \left[\left(\sqrt{k/m} \right) t \pm 90^\circ \right] = \pm \sqrt{(2E/k)} \cos \left[\left(\sqrt{k/m} \right) t \right].$$

Significance

A few paragraphs earlier, we referred to this mass-spring system as an example of a harmonic oscillator. Here, we anticipate that a harmonic oscillator executes sinusoidal oscillations with a maximum displacement of $\sqrt{(2E/k)}$ (called the amplitude) and a rate of oscillation of $(1/2\pi)\sqrt{k/m}$ (called the frequency). Further discussions about oscillations can be found in [Oscillations](#).

Note:

Exercise:

Problem:

Check Your Understanding Find $x(t)$ for the mass-spring system in [\[link\]](#) if the particle starts from $x_0 = 0$ at $t = 0$. What is the particle's initial velocity?

Solution:

$$x(t) = \pm \sqrt{(2E/k)} \sin \left[\left(\sqrt{k/m} \right) t \right] \text{ and } v_0 = \pm \sqrt{(2E/m)}$$

Summary

- Interpreting a one-dimensional potential energy diagram allows you to obtain qualitative, and some quantitative, information about the motion of a particle.
- At a turning point, the potential energy equals the mechanical energy and the kinetic energy is zero, indicating that the direction of the velocity reverses there.
- The negative of the slope of the potential energy curve, for a particle, equals the one-dimensional component of the conservative force on the particle. At an equilibrium point, the slope is zero and is a stable (unstable) equilibrium for a potential energy minimum (maximum).

Problems

Exercise:

Problem:

A mysterious constant force of 10 N acts horizontally on everything. The direction of the force is found to be always pointed toward a wall in a big hall. Find the potential energy of a particle due to this force when it is at a distance x from the wall, assuming the potential energy at the wall to be zero.

Solution:

10x with x-axis pointed away from the wall and origin at the wall

Exercise:

Problem:

A single force $F(x) = -4.0x$ (in newtons) acts on a 1.0-kg body. When $x = 3.5$ m, the speed of the body is 4.0 m/s. What is its speed at $x = 2.0$ m?

Exercise:**Problem:**

A particle of mass 4.0 kg is constrained to move along the x -axis under a single force $F(x) = -cx^3$, where $c = 8.0 \text{ N/m}^3$. The particle's speed at A, where $x_A = 1.0$ m, is 6.0 m/s. What is its speed at B, where $x_B = -2.0$ m?

Solution:

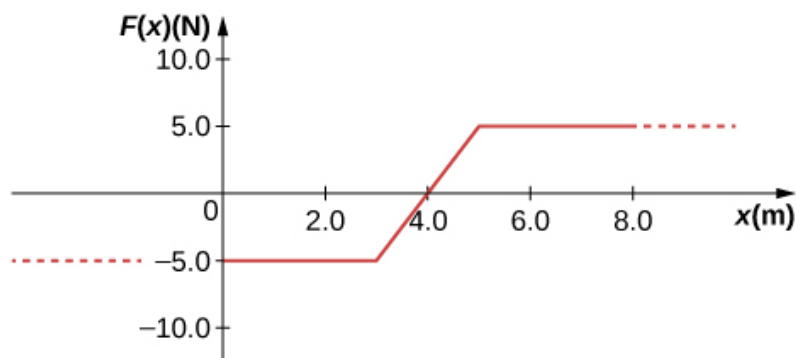
4.6 m/s

Exercise:**Problem:**

The force on a particle of mass 2.0 kg varies with position according to $F(x) = -3.0x^2$ (x in meters, $F(x)$ in newtons). The particle's velocity at $x = 2.0$ m is 5.0 m/s. Calculate the mechanical energy of the particle using (a) the origin as the reference point and (b) $x = 4.0$ m as the reference point. (c) Find the particle's velocity at $x = 1.0$ m. Do this part of the problem for each reference point.

Exercise:**Problem:**

A 4.0-kg particle moving along the x -axis is acted upon by the force whose functional form appears below. The velocity of the particle at $x = 0$ is $v = 6.0$ m/s. Find the particle's speed at $x =$ (a) 2.0 m, (b) 4.0 m, (c) 10.0 m, (d) Does the particle turn around at some point and head back toward the origin? (e) Repeat part (d) if $v = 2.0$ m/s at $x = 0$.

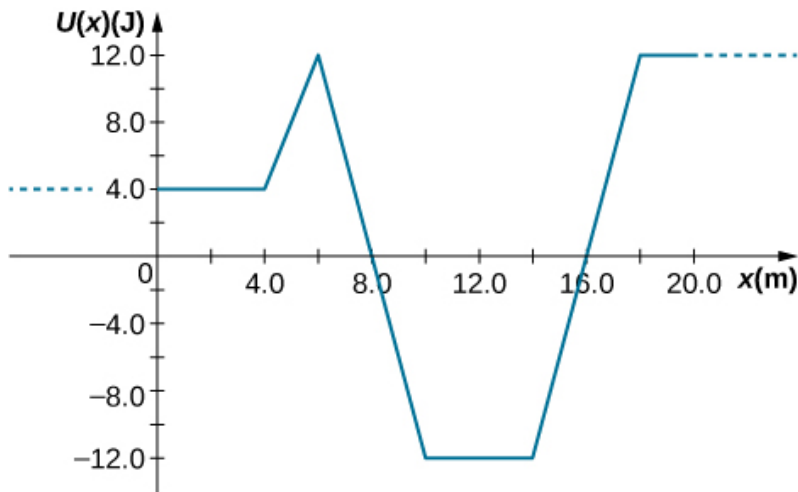


Solution:

a. 5.6 m/s; b. 5.2 m/s; c. 6.4 m/s; d. no; e. yes

Exercise:**Problem:**

A particle of mass 0.50 kg moves along the x -axis with a potential energy whose dependence on x is shown below. (a) What is the force on the particle at $x = 2.0, 5.0, 8.0$, and 12 m? (b) If the total mechanical energy E of the particle is -6.0 J, what are the minimum and maximum positions of the particle? (c) What are these positions if $E = 2.0$ J? (d) If $E = 16$ J, what are the speeds of the particle at the positions listed in part (a)?

**Exercise:****Problem:**

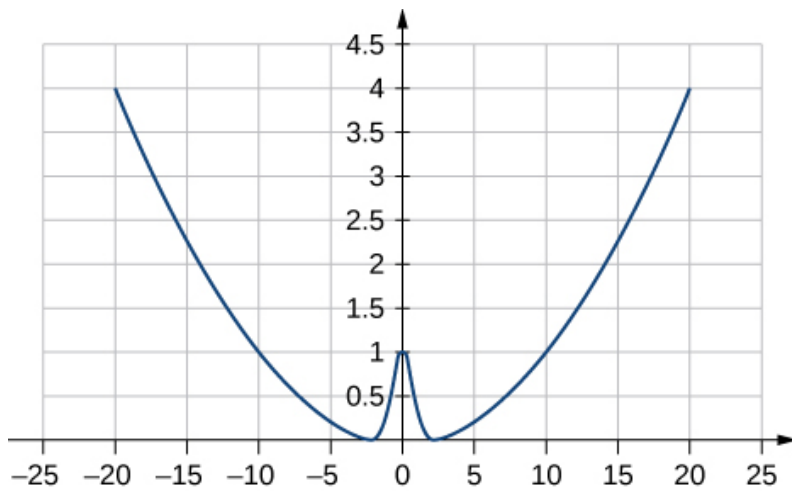
(a) Sketch a graph of the potential energy function $U(x) = kx^2/2 + Ae^{-\alpha x^2}$, where k , A , and α are constants. (b) What is the force corresponding to this potential energy? (c) Suppose a particle of mass m moving with this potential energy has a velocity v_a when its position is $x = a$. Show that the particle does not pass through the origin unless

Equation:

$$A \leq \frac{mv_a^2 + ka^2}{2(1 - e^{-\alpha a^2})}.$$

Solution:

a.



where $k = 0.02$, $A = 1$, $\alpha = 1$; b. $F = kx - \alpha x A e^{-\alpha x^2}$; c. The potential energy at $x = 0$ must be less than the kinetic plus potential energy at $x = a$ or $A \leq \frac{1}{2}mv^2 + \frac{1}{2}ka^2 + Ae^{-\alpha a^2}$. Solving this for A matches results in the problem.

Glossary

equilibrium point

position where the assumed conservative, net force on a particle, given by the slope of its potential energy curve, is zero

potential energy diagram

graph of a particle's potential energy as a function of position

turning point

position where the velocity of a particle, in one-dimensional motion, changes sign

Sources of Energy

By the end of this section, you will be able to:

- Describe energy transformations and conversions in general terms
- Explain what it means for an energy source to be renewable or nonrenewable

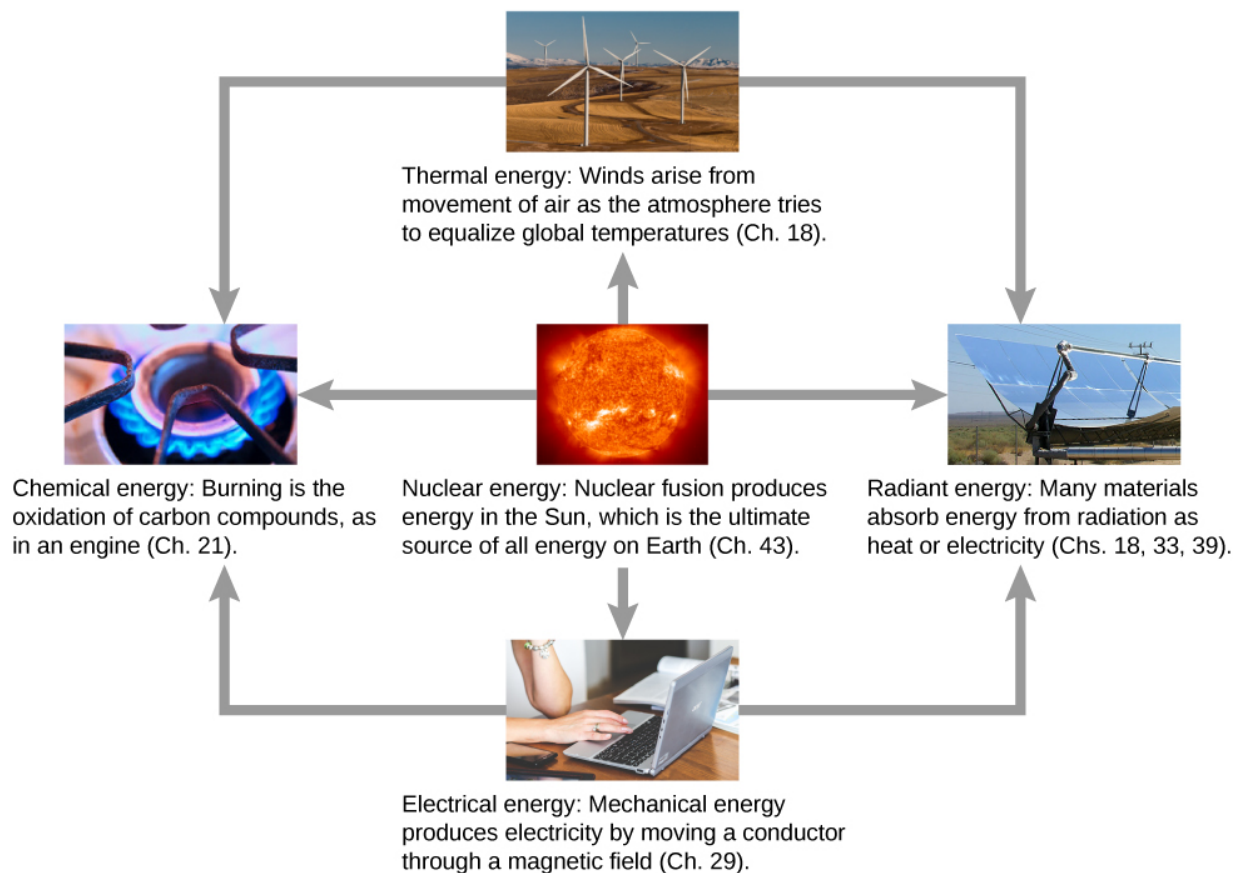
In this chapter, we have studied energy. We learned that energy can take different forms and can be transferred from one form to another. You will find that energy is discussed in many everyday, as well as scientific, contexts, because it is involved in all physical processes. It will also become apparent that many situations are best understood, or most easily conceptualized, by considering energy. So far, no experimental results have contradicted the conservation of energy. In fact, whenever measurements have appeared to conflict with energy conservation, new forms of energy have been discovered or recognized in accordance with this principle.

What are some other forms of energy? Many of these are covered in later chapters (also see [\[link\]](#)), but let's detail a few here:

- Atoms and molecules inside all objects are in random motion. The internal kinetic energy from these random motions is called *thermal energy*, because it is related to the temperature of the object. Note that thermal energy can also be transferred from one place to another, not transformed or converted, by the familiar processes of conduction, convection, and radiation. In this case, the energy is known as *heat energy*.
- *Electrical energy* is a common form that is converted to many other forms and does work in a wide range of practical situations.
- Fuels, such as gasoline and food, have *chemical energy*, which is potential energy arising from their molecular structure. Chemical energy can be converted into thermal energy by reactions like oxidation. Chemical reactions can also produce electrical energy, such as in batteries. Electrical energy can, in turn, produce thermal energy and light, such as in an electric heater or a light bulb.
- Light is just one kind of electromagnetic radiation, or *radiant energy*, which also includes radio, infrared, ultraviolet, X-rays, and gamma

rays. All bodies with thermal energy can radiate energy in electromagnetic waves.

- *Nuclear energy* comes from reactions and processes that convert measurable amounts of mass into energy. Nuclear energy is transformed into radiant energy in the Sun, into thermal energy in the boilers of nuclear power plants, and then into electrical energy in the generators of power plants. These and all other forms of energy can be transformed into one another and, to a certain degree, can be converted into mechanical work.



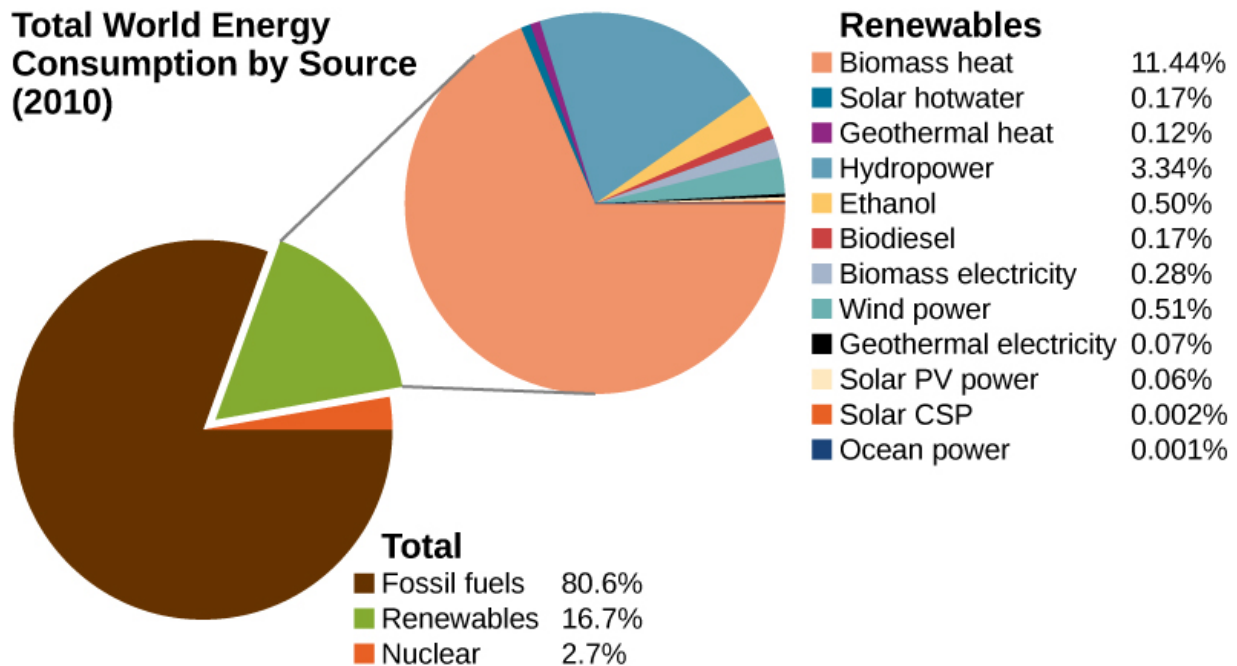
Energy that we use in society takes many forms, which be converted from one into another depending on the process involved. We will study many of these forms of energy in later chapters in this text.
(credit “sun”: modification of work by EIT - SOHO Consortium, ESA, NASA credit “solar panels”: “modification of work by

“kjkolb”/Wikimedia Commons; credit “gas burner”: modification of work by Steven Depolo)

The transformation of energy from one form into another happens all the time. The chemical energy in food is converted into thermal energy through metabolism; light energy is converted into chemical energy through photosynthesis. Another example of energy conversion occurs in a solar cell. Sunlight impinging on a solar cell produces electricity, which can be used to run electric motors or heat water. In an example encompassing many steps, the chemical energy contained in coal is converted into thermal energy as it burns in a furnace, to transform water into steam, in a boiler. Some of the thermal energy in the steam is then converted into mechanical energy as it expands and spins a turbine, which is connected to a generator to produce electrical energy. In these examples, not all of the initial energy is converted into the forms mentioned, because some energy is always transferred to the environment.

Energy is an important element at all levels of society. We live in a very interdependent world, and access to adequate and reliable energy resources is crucial for economic growth and for maintaining the quality of our lives. The principal energy resources used in the world are shown in [\[link\]](#). The figure distinguishes between two major types of energy sources: **renewable** and **non-renewable**, and further divides each type into a few more specific kinds. Renewable sources are energy sources that are replenished through naturally occurring, ongoing processes, on a time scale that is much shorter than the anticipated lifetime of the civilization using the source. Non-renewable sources are depleted once some of the energy they contain is extracted and converted into other kinds of energy. The natural processes by which non-renewable sources are formed typically take place over geological time scales.

Total World Energy Consumption by Source (2010)



World energy consumption by source; the percentage of renewables is increasing, accounting for 19% in 2012.

Our most important non-renewable energy sources are fossil fuels, such as coal, petroleum, and natural gas. These account for about 81% of the world's energy consumption, as shown in the figure. Burning fossil fuels creates chemical reactions that transform potential energy, in the molecular structures of the reactants, into thermal energy and products. This thermal energy can be used to heat buildings or to operate steam-driven machinery. Internal combustion and jet engines convert some of the energy of rapidly expanding gases, released from burning gasoline, into mechanical work. Electrical power generation is mostly derived from transferring energy in expanding steam, via turbines, into mechanical work, which rotates coils of wire in magnetic fields to generate electricity. Nuclear energy is the other non-renewable source shown in [\[link\]](#) and supplies about 3% of the world's consumption. Nuclear reactions release energy by transforming potential energy, in the structure of nuclei, into thermal energy, analogous to energy release in chemical reactions. The thermal energy obtained from nuclear reactions can be transferred and converted into other forms in the same ways that energy from fossil fuels are used.

An unfortunate byproduct of relying on energy produced from the combustion of fossil fuels is the release of carbon dioxide into the atmosphere and its contribution to global warming. Nuclear energy poses environmental problems as well, including the safety and disposal of nuclear waste. Besides these important consequences, reserves of non-renewable sources of energy are limited and, given the rapidly growing rate of world energy consumption, may not last for more than a few hundred years. Considerable effort is going on to develop and expand the use of renewable sources of energy, involving a significant percentage of the world's physicists and engineers.

Four of the renewable energy sources listed in [\[link\]](#)—those using material from plants as fuel (biomass heat, ethanol, biodiesel, and biomass electricity)—involve the same types of energy transformations and conversions as just discussed for fossil and nuclear fuels. The other major types of renewable energy sources are hydropower, wind power, geothermal power, and solar power.

Hydropower is produced by converting the gravitational potential energy of falling or flowing water into kinetic energy and then into work to run electric generators or machinery. Converting the mechanical energy in ocean surface waves and tides is in development. Wind power also converts kinetic energy into work, which can be used directly to generate electricity, operate mills, and propel sailboats.

The interior of Earth has a great deal of thermal energy, part of which is left over from its original formation (gravitational potential energy converted into thermal energy) and part of which is released from radioactive minerals (a form of natural nuclear energy). It will take a very long time for this geothermal energy to escape into space, so people generally regard it as a renewable source, when actually, it's just inexhaustible on human time scales.

The source of solar power is energy carried by the electromagnetic waves radiated by the Sun. Most of this energy is carried by visible light and infrared (heat) radiation. When suitable materials absorb electromagnetic waves, radiant energy is converted into thermal energy, which can be used to heat water, or when concentrated, to make steam and generate electricity

([link](#)). However, in another important physical process, known as the photoelectric effect, energetic radiation impinging on certain materials is directly converted into electricity. Materials that do this are called photovoltaics (PV in [link](#)). Some solar power systems use lenses or mirrors to concentrate the Sun's rays, before converting their energy through photovoltaics, and these are qualified as CSP in [link](#).



Solar cell arrays found in a sunny area converting the solar energy into stored electrical energy. (credit: modification of work by Sarah Swenty, U.S. Fish and Wildlife Service)

As we finish this chapter on energy and work, it is relevant to draw some distinctions between two sometimes misunderstood terms in the area of energy use. As we mentioned earlier, the “law of conservation of energy” is a very useful principle in analyzing physical processes. It cannot be proven from basic principles but is a very good bookkeeping device, and no exceptions have ever been found. It states that the total amount of energy in an isolated system always remains constant. Related to this principle, but remarkably different from it, is the important philosophy of energy conservation. This concept has to do with seeking to decrease the amount of

energy used by an individual or group through reducing activities (e.g., turning down thermostats, driving fewer kilometers) and/or increasing conversion efficiencies in the performance of a particular task, such as developing and using more efficient room heaters, cars that have greater miles-per-gallon ratings, energy-efficient compact fluorescent lights, etc.

Since energy in an isolated system is not destroyed, created, or generated, you might wonder why we need to be concerned about our energy resources, since energy is a conserved quantity. The problem is that the final result of most energy transformations is waste heat, that is, work that has been “degraded” in the energy transformation. We will discuss this idea in more detail in the chapters on thermodynamics.

Summary

- Energy can be transferred from one system to another and transformed or converted from one type into another. Some of the basic types of energy are kinetic, potential, thermal, and electromagnetic.
- Renewable energy sources are those that are replenished by ongoing natural processes, over human time scales. Examples are wind, water, geothermal, and solar power.
- Non-renewable energy sources are those that are depleted by consumption, over human time scales. Examples are fossil fuel and nuclear power.

Key Equations

Difference of potential energy	$\Delta U_{AB} = U_B - U_A = -W_{AB}$
Potential energy with respect to zero of	$\Delta U = U(\vec{r}) - U(\vec{r}_0)$

potential energy at $\vec{\mathbf{r}}_0$	
Gravitational potential energy near Earth's surface	$U(y) = mgy + \text{const.}$
Potential energy for an ideal spring	$U(x) = \frac{1}{2}kx^2 + \text{const.}$
Work done by conservative force over a closed path	$W_{\text{closed path}} = \int \vec{\mathbf{F}}_{\text{cons}} \cdot d\vec{\mathbf{r}} = 0$
Condition for conservative force in two dimensions	$\left(\frac{dF_x}{dy}\right) = \left(\frac{dF_y}{dx}\right)$
Conservative force is the negative derivative of potential energy	$F_l = -\frac{dU}{dl}$
Conservation of energy with no non-conservative forces	$0 = W_{nc,AB} = \Delta(K + U)_{AB} = \Delta E_{AB}.$

Problems

Exercise:

Problem:

In the cartoon movie [Pocahontas](#), Pocahontas runs to the edge of a cliff and jumps off, showcasing the fun side of her personality. (a) If she is running at 3.0 m/s before jumping off the cliff and she hits the water at the bottom of the cliff at 20.0 m/s, how high is the cliff? Assume negligible air drag in this cartoon. (b) If she jumped off the same cliff from a standstill, how fast would she be falling right before she hit the water?

Exercise:**Problem:**

In the reality television show [“Amazing Race”](#), a contestant is firing 12-kg watermelons from a slingshot to hit targets down the field. The slingshot is pulled back 1.5 m and the watermelon is considered to be at ground level. The launch point is 0.3 m from the ground and the targets are 10 m horizontally away. Calculate the spring constant of the slingshot.

Solution:

8700 N/m

Exercise:**Problem:**

In the [Back to the Future movies](#), a DeLorean car of mass 1230 kg travels at 88 miles per hour to venture back to the future. (a) What is the kinetic energy of the DeLorean? (b) What spring constant would be needed to stop this DeLorean in a distance of 0.1m?

Exercise:

Problem:

In the [Hunger Games movie](#), Katniss Everdeen fires a 0.0200-kg arrow from ground level to pierce an apple up on a stage. The spring constant of the bow is 330 N/m and she pulls the arrow back a distance of 0.55 m. The apple on the stage is 5.00 m higher than the launching point of the arrow. At what speed does the arrow (a) leave the bow? (b) strike the apple?

Solution:

a. 70.6 m/s; b. 69.9 m/s

Exercise:**Problem:**

In a [“Top Fail” video](#), two women run at each other and collide by hitting exercise balls together. If each woman has a mass of 50 kg, which includes the exercise ball, and one woman runs to the right at 2.0 m/s and the other is running toward her at 1.0 m/s, (a) how much total kinetic energy is there in the system? (b) If energy is conserved after the collision and each exercise ball has a mass of 2.0 kg, how fast would the balls fly off toward the camera?

Exercise:**Problem:**

In a [Coyote/Road Runner cartoon clip](#), a spring expands quickly and sends the coyote into a rock. If the spring extended 5 m and sent the coyote of mass 20 kg to a speed of 15 m/s, (a) what is the spring constant of this spring? (b) If the coyote were sent vertically into the air with the energy given to him by the spring, how high could he go if there were no non-conservative forces?

Solution:

a. 180 N/m; b. 11 m

Exercise:**Problem:**

In an iconic movie scene, [Forrest Gump](#) runs around the country. If he is running at a constant speed of 3 m/s, would it take him more or less energy to run uphill or downhill and why?

Exercise:**Problem:**

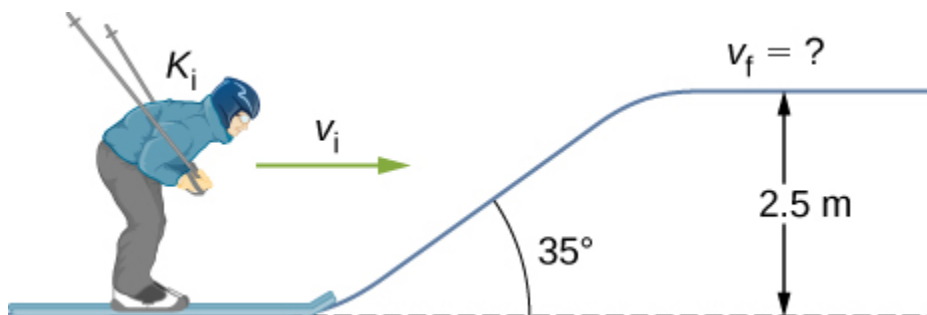
In the movie [Monty Python and the Holy Grail](#) a cow is catapulted from the top of a castle wall over to the people down below. The gravitational potential energy is set to zero at ground level. The cow is launched from a spring of spring constant $1.1 \times 10^4 \text{ N/m}$ that is expanded 0.5 m from equilibrium. If the castle is 9.1 m tall and the mass of the cow is 110 kg, (a) what is the gravitational potential energy of the cow at the top of the castle? (b) What is the elastic spring energy of the cow before the catapult is released? (c) What is the speed of the cow right before it lands on the ground?

Solution:

a. $9.8 \times 10^3 \text{ J}$; b. $1.4 \times 10^3 \text{ J}$; c. 14 m/s

Exercise:**Problem:**

A 60.0-kg skier with an initial speed of 12.0 m/s coasts up a 2.50-m high rise as shown. Find her final speed at the top, given that the coefficient of friction between her skis and the snow is 0.80.



Exercise:**Problem:**

(a) How high a hill can a car coast up (engines disengaged) if work done by friction is negligible and its initial speed is 110 km/h? (b) If, in actuality, a 750-kg car with an initial speed of 110 km/h is observed to coast up a hill to a height 22.0 m above its starting point, how much thermal energy was generated by friction? (c) What is the average force of friction if the hill has a slope of 2.5° above the horizontal?

Solution:

a. 47.6 m; b. 1.88×10^5 J; c. 373 N

Exercise:**Problem:**

A 5.00×10^5 -kg subway train is brought to a stop from a speed of 0.500 m/s in 0.400 m by a large spring bumper at the end of its track. What is the spring constant k of the spring?

Exercise:**Problem:**

A pogo stick has a spring with a spring constant of 2.5×10^4 N/m, which can be compressed 12.0 cm. To what maximum height from the uncompressed spring can a child jump on the stick using only the energy in the spring, if the child and stick have a total mass of 40 kg?

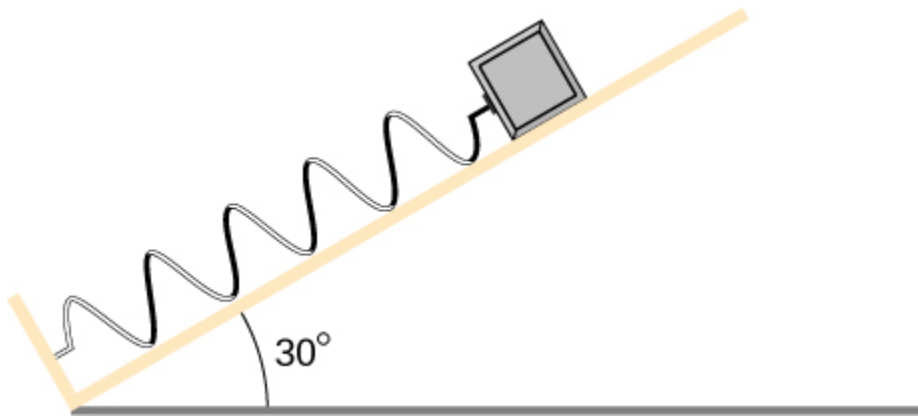
Solution:

33.9 cm

Exercise:

Problem:

A block of mass 500 g is attached to a spring of spring constant 80 N/m (see the following figure). The other end of the spring is attached to a support while the mass rests on a rough surface with a coefficient of friction of 0.20 that is inclined at angle of 30° . The block is pushed along the surface till the spring compresses by 10 cm and is then released from rest. (a) How much potential energy was stored in the block-spring-support system when the block was just released? (b) Determine the speed of the block when it crosses the point when the spring is neither compressed nor stretched. (c) Determine the position of the block where it just comes to rest on its way up the incline.

**Exercise:****Problem:**

A block of mass 200 g is attached at the end of a massless spring at equilibrium length of spring constant 50 N/m. The other end of the spring is attached to the ceiling and the mass is released at a height considered to be where the gravitational potential energy is zero. (a) What is the net potential energy of the block at the instant the block is at the lowest point? (b) What is the net potential energy of the block at the midpoint of its descent? (c) What is the speed of the block at the midpoint of its descent?

Solution:

a. Zero, since the total energy of the system is zero and the kinetic energy at the lowest point is zero; b. -0.038 J ; c. 0.62 m/s

Exercise:

Problem:

A T-shirt cannon launches a shirt at 5.00 m/s from a platform height of 3.00 m from ground level. How fast will the shirt be traveling if it is caught by someone whose hands are (a) 1.00 m from ground level? (b) 4.00 m from ground level? Neglect air drag.

Exercise:

Problem:

A child (32 kg) jumps up and down on a trampoline. The trampoline exerts a spring restoring force on the child with a constant of 5000 N/m . At the highest point of the bounce, the child is 1.0 m above the level surface of the trampoline. What is the compression distance of the trampoline? Neglect the bending of the legs or any transfer of energy of the child into the trampoline while jumping.

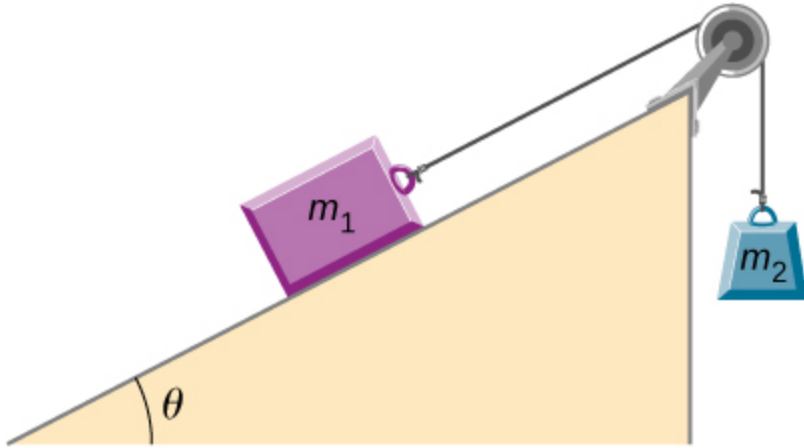
Solution:

42 cm

Exercise:

Problem:

Shown below is a box of mass m_1 that sits on a frictionless incline at an angle above the horizontal $\theta = 30^\circ$. This box is connected by a relatively massless string, over a frictionless pulley, and finally connected to a box at rest over the ledge, labeled m_2 . If m_1 and m_2 are a height h above the ground and $m_2 \gg m_1$: (a) What is the initial gravitational potential energy of the system? (b) What is the final kinetic energy of the system?



Additional Problems

Exercise:

Problem:

A massless spring with force constant $k = 200 \text{ N/m}$ hangs from the ceiling. A 2.0-kg block is attached to the free end of the spring and released. If the block falls 17 cm before starting back upwards, how much work is done by friction during its descent?

Solution:

-0.44 J

Exercise:

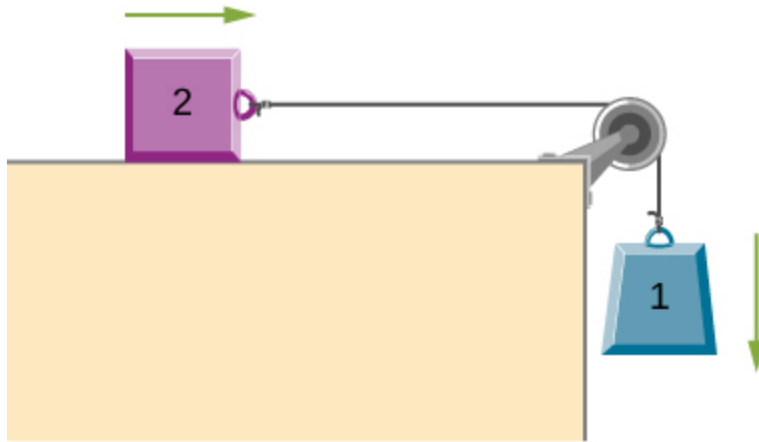
Problem:

A particle of mass 2.0 kg moves under the influence of the force $F(x) = (-5x^2 + 7x) \text{ N}$. Suppose a frictional force also acts on the particle. If the particle's speed when it starts at $x = -4.0 \text{ m}$ is 0.0 m/s and when it arrives at $x = 4.0 \text{ m}$ is 9.0 m/s , how much work is done on it by the frictional force between $x = -4.0 \text{ m}$ and $x = 4.0 \text{ m}$?

Exercise:

Problem:

Block 2 shown below slides along a frictionless table as block 1 falls. Both blocks are attached by a frictionless pulley. Find the speed of the blocks after they have each moved 2.0 m. Assume that they start at rest and that the pulley has negligible mass. Use $m_1 = 2.0$ kg and $m_2 = 4.0$ kg.

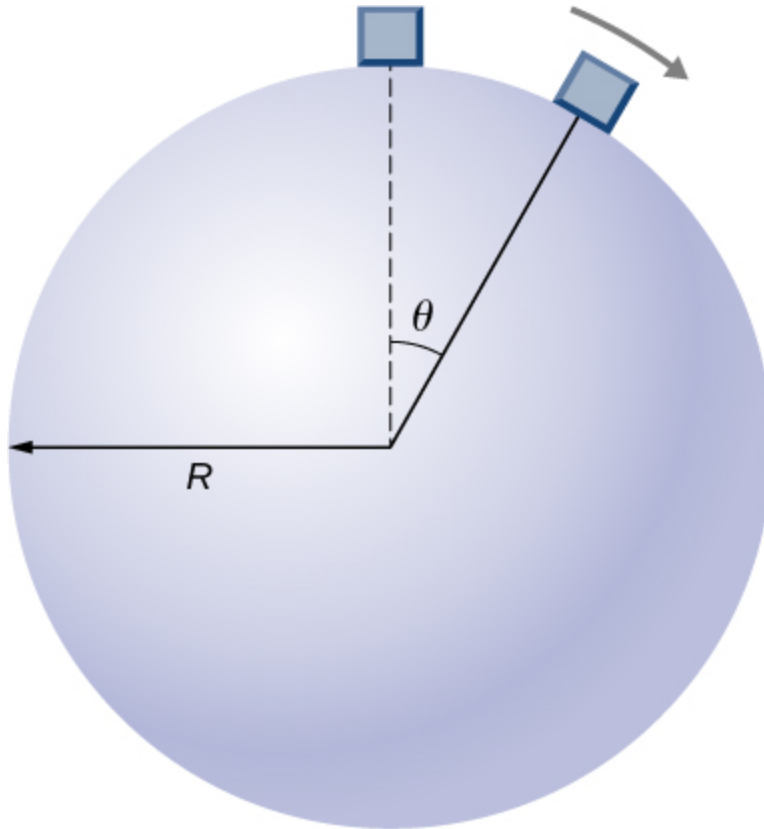


Solution:

3.6 m/s

Exercise:**Problem:**

A body of mass m and negligible size starts from rest and slides down the surface of a frictionless solid sphere of radius R . (See below.) Prove that the body leaves the sphere when $\theta = \cos^{-1}(2/3)$.



Exercise:

Problem:

A mysterious force acts on all particles along a particular line and always points towards a particular point P on the line. The magnitude of the force on a particle increases as the cube of the distance from that point; that is $F \propto r^3$, if the distance from P to the position of the particle is r . Let b be the proportionality constant, and write the magnitude of the force as $F = br^3$. Find the potential energy of a particle subjected to this force when the particle is at a distance D from P , assuming the potential energy to be zero when the particle is at P .

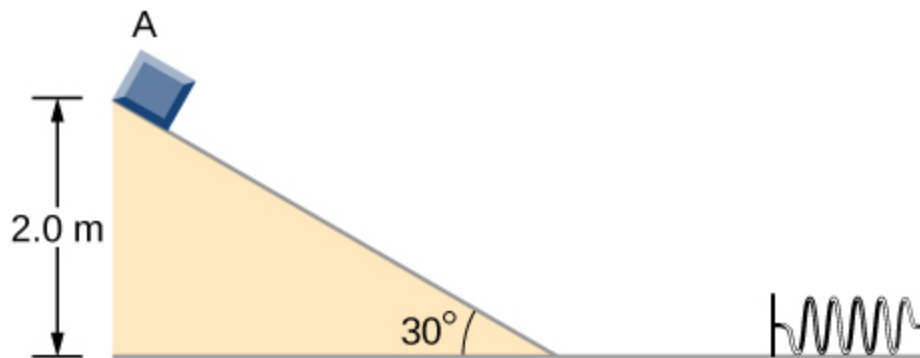
Solution:

$$bD^4/4$$

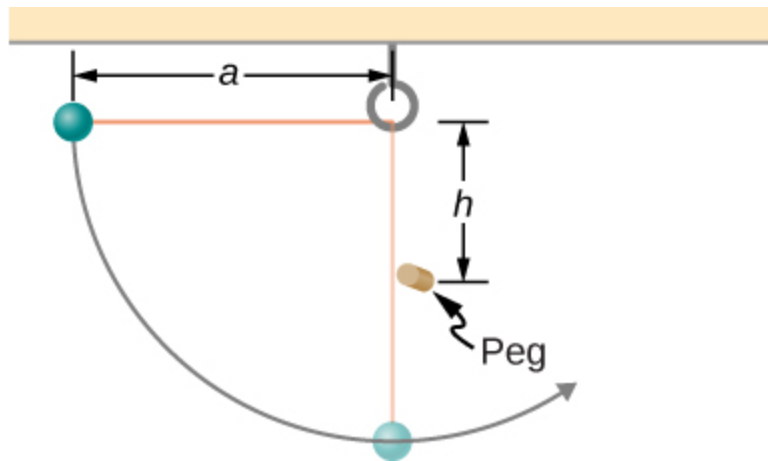
Exercise:

Problem:

An object of mass 10 kg is released at point A, slides to the bottom of the 30° incline, then collides with a horizontal massless spring, compressing it a maximum distance of 0.75 m. (See below.) The spring constant is 500 N/m, the height of the incline is 2.0 m, and the horizontal surface is frictionless. (a) What is the speed of the object at the bottom of the incline? (b) What is the work of friction on the object while it is on the incline? (c) The spring recoils and sends the object back toward the incline. What is the speed of the object when it reaches the base of the incline? (d) What vertical distance does it move back up the incline?

**Exercise:****Problem:**

Shown below is a small ball of mass m attached to a string of length a . A small peg is located a distance h below the point where the string is supported. If the ball is released when the string is horizontal, show that h must be greater than $3a/5$ if the ball is to swing completely around the peg.



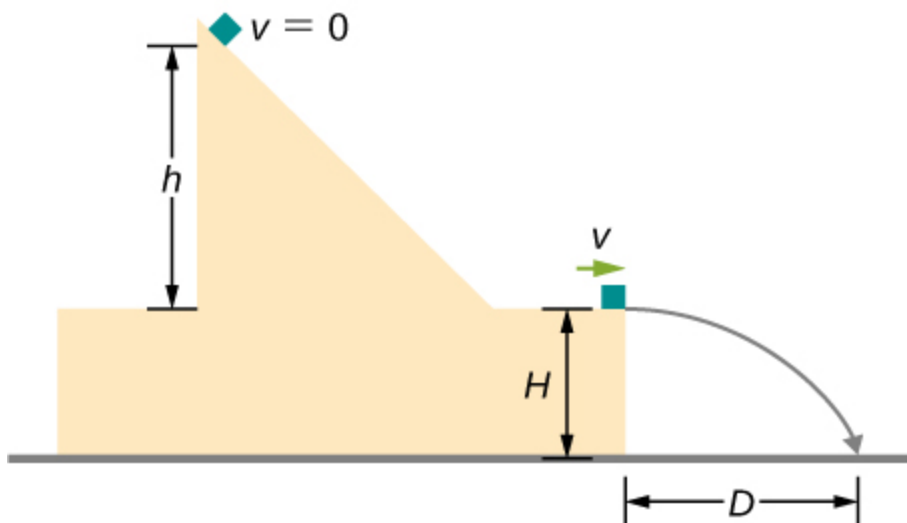
Solution:

proof

Exercise:

Problem:

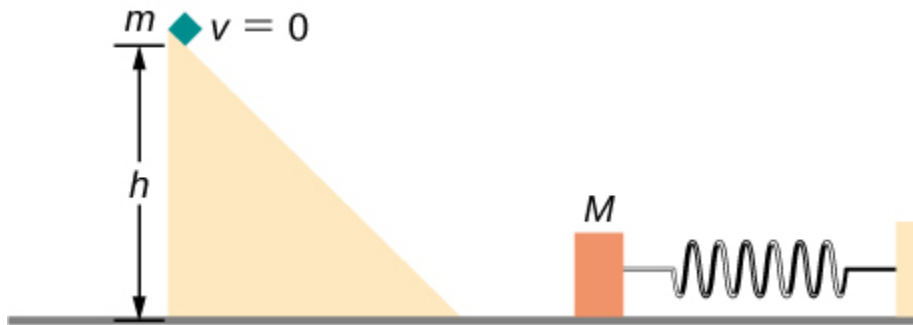
A block leaves a frictionless inclined surface horizontally after dropping off by a height h . Find the horizontal distance D where it will land on the floor, in terms of h , H , and g .



Exercise:

Problem:

A block of mass m , after sliding down a frictionless incline, strikes another block of mass M that is attached to a spring of spring constant k (see below). The blocks stick together upon impact and travel together. (a) Find the compression of the spring in terms of m , M , h , g , and k when the combination comes to rest. Hint: The speed of the combined blocks $m + M$ (v_2) is based on the speed of block m just prior to the collision with the block M (v_1) based on the equation $v_2 = (m/m) + M (v_1)$. This will be discussed further in the chapter on Linear Momentum and Collisions. (b) The loss of kinetic energy as a result of the bonding of the two masses upon impact is stored in the so-called binding energy of the two masses. Calculate the binding energy.



Solution:

a. $\sqrt{\frac{2m^2gh}{k(m+M)}}$; b. $\frac{mMgh}{m+M}$

Exercise:

Problem:

A block of mass 300 g is attached to a spring of spring constant 100 N/m. The other end of the spring is attached to a support while the block rests on a smooth horizontal table and can slide freely without any friction. The block is pushed horizontally till the spring compresses by 12 cm, and then the block is released from rest. (a) How much potential energy was stored in the block-spring support system when the block was just released? (b) Determine the speed of the block when it crosses the point when the spring is neither compressed nor stretched. (c) Determine the speed of the block when it has traveled a distance of 20 cm from where it was released.

Exercise:**Problem:**

Consider a block of mass 0.200 kg attached to a spring of spring constant 100 N/m. The block is placed on a frictionless table, and the other end of the spring is attached to the wall so that the spring is level with the table. The block is then pushed in so that the spring is compressed by 10.0 cm. Find the speed of the block as it crosses (a) the point when the spring is not stretched, (b) 5.00 cm to the left of point in (a), and (c) 5.00 cm to the right of point in (a).

Solution:

a. 2.24 m/s; b. 1.94 m/s; c. 1.94 m/s

Exercise:**Problem:**

A skier starts from rest and slides downhill. What will be the speed of the skier if he drops by 20 meters in vertical height? Ignore any air resistance (which will, in reality, be quite a lot), and any friction between the skis and the snow.

Exercise:

Problem:

Repeat the preceding problem, but this time, suppose that the work done by air resistance cannot be ignored. Let the work done by the air resistance when the skier goes from A to B along the given hilly path be -2000 J . The work done by air resistance is negative since the air resistance acts in the opposite direction to the displacement. Supposing the mass of the skier is 50 kg , what is the speed of the skier at point B ?

Solution:

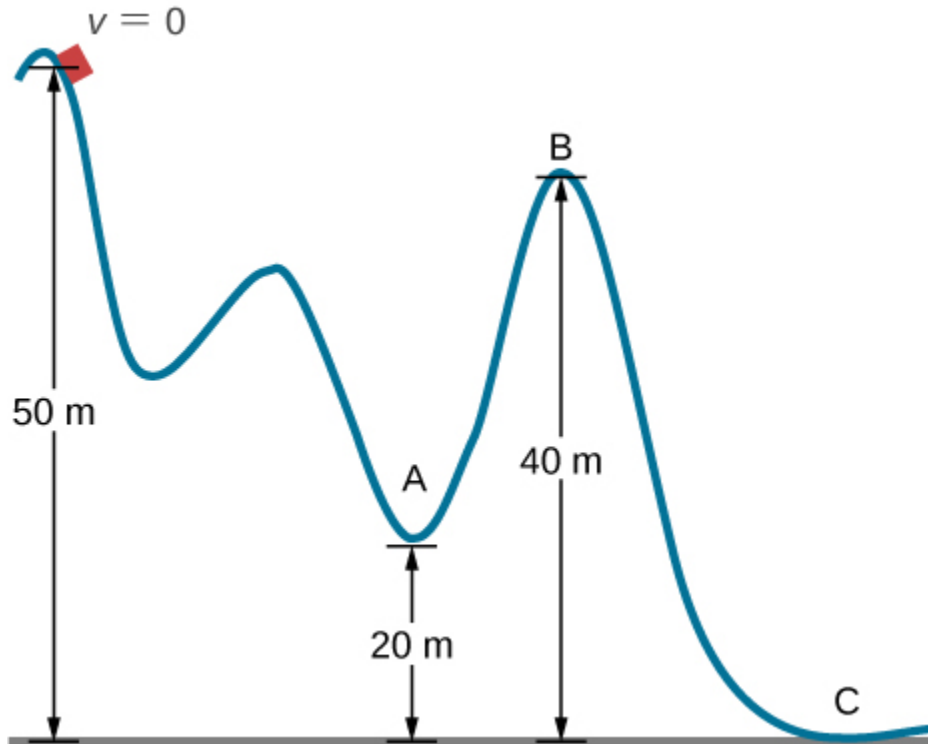
18 m/s

Exercise:**Problem:**

Two bodies are interacting by a conservative force. Show that the mechanical energy of an isolated system consisting of two bodies interacting with a conservative force is conserved. (*Hint:* Start by using Newton's third law and the definition of work to find the work done on each body by the conservative force.)

Exercise:**Problem:**

In an amusement park, a car rolls in a track as shown below. Find the speed of the car at A , B , and C . Note that the work done by the rolling friction is zero since the displacement of the point at which the rolling friction acts on the tires is momentarily at rest and therefore has a zero displacement.



Solution:

$$v_A = 24 \text{ m/s}; v_B = 14 \text{ m/s}; v_C = 31 \text{ m/s}$$

Exercise:

Problem:

A 200-g steel ball is tied to a 2.00-m “massless” string and hung from the ceiling to make a pendulum, and then, the ball is brought to a position making a 30° angle with the vertical direction and released from rest. Ignoring the effects of the air resistance, find the speed of the ball when the string (a) is vertically down, (b) makes an angle of 20° with the vertical and (c) makes an angle of 10° with the vertical.

Exercise:

Problem:

A 300 g hockey puck is shot across an ice-covered pond. Before the hockey puck was hit, the puck was at rest. After the hit, the puck has a speed of 40 m/s. The puck comes to rest after going a distance of 30 m. (a) Describe how the energy of the puck changes over time, giving the numerical values of any work or energy involved. (b) Find the magnitude of the net friction force.

Solution:

a. Loss of energy is $240 \text{ N} \cdot \text{m}$; b. $F = 8 \text{ N}$

Exercise:**Problem:**

A projectile of mass 2 kg is fired with a speed of 20 m/s at an angle of 30° with respect to the horizontal. (a) Calculate the initial total energy of the projectile given that the reference point of zero gravitational potential energy at the launch position. (b) Calculate the kinetic energy at the highest vertical position of the projectile. (c) Calculate the gravitational potential energy at the highest vertical position. (d) Calculate the maximum height that the projectile reaches. Compare this result by solving the same problem using your knowledge of projectile motion.

Exercise:**Problem:**

An artillery shell is fired at a target 200 m above the ground. When the shell is 100 m in the air, it has a speed of 100 m/s. What is its speed when it hits its target? Neglect air friction.

Solution:

89.7 m/s

Exercise:

Problem:

How much energy is lost to a dissipative drag force if a 60-kg person falls at a constant speed for 15 meters?

Exercise:**Problem:**

A box slides on a frictionless surface with a total energy of 50 J. It hits a spring and compresses the spring a distance of 25 cm from equilibrium. If the same box with the same initial energy slides on a rough surface, it only compresses the spring a distance of 15 cm, how much energy must have been lost by sliding on the rough surface?

Solution:

32 J

Glossary

non-renewable

energy source that is not renewable, but is depleted by human consumption

renewable

energy source that is replenished by natural processes, over human time scales

Introduction

class="introduction"

The
concepts of
impulse,
momentum,
and center
of mass are
crucial for a
major-
league
baseball
player to
successfully
get a hit. If
he
misjudges
these
quantities,
he might
break his
bat instead.
(credit:
modification
n of work
by “Cathy
T”/Flickr)



The concepts of work, energy, and the work-energy theorem are valuable for two primary reasons: First, they are powerful computational tools, making it much easier to analyze complex physical systems than is possible using Newton's laws directly (for example, systems with nonconstant forces); and second, the observation that the total energy of a closed system is conserved means that the system can only evolve in ways that are consistent with energy conservation. In other words, a system cannot evolve randomly; it can only change in ways that conserve energy.

In this chapter, we develop and define another conserved quantity, called *linear momentum*, and another relationship (the *impulse-momentum theorem*), which will put an additional constraint on how a system evolves in time. Conservation of momentum is useful for understanding collisions, such as that shown in the above image. It is just as powerful, just as important, and just as useful as conservation of energy and the work-energy theorem.

Linear Momentum

By the end of this section, you will be able to:

- Explain what momentum is, physically
- Calculate the momentum of a moving object

Our study of kinetic energy showed that a complete understanding of an object's motion must include both its mass and its velocity ($K = (1/2)mv^2$). However, as powerful as this concept is, it does not include any information about the direction of the moving object's velocity vector. We'll now define a physical quantity that includes direction.

Like kinetic energy, this quantity includes both mass and velocity; like kinetic energy, it is a way of characterizing the “quantity of motion” of an object. It is given the name **momentum** (from the Latin word *movimentum*, meaning “movement”), and it is represented by the symbol p .

Note:

Momentum

The momentum p of an object is the product of its mass and its velocity:

Equation:

$$\vec{p} = m\vec{v}.$$

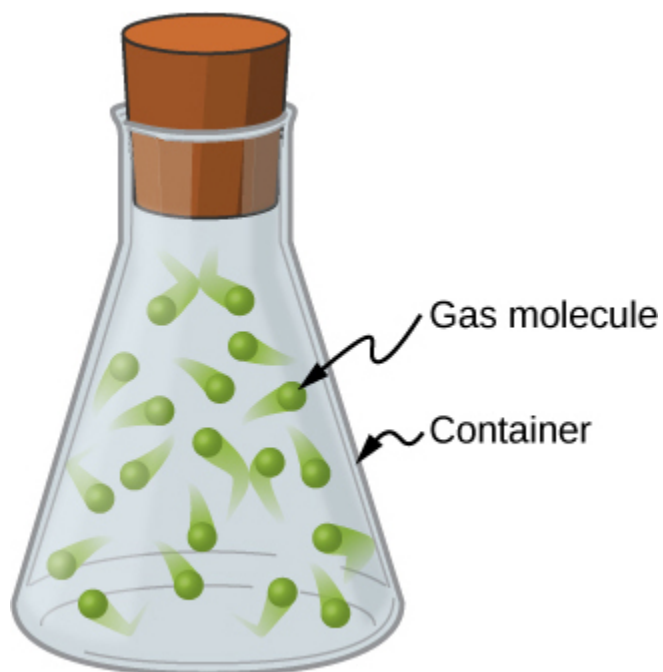


The velocity and momentum vectors for the ball are in the same direction. The mass of the ball is about 0.5 kg, so the momentum vector is about half the length of the velocity vector because momentum is velocity time mass. (credit: modification of work by Ben Sutherland)

As shown in [\[link\]](#), momentum is a vector quantity (since velocity is). This is one of the things that makes momentum useful and not a duplication of kinetic energy. It is perhaps most useful when determining whether an object's motion is difficult to change ([\[link\]](#)) or easy to change ([\[link\]](#)).



This supertanker transports a huge mass of oil; as a consequence, it takes a long time for a force to change its (comparatively small) velocity. (credit: modification of work by “the_tahoe_guy”/Flickr)



Gas molecules can have very large velocities, but these velocities change nearly instantaneously when they collide with the container walls or with each other. This is primarily because their masses are so tiny.

Unlike kinetic energy, momentum depends equally on an object's mass and velocity. For example, as you will learn when you study thermodynamics, the average speed of an air molecule at room temperature is approximately 500 m/s, with an average molecular mass of 6×10^{-25} kg; its momentum is thus

Equation:

$$p_{\text{molecule}} = (6 \times 10^{-25} \text{ kg}) \left(500 \frac{\text{m}}{\text{s}} \right) = 3 \times 10^{-22} \frac{\text{kg} \cdot \text{m}}{\text{s}}.$$

For comparison, a typical automobile might have a speed of only 15 m/s, but a mass of 1400 kg, giving it a momentum of

Equation:

$$p_{\text{car}} = (1400 \text{ kg}) \left(15 \frac{\text{m}}{\text{s}} \right) = 21,000 \frac{\text{kg} \cdot \text{m}}{\text{s}}.$$

These momenta are different by 27 orders of magnitude, or a factor of a billion billion billion!

Summary

- The motion of an object depends on its mass as well as its velocity. Momentum is a concept that describes this. It is a useful and powerful concept, both computationally and theoretically. The SI unit for momentum is $\text{kg} \cdot \text{m/s}$.

Conceptual Questions

Exercise:

Problem:

An object that has a small mass and an object that has a large mass have the same momentum. Which object has the largest kinetic energy?

Solution:

Since $K = p^2/2m$, then if the momentum is fixed, the object with smaller mass has more kinetic energy.

Exercise:

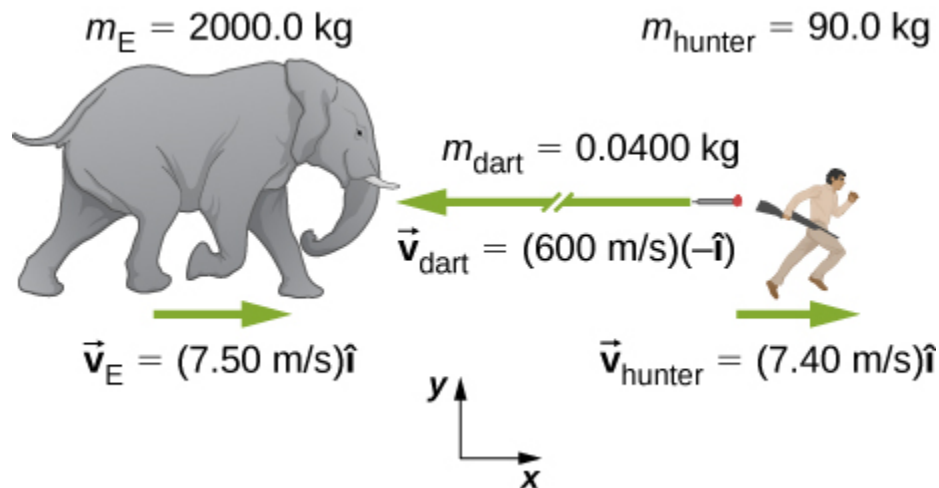
Problem:

An object that has a small mass and an object that has a large mass have the same kinetic energy. Which mass has the largest momentum?

Problems

Exercise:

Problem: An elephant and a hunter are having a confrontation.



- Calculate the momentum of the 2000.0-kg elephant charging the hunter at a speed of 7.50 m/s.
- Calculate the ratio of the elephant's momentum to the momentum of a 0.0400-kg tranquilizer dart fired at a speed of 600 m/s.
- What is the momentum of the 90.0-kg hunter running at 7.40 m/s after missing the elephant?

Exercise:

Problem:

A skater of mass 40 kg is carrying a box of mass 5 kg. The skater has a speed of 5 m/s with respect to the floor and is gliding without any friction on a smooth surface.

- Find the momentum of the box with respect to the floor.
- Find the momentum of the box with respect to the floor after she puts the box down on the frictionless skating surface.

Solution:

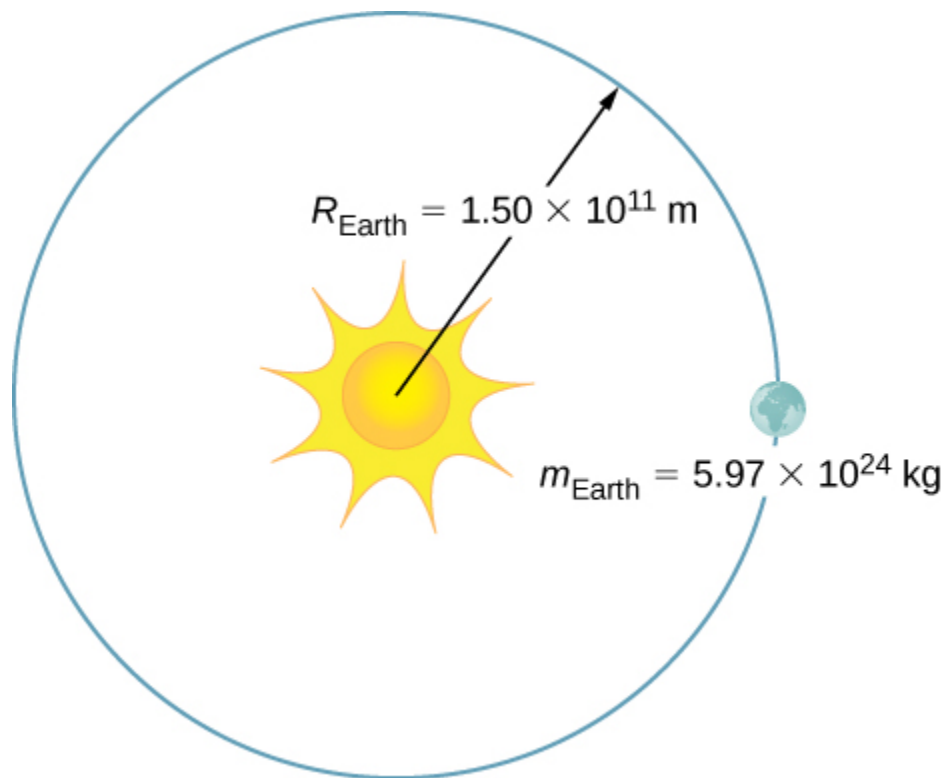
a. magnitude: $25 \text{ kg} \cdot \text{m/s}$; b. same as a.

Exercise:**Problem:**

A car of mass 2000 kg is moving with a constant velocity of 10 m/s due east. What is the momentum of the car?

Exercise:**Problem:**

The mass of Earth is $5.97 \times 10^{24} \text{ kg}$ and its orbital radius is an average of $1.50 \times 10^{11} \text{ m}$. Calculate the magnitude of its linear momentum at the location in the diagram.



Solution:

$$1.78 \times 10^{29} \text{ kg} \cdot \text{m/s}$$

Exercise:**Problem:**

If a rainstorm drops 1 cm of rain over an area of 10 km^2 in the period of 1 hour, what is the momentum of the rain that falls in one second? Assume the terminal velocity of a raindrop is 10 m/s.

Exercise:**Problem:**

What is the average momentum of an avalanche that moves a 40-cm-thick layer of snow over an area of 100 m by 500 m over a distance of 1 km down a hill in 5.5 s? Assume a density of 350 kg/m^3 for the snow.

Solution:

$$1.3 \times 10^9 \text{ kg} \cdot \text{m/s}$$

Exercise:**Problem:**

What is the average momentum of a 70.0-kg sprinter who runs the 100-m dash in 9.65 s?

Glossary**momentum**

measure of the quantity of motion that an object has; it takes into account both how fast the object is moving, and its mass; specifically, it is the product of mass and velocity; it is a vector quantity

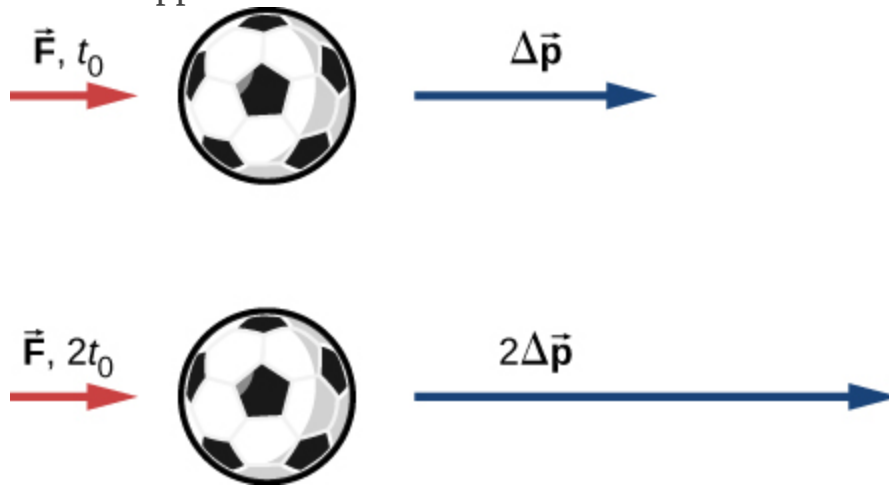
Impulse and Collisions

By the end of this section, you will be able to:

- Explain what an impulse is, physically
- Describe what an impulse does
- Relate impulses to collisions
- Apply the impulse-momentum theorem to solve problems

We have defined momentum to be the product of mass and velocity. Therefore, if an object's velocity should change (due to the application of a force on the object), then necessarily, its momentum changes as well. This indicates a connection between momentum and force. The purpose of this section is to explore and describe that connection.

Suppose you apply a force on a free object for some amount of time. Clearly, the larger the force, the larger the object's change of momentum will be. Alternatively, the more time you spend applying this force, again the larger the change of momentum will be, as depicted in [\[link\]](#). The amount by which the object's motion changes is therefore proportional to the magnitude of the force, and also to the time interval over which the force is applied.



The change in momentum of an object is proportional to the length of time during which the force is applied. If a force is exerted on the lower ball for twice as long as on the upper ball,

then the change in the momentum of the lower ball is twice that of the upper ball.

Mathematically, if a quantity is proportional to two (or more) things, then it is proportional to the product of those things. The product of a force and a time interval (over which that force acts) is called **impulse**, and is given the symbol \vec{J} .

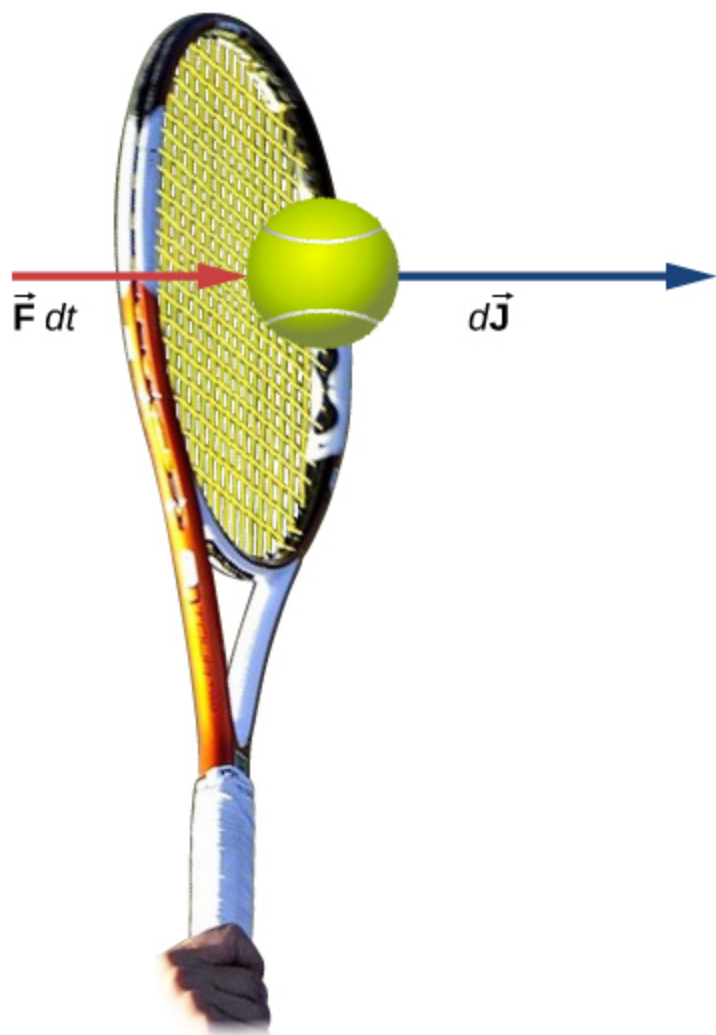
Note:

Impulse

Let $\vec{F}(t)$ be the force applied to an object over some differential time interval dt ([link](#)). The resulting impulse on the object is defined as

Equation:

$$d\vec{J} \equiv \vec{F}(t)dt.$$



A force applied by a tennis racquet to a tennis ball over a time interval generates an impulse acting on the ball.

The total impulse over the interval $t_f - t_i$ is

Note:
Equation:

$$\vec{\mathbf{J}} = \int_{t_i}^{t_f} d\vec{\mathbf{J}} \text{ or } \vec{\mathbf{J}} \equiv \int_{t_i}^{t_f} \vec{\mathbf{F}}(t)dt.$$

[\[link\]](#) and [\[link\]](#) together say that when a force is applied for an infinitesimal time interval dt , it causes an infinitesimal impulse $d\vec{\mathbf{J}}$, and the total impulse given to the object is defined to be the sum (integral) of all these infinitesimal impulses.

To calculate the impulse using [\[link\]](#), we need to know the force function $F(t)$, which we often don't. However, a result from calculus is useful here: Recall that the average value of a function over some interval is calculated by

Equation:

$$f(x)_{\text{ave}} = \frac{1}{\Delta x} \int_{x_i}^{x_f} f(x)dx$$

where $\Delta x = x_f - x_i$. Applying this to the time-dependent force function, we obtain

Equation:

$$\vec{\mathbf{F}}_{\text{ave}} = \frac{1}{\Delta t} \int_{t_i}^{t_f} \vec{\mathbf{F}}(t)dt.$$

Therefore, from [\[link\]](#),

Note:

Equation:

$$\vec{\mathbf{J}} = \vec{\mathbf{F}}_{\text{ave}}\Delta t.$$

The idea here is that you can calculate the impulse on the object even if you don't know the details of the force as a function of time; you only need the average force. In fact, though, the process is usually reversed: You determine the impulse (by measurement or calculation) and then calculate the average force that caused that impulse.

To calculate the impulse, a useful result follows from writing the force in [\[link\]](#) as $\vec{\mathbf{F}}(t) = m\vec{\mathbf{a}}(t)$:

Equation:

$$\vec{\mathbf{J}} = \int_{t_i}^{t_f} \vec{\mathbf{F}}(t) dt = m \int_{t_i}^{t_f} \vec{\mathbf{a}}(t) dt = m [\vec{\mathbf{v}}(t_f) - \vec{\mathbf{v}}_i].$$

For a constant force $\vec{\mathbf{F}}_{\text{ave}} = \vec{\mathbf{F}} = m\vec{\mathbf{a}}$, this simplifies to

Equation:

$$\vec{\mathbf{J}} = m\vec{\mathbf{a}}\Delta t = m\vec{\mathbf{v}}_f - m\vec{\mathbf{v}}_i = m(\vec{\mathbf{v}}_f - \vec{\mathbf{v}}_i).$$

That is,

Equation:

$$\vec{\mathbf{J}} = m\Delta\vec{\mathbf{v}}.$$

Note that the integral form, [\[link\]](#), applies to constant forces as well; in that case, since the force is independent of time, it comes out of the integral, which can then be trivially evaluated.

Example:

The Arizona Meteor Crater

Approximately 50,000 years ago, a large (radius of 25 m) iron-nickel meteorite collided with Earth at an estimated speed of $1.28 \times 10^4 \text{ m/s}$ in what is now the northern Arizona desert, in the United States. The impact produced a crater that is still visible today ([link](#)); it is approximately 1200 m (three-quarters of a mile) in diameter, 170 m deep, and has a rim that rises 45 m above the surrounding desert plain. Iron-nickel meteorites typically have a density of $\rho = 7970 \text{ kg/m}^3$. Use impulse considerations to estimate the average force and the maximum force that the meteor applied to Earth during the impact.



The Arizona Meteor Crater in Flagstaff, Arizona (often referred to as the Barringer Crater after the person who first suggested its origin and whose family owns the land). (credit: modification of work by “Shane.torgerson”/Wikimedia Commons)

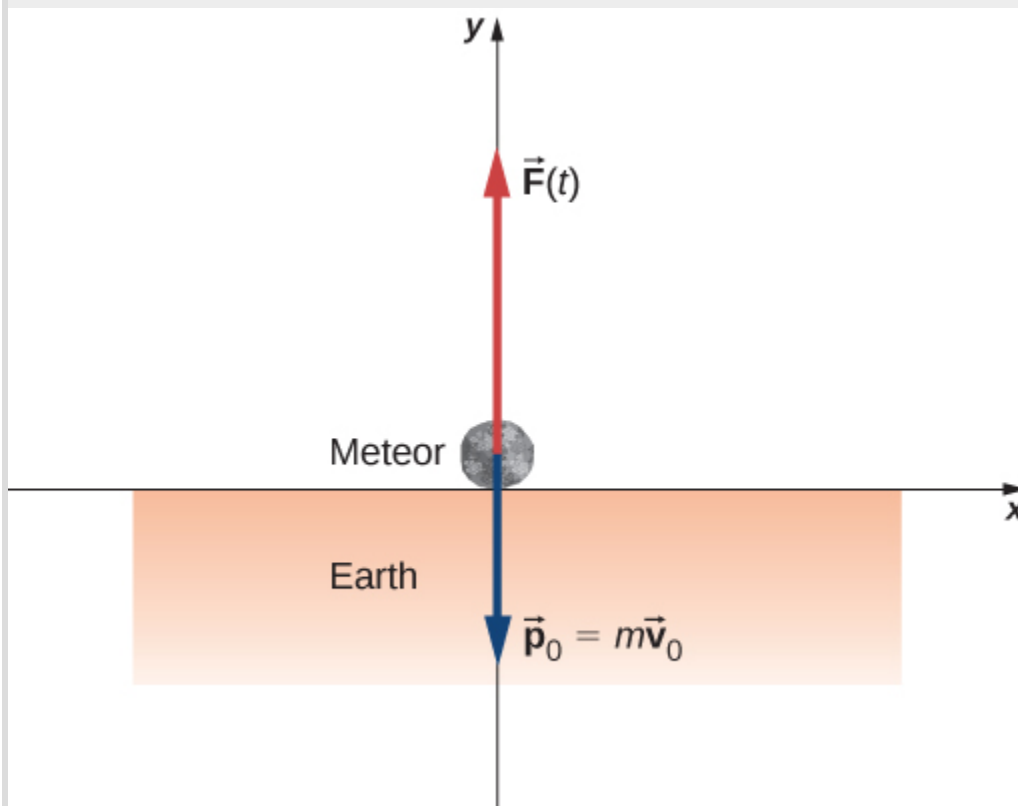
Strategy

It is conceptually easier to reverse the question and calculate the force that Earth applied on the meteor in order to stop it. Therefore, we'll calculate the force on the meteor and then use Newton's third law to argue that the force from the meteor on Earth was equal in magnitude and opposite in direction.

Using the given data about the meteor, and making reasonable guesses about the shape of the meteor and impact time, we first calculate the impulse using [\[link\]](#). We then use the relationship between force and impulse [\[link\]](#) to estimate the average force during impact. Next, we choose a reasonable force function for the impact event, calculate the average value of that function [\[link\]](#), and set the resulting expression equal to the calculated average force. This enables us to solve for the maximum force.

Solution

Define upward to be the $+y$ -direction. For simplicity, assume the meteor is traveling vertically downward prior to impact. In that case, its initial velocity is $\vec{v}_i = -v_i\hat{j}$, and the force Earth exerts on the meteor points upward, $\vec{F}(t) = +F(t)\hat{j}$. The situation at $t = 0$ is depicted below.



The average force during the impact is related to the impulse by
Equation:

$$\vec{\mathbf{F}}_{\text{ave}} = \frac{\vec{\mathbf{J}}}{\Delta t}.$$

From [\[link\]](#), $\vec{\mathbf{J}} = m\Delta\vec{\mathbf{v}}$, so we have

Equation:

$$\vec{\mathbf{F}}_{\text{ave}} = \frac{m\Delta\vec{\mathbf{v}}}{\Delta t}.$$

The mass is equal to the product of the meteor's density and its volume:

Equation:

$$m = \rho V.$$

If we assume (guess) that the meteor was roughly spherical, we have

Equation:

$$V = \frac{4}{3}\pi R^3.$$

Thus we obtain

Equation:

$$\vec{\mathbf{F}}_{\text{ave}} = \frac{\rho V \Delta\vec{\mathbf{v}}}{\Delta t} = \frac{\rho \left(\frac{4}{3}\pi R^3 \right) (\vec{\mathbf{v}}_{\text{f}} - \vec{\mathbf{v}}_{\text{i}})}{\Delta t}.$$

The problem says the velocity at impact was $-1.28 \times 10^4 \text{ m/s}\hat{\mathbf{j}}$ (the final velocity is zero); also, we guess that the primary impact lasted about $t_{\text{max}} = 2 \text{ s}$. Substituting these values gives

Equation:

$$\begin{aligned}\vec{\mathbf{F}}_{\text{ave}} &= \frac{\left(7970 \frac{\text{kg}}{\text{m}^3} \right) \left[\frac{4}{3}\pi (25 \text{ m})^3 \right] \left[0 \frac{\text{m}}{\text{s}} - \left(-1.28 \times 10^4 \frac{\text{m}}{\text{s}}\hat{\mathbf{j}} \right) \right]}{2 \text{ s}} \\ &= + (3.33 \times 10^{12} \text{ N})\hat{\mathbf{j}}\end{aligned}$$

This is the average force applied during the collision. Notice that this force vector points in the same direction as the change of velocity vector $\Delta \vec{v}$. Next, we calculate the maximum force. The impulse is related to the force function by

Equation:

$$\vec{J} = \int_{t_i}^{t_{\max}} \vec{F}(t) dt.$$

We need to make a reasonable choice for the force as a function of time. We define $t = 0$ to be the moment the meteor first touches the ground. Then we assume the force is a maximum at impact, and rapidly drops to zero. A function that does this is

Equation:

$$F(t) = F_{\max} e^{-t^2/(2\tau^2)}.$$

(The parameter τ represents how rapidly the force decreases to zero.) The average force is

Equation:

$$F_{\text{ave}} = \frac{1}{\Delta t} \int_0^{t_{\max}} F_{\max} e^{-t^2/(2\tau^2)} dt$$

where $\Delta t = t_{\max} - 0$ s. Since we already have a numeric value for F_{ave} , we can use the result of the integral to obtain F_{\max} .

Choosing $\tau = \frac{1}{e} t_{\max}$ (this is a common choice, as you will see in later chapters), and guessing that $t_{\max} = 2$ s, this integral evaluates to

Equation:

$$F_{\text{avg}} = 0.458 F_{\max}.$$

Thus, the maximum force has a magnitude of

Equation:

$$\begin{aligned} 0.458 F_{\max} &= 3.33 \times 10^{12} \text{ N} \\ F_{\max} &= 7.27 \times 10^{12} \text{ N} \end{aligned}$$

The complete force function, including the direction, is

Equation:

$$\vec{\mathbf{F}}(t) = (7.27 \times 10^{12} \text{ N})e^{-t^2/(8\text{s}^2)}\hat{\mathbf{j}}.$$

This is the force Earth applied to the meteor; by Newton's third law, the force the meteor applied to Earth is

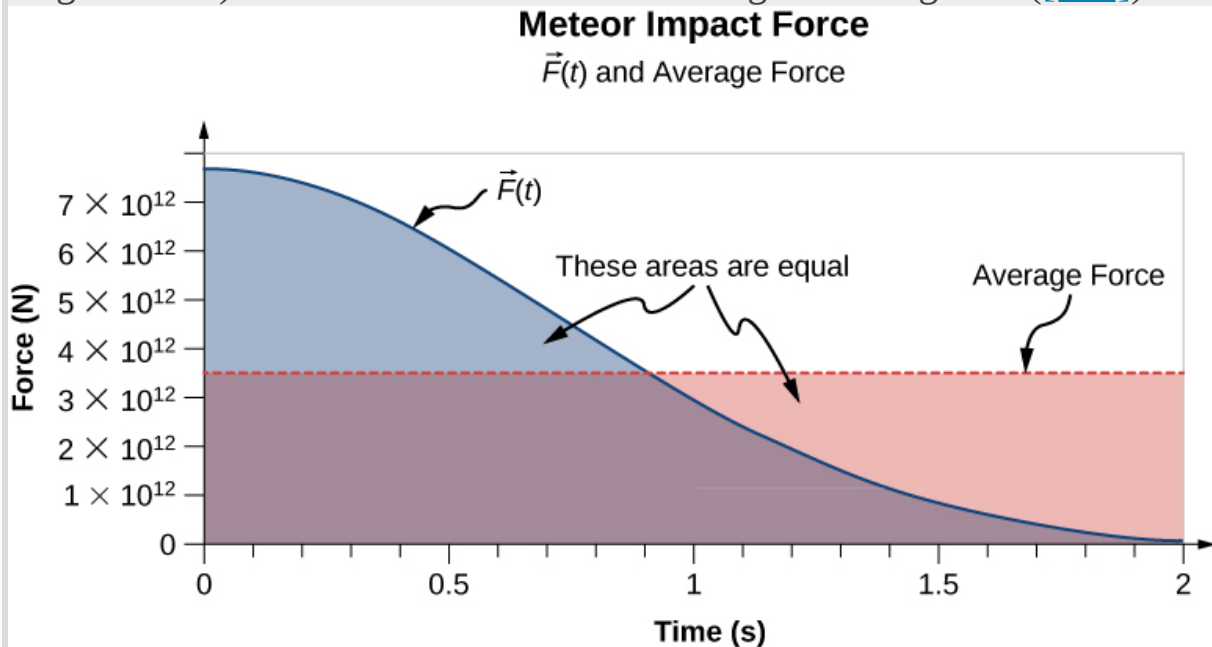
Equation:

$$\vec{\mathbf{F}}(t) = - (7.27 \times 10^{12} \text{ N})e^{-t^2/(8\text{s}^2)}\hat{\mathbf{j}}$$

which is the answer to the original question.

Significance

The graph of this function contains important information. Let's graph (the magnitude of) both this function and the average force together ([link](#)).



A graph of the average force (in red) and the force as a function of time (blue) of the meteor impact. The areas under the curves are equal to each other, and are numerically equal to the applied impulse.

Notice that the area under each plot has been filled in. For the plot of the (constant) force F_{ave} , the area is a rectangle, corresponding to $F_{\text{ave}}\Delta t = J$. As for the plot of $F(t)$, recall from calculus that the area under the plot of a function is numerically equal to the integral of that function, over the specified interval; so here, that is $\int_0^{t_{\text{max}}} F(t)dt = J$. Thus, the areas are equal, and both represent the impulse that the meteor applied to Earth during the two-second impact. The average force on Earth sounds like a huge force, and it is. Nevertheless, Earth barely noticed it. The acceleration Earth obtained was just

Equation:

$$\vec{a} = \frac{-\vec{F}_{\text{ave}}}{M_{\text{Earth}}} = \frac{-(3.33 \times 10^{12} \text{ N})\hat{\mathbf{j}}}{5.97 \times 10^{24} \text{ kg}} = -\left(5.6 \times 10^{-13} \frac{\text{m}}{\text{s}^2}\right)\hat{\mathbf{j}}$$

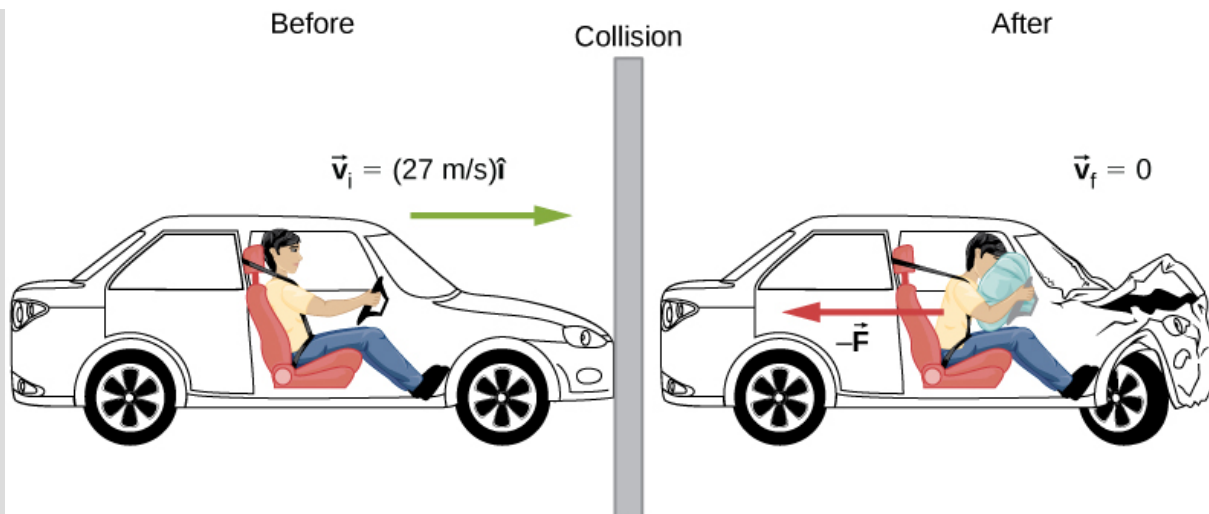
which is completely immeasurable. That said, the impact created seismic waves that nowadays could be detected by modern monitoring equipment.

Example:

The Benefits of Impulse

A car traveling at 27 m/s collides with a building. The collision with the building causes the car to come to a stop in approximately 1 second. The driver, who weighs 860 N, is protected by a combination of a variable-tension seatbelt and an airbag ([link](#)). (In effect, the driver collides with the seatbelt and airbag and *not* with the building.) The airbag and seatbelt slow his velocity, such that he comes to a stop in approximately 2.5 s.

- What average force does the driver experience during the collision?
- Without the seatbelt and airbag, his collision time (with the steering wheel) would have been approximately 0.20 s. What force would he experience in this case?



The motion of a car and its driver at the instant before and the instant after colliding with the wall. The restrained driver experiences a large backward force from the seatbelt and airbag, which causes his velocity to decrease to zero. (The forward force from the seatback is much smaller than the backward force, so we neglect it in the solution.)

Strategy

We are given the driver's weight, his initial and final velocities, and the time of collision; we are asked to calculate a force. Impulse seems the right way to tackle this; we can combine [\[link\]](#) and [\[link\]](#).

Solution

- Define the $+x$ -direction to be the direction the car is initially moving.

We know

Equation:

$$\vec{J} = \vec{F}\Delta t$$

and

Equation:

$$\vec{J} = m\Delta\vec{v}.$$

Since J is equal to both those things, they must be equal to each other:
Equation:

$$\vec{F}\Delta t = m\Delta\vec{v}.$$

We need to convert this weight to the equivalent mass, expressed in SI units:

Equation:

$$\frac{860 \text{ N}}{9.8 \text{ m/s}^2} = 87.8 \text{ kg}.$$

Remembering that $\Delta\vec{v} = \vec{v}_f - \vec{v}_i$, and noting that the final velocity is zero, we solve for the force:

Equation:

$$\vec{F} = m \frac{0 - v_i \hat{i}}{\Delta t} = (87.8 \text{ kg}) \left(\frac{-(27 \text{ m/s}) \hat{i}}{2.5 \text{ s}} \right) = -(948 \text{ N}) \hat{i}.$$

The negative sign implies that the force slows him down. For perspective, this is about 1.1 times his own weight.

b. Same calculation, just the different time interval:

Equation:

$$\vec{F} = (87.8 \text{ kg}) \left(\frac{-(27 \text{ m/s}) \hat{i}}{0.20 \text{ s}} \right) = -(11,853 \text{ N}) \hat{i}$$

which is about 14 times his own weight. Big difference!

Significance

You see that the value of an airbag is how greatly it reduces the force on the vehicle occupants. For this reason, they have been required on all passenger vehicles in the United States since 1991, and have been commonplace throughout Europe and Asia since the mid-1990s. The change of momentum in a crash is the same, with or without an airbag; the force, however, is vastly different.

Effect of Impulse

Since an impulse is a force acting for some amount of time, it causes an object's motion to change. Recall [\[link\]](#):

Equation:

$$\vec{\mathbf{J}} = m\Delta\vec{\mathbf{v}}.$$

Because $m\vec{\mathbf{v}}$ is the momentum of a system, $m\Delta\vec{\mathbf{v}}$ is the *change* of momentum $\Delta\vec{\mathbf{p}}$. This gives us the following relation, called the **impulse-momentum theorem** (or relation).

Note:

Impulse-Momentum Theorem

An impulse applied to a system changes the system's momentum, and that change of momentum is exactly equal to the impulse that was applied:

Equation:

$$\vec{\mathbf{J}} = \Delta\vec{\mathbf{p}}.$$

The impulse-momentum theorem is depicted graphically in [\[link\]](#).

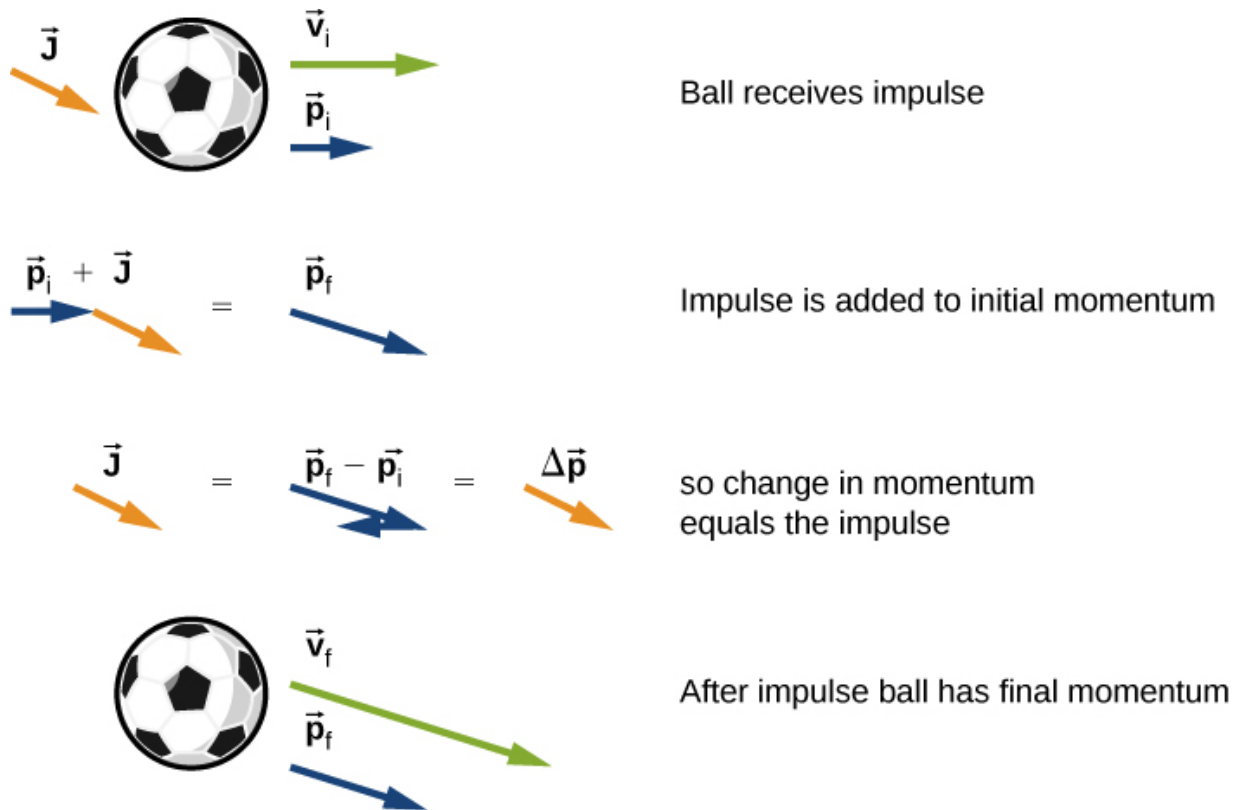


Illustration of impulse-momentum theorem. (a) A ball with initial velocity \vec{v}_0 and momentum \vec{p}_0 receives an impulse \vec{J} . (b) This impulse is added vectorially to the initial momentum. (c) Thus, the impulse equals the change in momentum, $\vec{J} = \Delta\vec{p}$. (d) After the impulse, the ball moves off with its new momentum \vec{p}_f .

There are two crucial concepts in the impulse-momentum theorem:

1. Impulse is a vector quantity; an impulse of, say, $-(10 \text{ N} \cdot \text{s})\hat{i}$ is very different from an impulse of $+(10 \text{ N} \cdot \text{s})\hat{i}$; they cause completely opposite changes of momentum.
2. An impulse does not cause momentum; rather, it causes a *change* in the momentum of an object. Thus, you must subtract the final momentum from the initial momentum, and—since momentum is also a vector quantity—you must take careful account of the signs of the momentum vectors.

The most common questions asked in relation to impulse are to calculate the applied force, or the change of velocity that occurs as a result of applying an impulse. The general approach is the same.

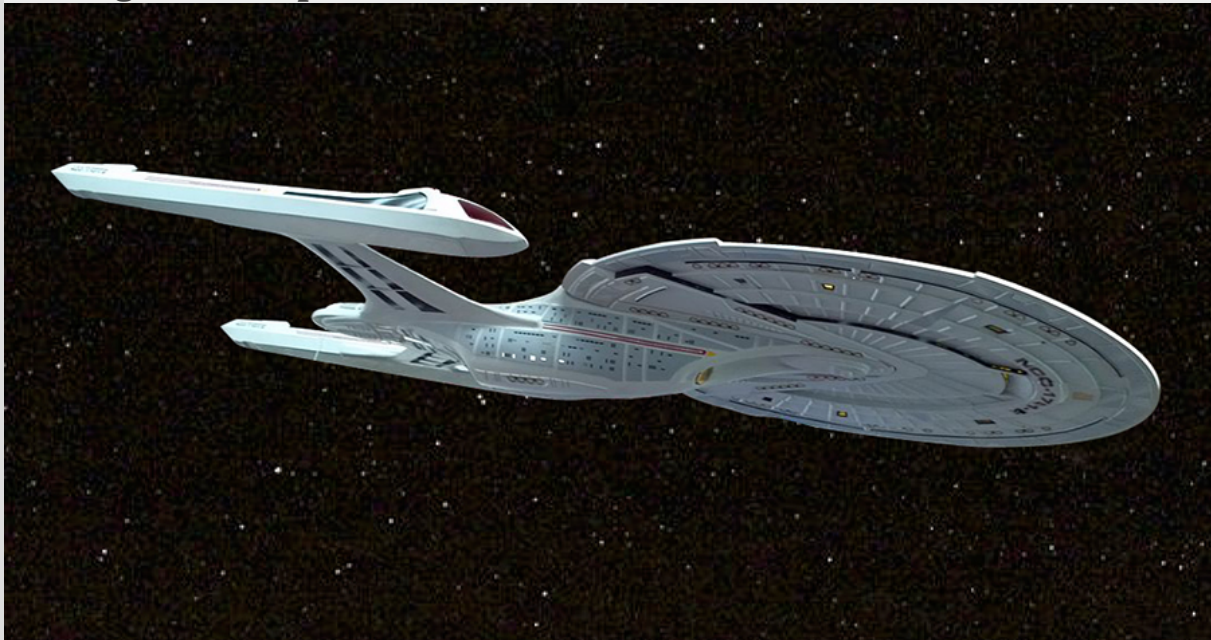
Note:

Impulse-Momentum Theorem

1. Express the impulse as force times the relevant time interval.
2. Express the impulse as the change of momentum, usually $m\Delta v$.
3. Equate these and solve for the desired quantity.

Example:

Moving the *Enterprise*



The fictional starship *Enterprise* from the Star Trek adventures operated on so-called “impulse engines” that combined matter with antimatter to produce energy.

When Captain Picard commands, “Take us out; ahead one-quarter impulse,” the starship *Enterprise* ([\[link\]](#)) starts from rest to a final speed of $v_f = 1/4 (3.0 \times 10^8 \text{ m/s})$. Assuming this maneuver is completed in 60 s, what average force did the impulse engines apply to the ship?

Strategy

We are asked for a force; we know the initial and final speeds (and hence the change in speed), and we know the time interval over which this all happened. In particular, we know the amount of time that the force acted. This suggests using the impulse-momentum relation. To use that, though, we need the mass of the *Enterprise*. An internet search gives a best estimate of the mass of the *Enterprise* (in the 2009 movie) as $2 \times 10^9 \text{ kg}$.

Solution

Because this problem involves only one direction (i.e., the direction of the force applied by the engines), we only need the scalar form of the impulse-momentum theorem [\[link\]](#), which is

Equation:

$$\Delta p = J$$

with

Equation:

$$\Delta p = m\Delta v$$

and

Equation:

$$J = F\Delta t.$$

Equating these expressions gives

Equation:

$$F\Delta t = m\Delta v.$$

Solving for the magnitude of the force and inserting the given values leads to

Equation:

$$F = \frac{m\Delta v}{\Delta t} = \frac{(2 \times 10^9 \text{ kg})(7.5 \times 10^7 \text{ m/s})}{60 \text{ s}} = 2.5 \times 10^{15} \text{ N}.$$

Significance

This is an unimaginably huge force. It goes almost without saying that such a force would kill everyone on board instantly, as well as destroying every piece of equipment. Fortunately, the *Enterprise* has “inertial dampeners.” It is left as an exercise for the reader’s imagination to determine how these work.

Note:

Exercise:

Problem:

Check Your Understanding The U.S. Air Force uses “10gs” (an acceleration equal to $10 \times 9.8 \text{ m/s}^2$) as the maximum acceleration a human can withstand (but only for several seconds) and survive. How much time must the *Enterprise* spend accelerating if the humans on board are to experience an average of at most 10gs of acceleration? (Assume the inertial dampeners are offline.)

Solution:

To reach a final speed of $v_f = \frac{1}{4}(3.0 \times 10^8 \text{ m/s})$ at an acceleration of $10g$, the time required is

$$10g = \frac{v_f}{\Delta t}$$

$$\Delta t = \frac{v_f}{10g} = \frac{\frac{1}{4}(3.0 \times 10^8 \text{ m/s})}{10g} = 7.7 \times 10^5 \text{ s} = 8.9 \text{ d}$$

Example:

The iPhone Drop

Apple released its iPhone 6 Plus in November 2014. According to many reports, it was originally supposed to have a screen made from sapphire, but that was changed at the last minute for a hardened glass screen.

Reportedly, this was because the sapphire screen cracked when the phone was dropped. What force did the iPhone 6 Plus experience as a result of being dropped?

Strategy

The force the phone experiences is due to the impulse applied to it by the floor when the phone collides with the floor. Our strategy then is to use the impulse-momentum relationship. We calculate the impulse, estimate the impact time, and use this to calculate the force.

We need to make a couple of reasonable estimates, as well as find technical data on the phone itself. First, let's suppose that the phone is most often dropped from about chest height on an average-height person. Second, assume that it is dropped from rest, that is, with an initial vertical velocity of zero. Finally, we assume that the phone bounces very little—the height of its bounce is assumed to be negligible.

Solution

Define upward to be the $+y$ -direction. A typical height is approximately $h = 1.5$ m and, as stated, $\vec{v}_i = (0 \text{ m/s})\hat{i}$. The average force on the phone is related to the impulse the floor applies on it during the collision:

Equation:

$$\vec{F}_{\text{ave}} = \frac{\vec{J}}{\Delta t}.$$

The impulse \vec{J} equals the change in momentum,

Equation:

$$\vec{J} = \Delta \vec{p}$$

so

Equation:

$$\vec{F}_{\text{ave}} = \frac{\Delta \vec{p}}{\Delta t}.$$

Next, the change of momentum is

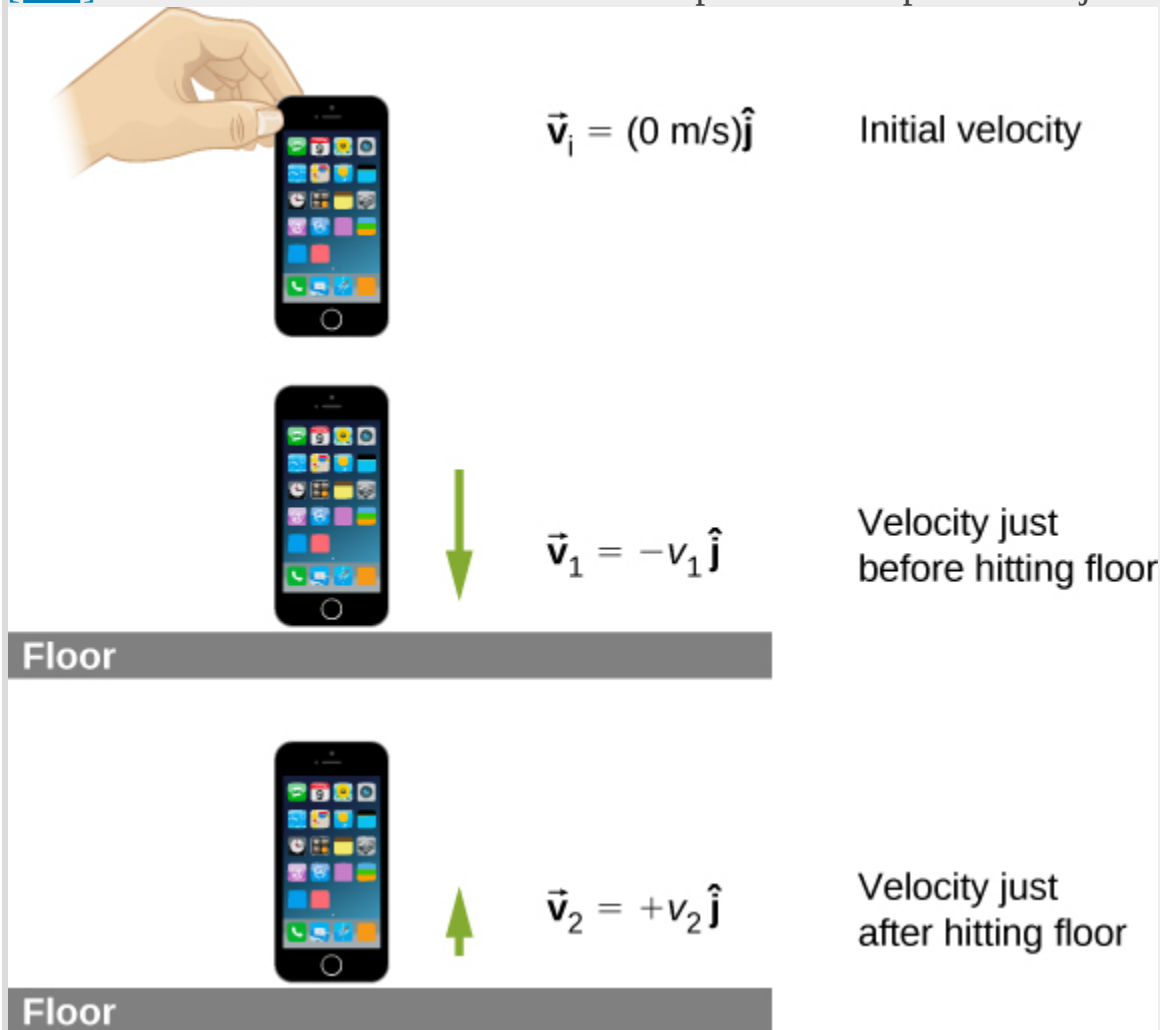
Equation:

$$\Delta \vec{p} = m \Delta \vec{v}.$$

We need to be careful with the velocities here; this is the change of velocity due to the collision with the floor. But the phone also has an initial drop velocity [$\vec{v}_i = (0 \text{ m/s})\hat{j}$], so we label our velocities. Let:

- \vec{v}_i = the initial velocity with which the phone was dropped (zero, in this example)
- \vec{v}_1 = the velocity the phone had the instant just before it hit the floor
- \vec{v}_2 = the final velocity of the phone as a result of hitting the floor

[\[link\]](#) shows the velocities at each of these points in the phone's trajectory.



(a) The initial velocity of the phone is zero, just after the person drops it. (b) Just before the phone hits the floor, its velocity is \vec{v}_1 , which is unknown at the moment, except for its direction, which is downward ($-\hat{j}$). (c) After bouncing off the floor, the phone has a velocity \vec{v}_2 , which is also unknown, except for its direction, which is upward ($+\hat{j}$).

With these definitions, the change of momentum of the phone during the collision with the floor is

Equation:

$$m\Delta\vec{v} = m(\vec{v}_2 - \vec{v}_1).$$

Since we assume the phone doesn't bounce at all when it hits the floor (or at least, the bounce height is negligible), then \vec{v}_2 is zero, so

Equation:

$$\begin{aligned} m\Delta\vec{v} &= m\left[0 - (-v_1\hat{j})\right] \\ m\Delta\vec{v} &= +mv_1\hat{j}. \end{aligned}$$

We can get the speed of the phone just before it hits the floor using either kinematics or conservation of energy. We'll use conservation of energy here; you should re-do this part of the problem using kinematics and prove that you get the same answer.

First, define the zero of potential energy to be located at the floor.

Conservation of energy then gives us:

Equation:

$$\begin{aligned} E_i &= E_f \\ K_i + U_i &= K_f + U_f \\ \frac{1}{2}mv_i^2 + mgh_{\text{drop}} &= \frac{1}{2}mv_f^2 + mgh_{\text{floor}}. \end{aligned}$$

Defining $h_{\text{floor}} = 0$ and using $\vec{v}_i = (0 \text{ m/s})\hat{j}$ gives

Equation:

$$\begin{aligned}\frac{1}{2}mv_1^2 &= mgh_{\text{drop}} \\ v_1 &= \pm\sqrt{2gh_{\text{drop}}}.\end{aligned}$$

Because v_1 is a vector magnitude, it must be positive. Thus, $m\Delta v = mv_1 = m\sqrt{2gh_{\text{drop}}}$. Inserting this result into the expression for force gives

Equation:

$$\begin{aligned}\vec{\mathbf{F}} &= \frac{\Delta\vec{\mathbf{p}}}{\Delta t} \\ &= \frac{m\Delta\vec{\mathbf{v}}}{\Delta t} \\ &= \frac{+mv_1\hat{\mathbf{j}}}{\Delta t} \\ &= \frac{m\sqrt{2gh}}{\Delta t}\hat{\mathbf{j}}.\end{aligned}$$

Finally, we need to estimate the collision time. One common way to estimate a collision time is to calculate how long the object would take to travel its own length. The phone is moving at 5.4 m/s just before it hits the floor, and it is 0.14 m long, giving an estimated collision time of 0.026 s. Inserting the given numbers, we obtain

Equation:

$$\vec{\mathbf{F}} = \frac{(0.172 \text{ kg})\sqrt{2(9.8 \text{ m/s}^2)(1.5 \text{ m})}}{0.026 \text{ s}}\hat{\mathbf{j}} = (36 \text{ N})\hat{\mathbf{j}}.$$

Significance

The iPhone itself weighs just $(0.172 \text{ kg})(9.81 \text{ m/s}^2) = 1.68 \text{ N}$; the force the floor applies to it is therefore over 20 times its weight.

Note:**Exercise:**

Problem:

Check Your Understanding What if we had assumed the phone *did* bounce on impact? Would this have increased the force on the iPhone, decreased it, or made no difference?

Solution:

If the phone bounces up with approximately the same initial speed as its impact speed, the change in momentum of the phone will be $\Delta\vec{p} = m\Delta\vec{v} - (-m\Delta\vec{v}) = 2m\Delta\vec{v}$. This is twice the momentum change than when the phone does not bounce, so the impulse-momentum theorem tells us that more force must be applied to the phone.

Momentum and Force

In [\[link\]](#), we obtained an important relationship:

Note:**Equation:**

$$\vec{\mathbf{F}}_{\text{ave}} = \frac{\Delta\vec{p}}{\Delta t}.$$

In words, the average force applied to an object is equal to the change of the momentum that the force causes, divided by the time interval over which this change of momentum occurs. This relationship is very useful in situations where the collision time Δt is small, but measureable; typical values would be 1/10th of a second, or even one thousandth of a second.

Car crashes, punting a football, or collisions of subatomic particles would meet this criterion.

For a *continuously* changing momentum—due to a continuously changing force—this becomes a powerful conceptual tool. In the limit $\Delta t \rightarrow dt$, [\[link\]](#) becomes

Note:

Equation:

$$\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}}{dt}.$$

This says that the rate of change of the system’s momentum (implying that momentum is a function of time) is exactly equal to the net applied force (also, in general, a function of time). This is, in fact, Newton’s second law, written in terms of momentum rather than acceleration. This is the relationship Newton himself presented in his *Principia Mathematica* (although he called it “quantity of motion” rather than “momentum”).

If the mass of the system remains constant, [\[link\]](#) reduces to the more familiar form of Newton’s second law. We can see this by substituting the definition of momentum:

Equation:

$$\vec{\mathbf{F}} = \frac{d(m\vec{\mathbf{v}})}{dt} = m \frac{d\vec{\mathbf{v}}}{dt} = m\vec{\mathbf{a}}.$$

The assumption of constant mass allowed us to pull m out of the derivative. If the mass is not constant, we cannot use this form of the second law, but instead must start from [\[link\]](#). Thus, one advantage to expressing force in terms of changing momentum is that it allows for the mass of the system to

change, as well as the velocity; this is a concept we'll explore when we study the motion of rockets.

Note:

Newton's Second Law of Motion in Terms of Momentum

The net external force on a system is equal to the rate of change of the momentum of that system caused by the force:

Equation:

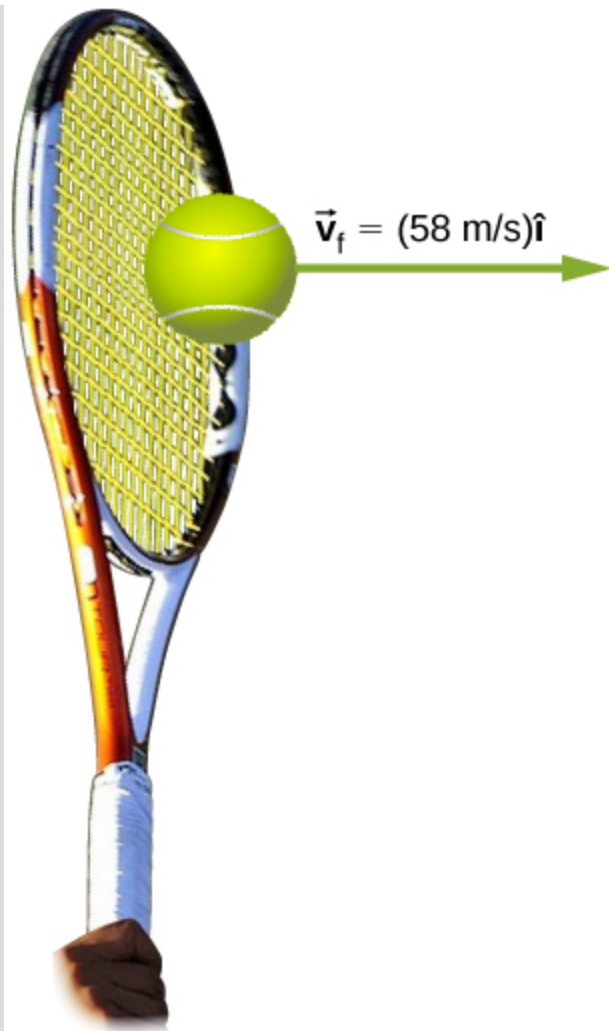
$$\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}}{dt}.$$

Although [\[link\]](#) allows for changing mass, as we will see in [Rocket Propulsion](#), the relationship between momentum and force remains useful when the mass of the system is constant, as in the following example.

Example:

Calculating Force: Venus Williams' Tennis Serve

During the 2007 French Open, Venus Williams hit the fastest recorded serve in a premier women's match, reaching a speed of 58 m/s (209 km/h). What is the average force exerted on the 0.057-kg tennis ball by Venus Williams' racquet? Assume that the ball's speed just after impact is 58 m/s, as shown in [\[link\]](#), that the initial horizontal component of the velocity before impact is negligible, and that the ball remained in contact with the racquet for 5.0 ms.



The final velocity of the tennis ball is $\vec{v}_f = (58 \text{ m/s})\hat{i}$.

Strategy

This problem involves only one dimension because the ball starts from having no horizontal velocity component before impact. Newton's second law stated in terms of momentum is then written as

Equation:

$$\vec{F} = \frac{d\vec{p}}{dt}.$$

As noted above, when mass is constant, the change in momentum is given by

Equation:

$$\Delta p = m\Delta v = m(v_f - v_i)$$

where we have used scalars because this problem involves only one dimension. In this example, the velocity just after impact and the time interval are given; thus, once Δp is calculated, we can use $F = \frac{\Delta p}{\Delta t}$ to find the force.

Solution

To determine the change in momentum, insert the values for the initial and final velocities into the equation above:

Equation:

$$\begin{aligned}\Delta p &= m(v_f - v_i) \\ &= (0.057 \text{ kg})(58 \text{ m/s} - 0 \text{ m/s}) \\ &= 3.3 \frac{\text{kg}\cdot\text{m}}{\text{s}}.\end{aligned}$$

Now the magnitude of the net external force can be determined by using

Equation:

$$F = \frac{\Delta p}{\Delta t} = \frac{3.3 \frac{\text{kg}\cdot\text{m}}{\text{s}}}{5.0 \times 10^{-3} \text{ s}} = 6.6 \times 10^2 \text{ N}.$$

where we have retained only two significant figures in the final step.

Significance

This quantity was the average force exerted by Venus Williams' racquet on the tennis ball during its brief impact (note that the ball also experienced the 0.57-N force of gravity, but that force was not due to the racquet). This problem could also be solved by first finding the acceleration and then using $F = ma$, but one additional step would be required compared with the strategy used in this example.

Summary

- When a force is applied on an object for some amount of time, the object experiences an impulse.
- This impulse is equal to the object's change of momentum.
- Newton's second law in terms of momentum states that the net force applied to a system equals the rate of change of the momentum that the force causes.

Conceptual Questions

Exercise:

Problem:

Is it possible for a small force to produce a larger impulse on a given object than a large force? Explain.

Solution:

Yes; impulse is the force applied multiplied by the time during which it is applied ($J = F\Delta t$), so if a small force acts for a long time, it may result in a larger impulse than a large force acting for a small time.

Exercise:

Problem:

Why is a 10-m fall onto concrete far more dangerous than a 10-m fall onto water?

Exercise:

Problem:

What external force is responsible for changing the momentum of a car moving along a horizontal road?

Solution:

By friction, the road exerts a horizontal force on the tires of the car, which changes the momentum of the car.

Exercise:

Problem:

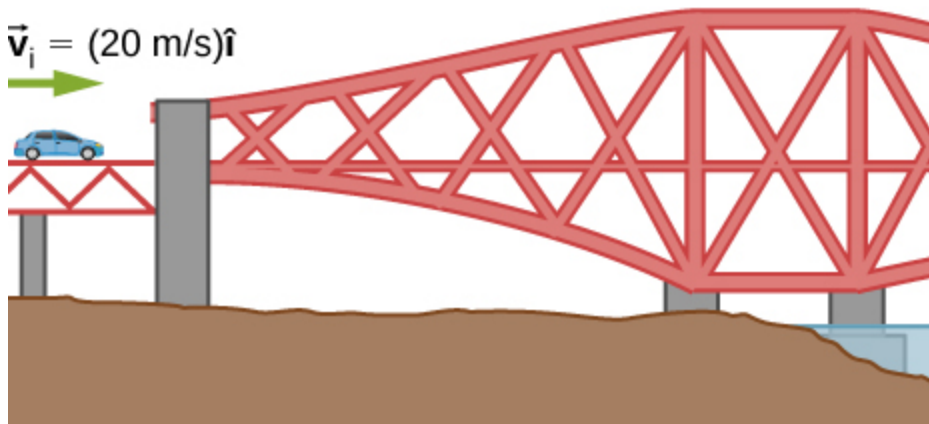
A piece of putty and a tennis ball with the same mass are thrown against a wall with the same velocity. Which object experiences a greater force from the wall or are the forces equal? Explain.

Problems

Exercise:

Problem:

A 75.0-kg person is riding in a car moving at 20.0 m/s when the car runs into a bridge abutment (see the following figure).



- Calculate the average force on the person if he is stopped by a padded dashboard that compresses an average of 1.00 cm.
- Calculate the average force on the person if he is stopped by an air bag that compresses an average of 15.0 cm.

Solution:

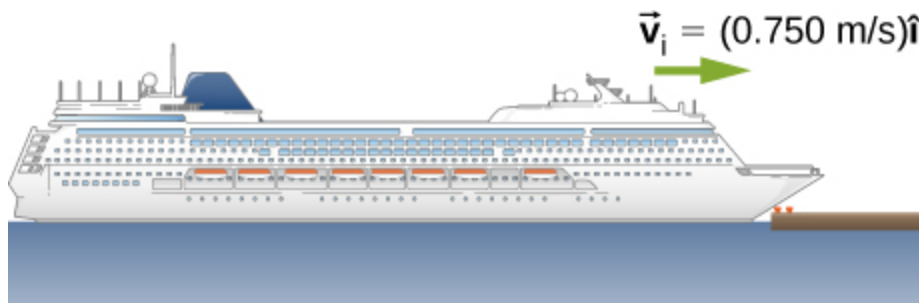
- a. $1.50 \times 10^6 \text{ N}$; b. $1.00 \times 10^5 \text{ N}$

Exercise:**Problem:**

One hazard of space travel is debris left by previous missions. There are several thousand objects orbiting Earth that are large enough to be detected by radar, but there are far greater numbers of very small objects, such as flakes of paint. Calculate the force exerted by a 0.100-mg chip of paint that strikes a spacecraft window at a relative speed of $4.00 \times 10^3 \text{ m/s}$, given the collision lasts $6.00 \times 10^{-8} \text{ s}$.

Exercise:**Problem:**

A cruise ship with a mass of $1.00 \times 10^7 \text{ kg}$ strikes a pier at a speed of 0.750 m/s . It comes to rest after traveling 6.00 m , damaging the ship, the pier, and the tugboat captain's finances. Calculate the average force exerted on the pier using the concept of impulse. (*Hint:* First calculate the time it took to bring the ship to rest, assuming a constant force.)



Solution:

$$4.69 \times 10^5 \text{ N}$$

Exercise:**Problem:**

Calculate the final speed of a 110-kg rugby player who is initially running at 8.00 m/s but collides head-on with a padded goalpost and experiences a backward force of $1.76 \times 10^4 \text{ N}$ for $5.50 \times 10^{-2} \text{ s}$.

Exercise:**Problem:**

Water from a fire hose is directed horizontally against a wall at a rate of 50.0 kg/s and a speed of 42.0 m/s. Calculate the force exerted on the wall, assuming the water's horizontal momentum is reduced to zero.

Solution:

$$2.10 \times 10^3 \text{ N}$$

Exercise:**Problem:**

A 0.450-kg hammer is moving horizontally at 7.00 m/s when it strikes a nail and comes to rest after driving the nail 1.00 cm into a board. Assume constant acceleration of the hammer-nail pair.

- Calculate the duration of the impact.
- What was the average force exerted on the nail?

Exercise:**Problem:**

What is the momentum (as a function of time) of a 5.0-kg particle moving with a velocity $\vec{v}(t) = (2.0\hat{i} + 4.0t\hat{j}) \text{ m/s}$? What is the net force acting on this particle?

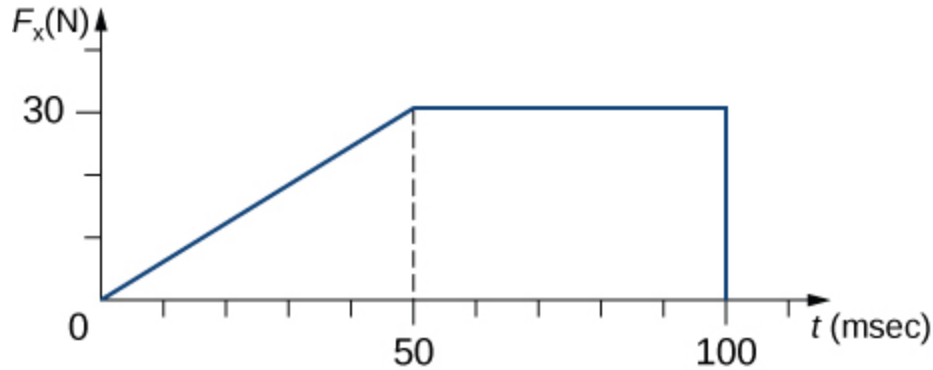
Solution:

$$\vec{p}(t) = (10\hat{i} + 20t\hat{j}) \text{ kg} \cdot \text{m/s}; \vec{F} = (20 \text{ N})\hat{j}$$

Exercise:

Problem:

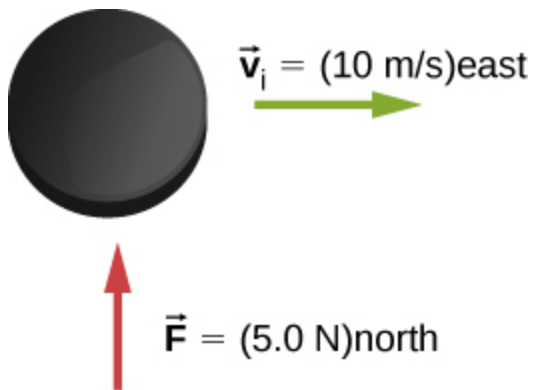
The x -component of a force on a 46-g golf ball by a 7-iron versus time is plotted in the following figure:



- a. Find the x -component of the impulse during the intervals
 - i. $[0, 50 \text{ ms}]$, and
 - ii. $[50 \text{ ms}, 100 \text{ ms}]$
- b. Find the change in the x -component of the momentum during the intervals
 - iii. $[0, 50 \text{ ms}]$, and
 - iv. $[50 \text{ ms}, 100 \text{ ms}]$

Exercise:**Problem:**

A hockey puck of mass 150 g is sliding due east on a frictionless table with a speed of 10 m/s. Suddenly, a constant force of magnitude 5 N and direction due north is applied to the puck for 1.5 s. Find the north and east components of the momentum at the end of the 1.5-s interval.

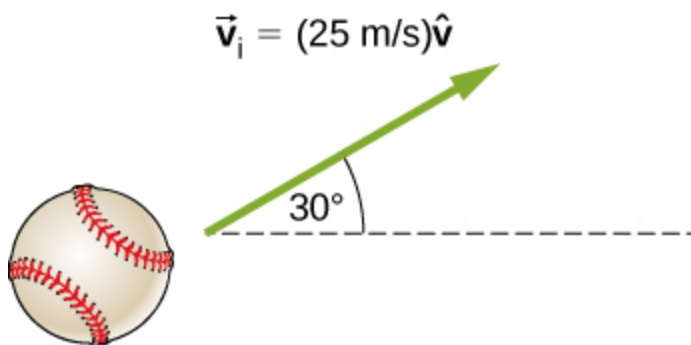


Solution:

Let the positive x -axis be in the direction of the original momentum.
Then $p_x = 1.5 \text{ kg} \cdot \text{m/s}$ and $p_y = 7.5 \text{ kg} \cdot \text{m/s}$

Exercise:**Problem:**

A ball of mass 250 g is thrown with an initial velocity of 25 m/s at an angle of 30° with the horizontal direction. Ignore air resistance. What is the momentum of the ball after 0.2 s? (Do this problem by finding the components of the momentum first, and then constructing the magnitude and direction of the momentum vector from the components.)

**Glossary**

impulse

effect of applying a force on a system for a time interval; this time interval is usually small, but does not have to be

impulse-momentum theorem

change of momentum of a system is equal to the impulse applied to the system

Conservation of Linear Momentum

By the end of this section, you will be able to:

- Explain the meaning of “conservation of momentum”
- Correctly identify if a system is, or is not, closed
- Define a system whose momentum is conserved
- Mathematically express conservation of momentum for a given system
- Calculate an unknown quantity using conservation of momentum

Recall Newton’s third law: When two objects of masses m_1 and m_2 interact (meaning that they apply forces on each other), the force that object 2 applies to object 1 is equal in magnitude and opposite in direction to the force that object 1 applies on object 2. Let:

- $\vec{\mathbf{F}}_{21}$ = the force on m_1 from m_2
- $\vec{\mathbf{F}}_{12}$ = the force on m_2 from m_1

Then, in symbols, Newton’s third law says

Equation:

$$\begin{aligned}\vec{\mathbf{F}}_{21} &= -\vec{\mathbf{F}}_{12} \\ m_1 \vec{\mathbf{a}}_1 &= -m_2 \vec{\mathbf{a}}_2.\end{aligned}$$

(Recall that these two forces do not cancel because they are applied to different objects. F_{21} causes m_1 to accelerate, and F_{12} causes m_2 to accelerate.)

Although the magnitudes of the forces on the objects are the same, the accelerations are not, simply because the masses (in general) are different. Therefore, the changes in velocity of each object are different:

Equation:

$$\frac{d\vec{\mathbf{v}}_1}{dt} \neq \frac{d\vec{\mathbf{v}}_2}{dt}.$$

However, the products of the mass and the change of velocity *are* equal (in magnitude):

Note:

Equation:

$$m_1 \frac{d\vec{v}_1}{dt} = -m_2 \frac{d\vec{v}_2}{dt}.$$

It's a good idea, at this point, to make sure you're clear on the physical meaning of the derivatives in [\[link\]](#). Because of the interaction, each object ends up getting its velocity changed, by an amount $d\vec{v}$. Furthermore, the interaction occurs over a time interval dt , which means that the change of velocities also occurs over dt . This time interval is the same for each object.

Let's assume, for the moment, that the masses of the objects do not change during the interaction. (We'll relax this restriction later.) In that case, we can pull the masses inside the derivatives:

Equation:

$$\frac{d}{dt}(m_1 \vec{v}_1) = -\frac{d}{dt}(m_2 \vec{v}_2)$$

and thus

Note:

Equation:

$$\frac{d\vec{p}_1}{dt} = -\frac{d\vec{p}_2}{dt}.$$

This says that *the rate at which momentum changes is the same for both objects*. The masses are different, and the changes of velocity are different, but the rate of change of the product of m and \vec{v} are the same.

Physically, this means that during the interaction of the two objects (m_1 and m_2), both objects have their momentum changed; but those changes are identical in magnitude, though opposite in sign. For example, the momentum of object 1 might increase, which means that the momentum of object 2 decreases by exactly the same amount.

In light of this, let's re-write [\[link\]](#) in a more suggestive form:

Note:

Equation:

$$\frac{d\vec{p}_1}{dt} + \frac{d\vec{p}_2}{dt} = 0.$$

This says that during the interaction, although object 1's momentum changes, and object 2's momentum also changes, these two changes cancel each other out, so that the total change of momentum of the two objects together is zero.

Since the total combined momentum of the two objects together never changes, then we could write

Equation:

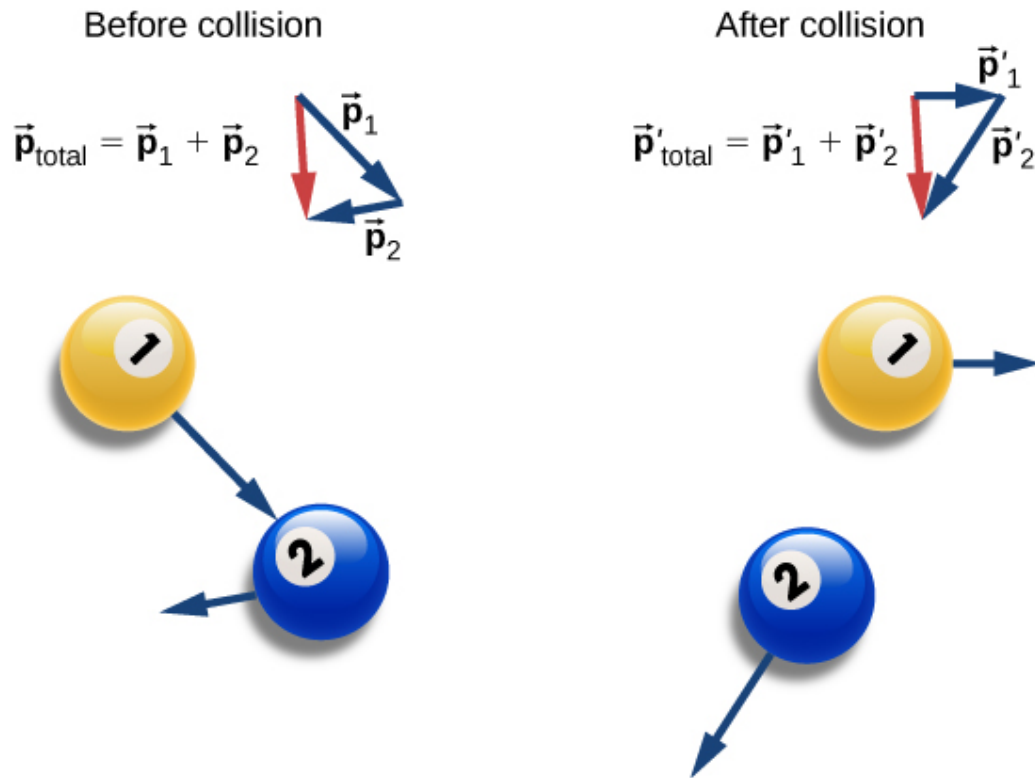
$$\frac{d}{dt}(\vec{p}_1 + \vec{p}_2) = 0$$

from which it follows that

Equation:

$$\vec{p}_1 + \vec{p}_2 = \text{constant}.$$

As shown in [\[link\]](#), the total momentum of the system before and after the collision remains the same.



Before the collision, the two billiard balls travel with momenta \vec{p}_1 and \vec{p}_2 . The total momentum of the system is the sum of these, as shown by the red vector labeled \vec{p}_{total} on the left. After the collision, the two billiard balls travel with different momenta \vec{p}'_1 and \vec{p}'_2 . The total momentum, however, has not changed, as shown by the red vector arrow \vec{p}'_{total} on the right.

Generalizing this result to N objects, we obtain

Note:

Equation:

$$\vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \cdots + \vec{p}_N = \text{constant}$$

$$\sum_{j=1}^N \vec{p}_j = \text{constant.}$$

[\[link\]](#) is the definition of the total (or net) momentum of a system of N interacting objects, along with the statement that the total momentum of a system of objects is constant in time—or better, is conserved.

Note:

Conservation Laws

If the value of a physical quantity is constant in time, we say that the quantity is conserved.

Requirements for Momentum Conservation

There is a complication, however. A system must meet two requirements for its momentum to be conserved:

1. *The mass of the system must remain constant during the interaction.*
As the objects interact (apply forces on each other), they may *transfer* mass from one to another; but any mass one object gains is balanced by the loss of that mass from another. The total mass of the system of objects, therefore, remains unchanged as time passes:

Equation:

$$\left[\frac{dm}{dt} \right]_{\text{system}} = 0.$$

2. *The net external force on the system must be zero.*
As the objects collide, or explode, and move around, they exert forces on each other. However, all of these forces are internal to the system, and thus each of these internal forces is balanced by another internal force that is equal in magnitude and opposite in sign. As a result, the change in momentum caused by each internal force is cancelled by another momentum change that is equal in magnitude and opposite in direction. Therefore, internal forces cannot change the total momentum of a system because the changes sum to zero. However, if there is some external force

that acts on all of the objects (gravity, for example, or friction), then this force changes the momentum of the system as a whole; that is to say, the momentum of the system is changed by the external force. Thus, for the momentum of the system to be conserved, we must have

Equation:

$$\vec{\mathbf{F}}_{\text{ext}} = \vec{\mathbf{0}}.$$

A system of objects that meets these two requirements is said to be a **closed system** (also called an isolated system). Thus, the more compact way to express this is shown below.

Note:

Law of Conservation of Momentum

The total momentum of a closed system is conserved:

Equation:

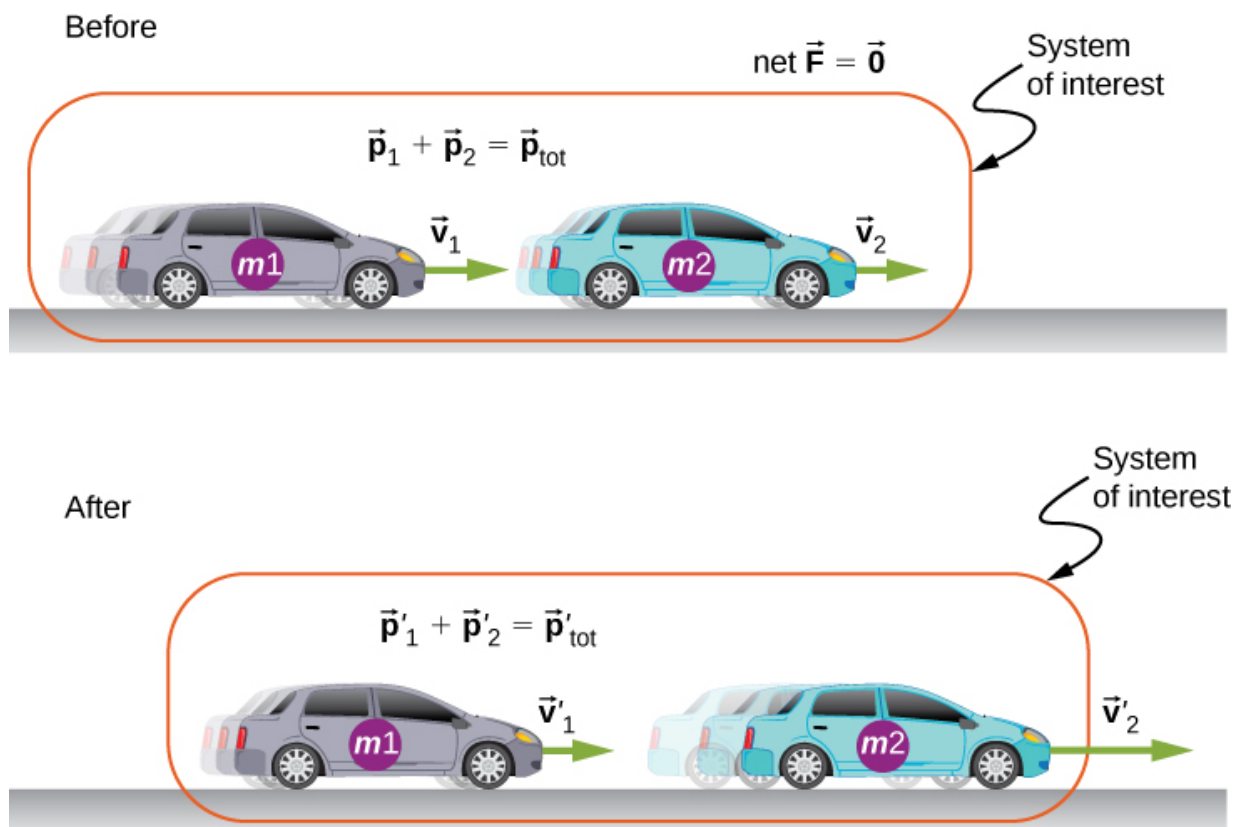
$$\sum_{j=1}^N \vec{\mathbf{p}}_j = \text{constant}.$$

This statement is called the **Law of Conservation of Momentum**. Along with the conservation of energy, it is one of the foundations upon which all of physics stands. All our experimental evidence supports this statement: from the motions of galactic clusters to the quarks that make up the proton and the neutron, and at every scale in between. *In a closed system, the total momentum never changes.*

Note that there absolutely *can* be external forces acting on the system; but for the system's momentum to remain constant, these external forces have to cancel, so that the *net* external force is zero. Billiard balls on a table all have a weight force acting on them, but the weights are balanced (canceled) by the normal forces, so there is no *net* force.

The Meaning of ‘System’

A **system** (mechanical) is the collection of objects in whose motion (kinematics and dynamics) you are interested. If you are analyzing the bounce of a ball on the ground, you are probably only interested in the motion of the ball, and not of Earth; thus, the ball is your system. If you are analyzing a car crash, the two cars together compose your system ([link](#)).



The two cars together form the system that is to be analyzed. It is important to remember that the contents (the mass) of the system do not change before, during, or after the objects in the system interact.

Note:

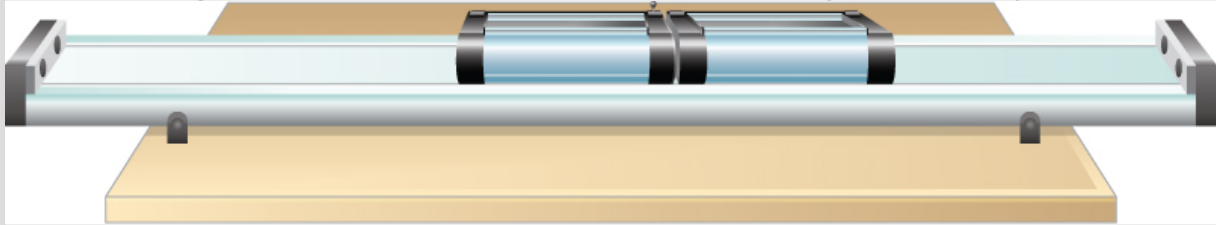
Conservation of Momentum

Using conservation of momentum requires four basic steps. The first step is crucial:

1. Identify a closed system (total mass is constant, no net external force acts on the system).
2. Write down an expression representing the total momentum of the system before the “event” (explosion or collision).
3. Write down an expression representing the total momentum of the system after the “event.”
4. Set these two expressions equal to each other, and solve this equation for the desired quantity.

Example:
Colliding Carts

Two carts in a physics lab roll on a level track, with negligible friction. These carts have small magnets at their ends, so that when they collide, they stick together ([\[link\]](#)). The first cart has a mass of 675 grams and is rolling at 0.75 m/s to the right; the second has a mass of 500 grams and is rolling at 1.33 m/s, also to the right. After the collision, what is the velocity of the two joined carts?



Two lab carts collide and stick together after the collision.

Strategy

We have a collision. We’re given masses and initial velocities; we’re asked for the final velocity. This all suggests using conservation of momentum as a method of solution. However, we can only use it if we have a closed system. So we need to be sure that the system we choose has no net external force on it, and that its mass is not changed by the collision.

Defining the system to be the two carts meets the requirements for a closed system: The combined mass of the two carts certainly doesn’t change, and while the carts definitely exert forces on each other, those forces are internal to the system, so they do not change the momentum of the system as a whole. In the

vertical direction, the weights of the carts are canceled by the normal forces on the carts from the track.

Solution

Conservation of momentum is

Equation:

$$\vec{\mathbf{p}}_{\mathbf{f}} = \vec{\mathbf{p}}_{\mathbf{i}}.$$

Define the direction of their initial velocity vectors to be the +x-direction. The initial momentum is then

Equation:

$$\vec{\mathbf{p}}_{\mathbf{i}} = m_1 v_1 \hat{\mathbf{i}} + m_2 v_2 \hat{\mathbf{i}}.$$

The final momentum of the now-linked carts is

Equation:

$$\vec{\mathbf{p}}_{\mathbf{f}} = (m_1 + m_2) \vec{\mathbf{v}}_{\mathbf{f}}.$$

Equating:

Equation:

$$\begin{aligned} (m_1 + m_2) \vec{\mathbf{v}}_{\mathbf{f}} &= m_1 v_1 \hat{\mathbf{i}} + m_2 v_2 \hat{\mathbf{i}} \\ \vec{\mathbf{v}}_{\mathbf{f}} &= \left(\frac{m_1 v_1 + m_2 v_2}{m_1 + m_2} \right) \hat{\mathbf{i}}. \end{aligned}$$

Substituting the given numbers:

Equation:

$$\begin{aligned} \vec{\mathbf{v}}_{\mathbf{f}} &= \left[\frac{(0.675 \text{ kg})(0.75 \text{ m/s}) + (0.5 \text{ kg})(1.33 \text{ m/s})}{1.175 \text{ kg}} \right] \hat{\mathbf{i}} \\ &= (0.997 \text{ m/s}) \hat{\mathbf{i}}. \end{aligned}$$

Significance

The principles that apply here to two laboratory carts apply identically to all objects of whatever type or size. Even for photons, the concepts of momentum and conservation of momentum are still crucially important even at that scale. (Since they are massless, the momentum of a photon is defined very differently from the momentum of ordinary objects. You will learn about this when you study quantum physics.)

Note:**Exercise:****Problem:**

Check Your Understanding Suppose the second, smaller cart had been initially moving to the left. What would the sign of the final velocity have been in this case?

Solution:

If the smaller cart were rolling at 1.33 m/s to the left, then conservation of momentum gives

$$\begin{aligned}(m_1 + m_2)\vec{v}_f &= m_1 v_1 \hat{\mathbf{i}} - m_2 v_2 \hat{\mathbf{i}} \\ \vec{v}_f &= \left(\frac{m_1 v_1 - m_2 v_2}{m_1 + m_2} \right) \hat{\mathbf{i}} \\ &= \left[\frac{(0.675 \text{ kg})(0.75 \text{ m/s}) - (0.500 \text{ kg})(1.33 \text{ m/s})}{1.175 \text{ kg}} \right] \hat{\mathbf{i}} \\ &= -(0.135 \text{ m/s}) \hat{\mathbf{i}}\end{aligned}$$

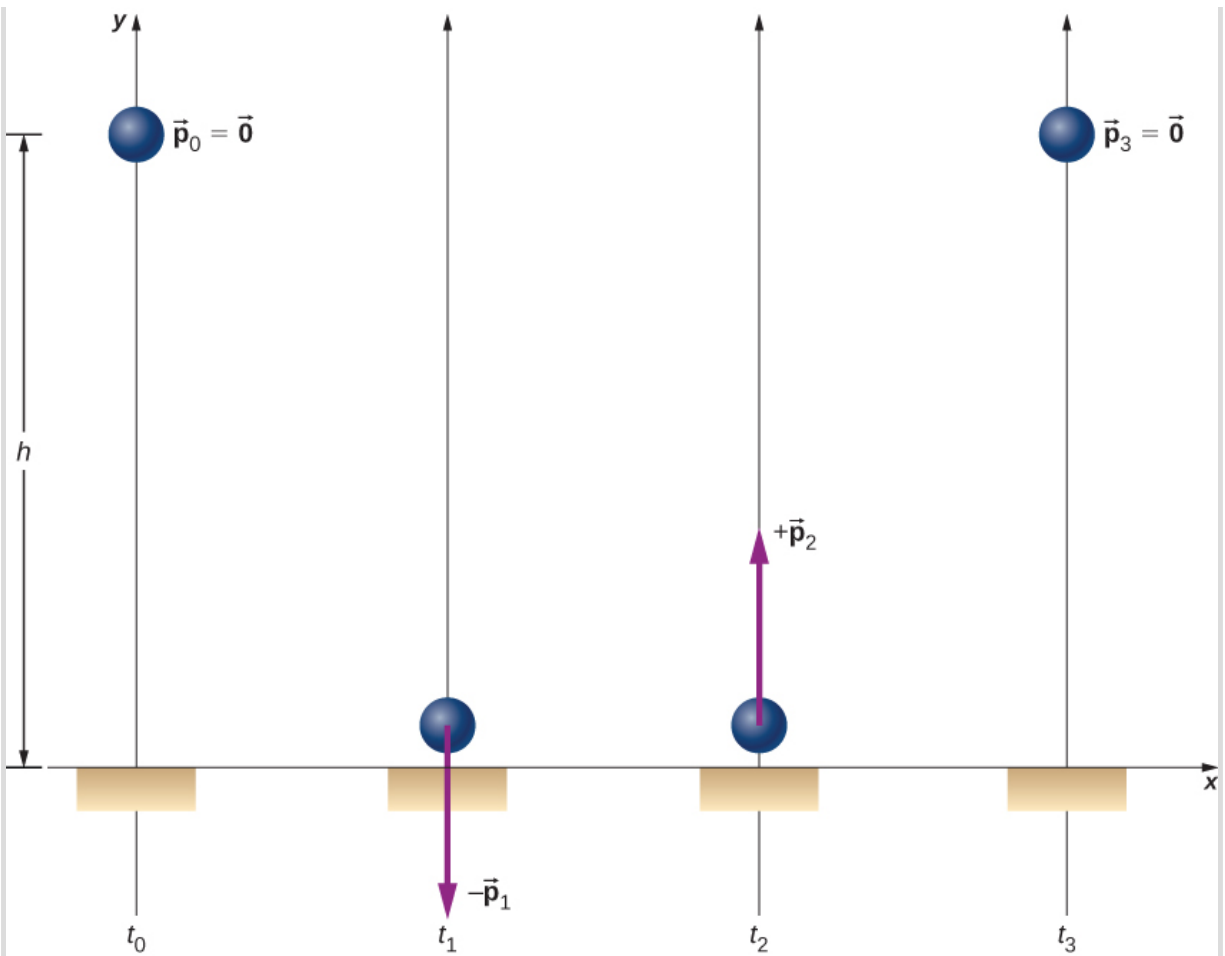
Thus, the final velocity is 0.135 m/s to the left.

Example:**A Bouncing Superball**

A superball of mass 0.25 kg is dropped from rest from a height of $h = 1.50$ m above the floor. It bounces with no loss of energy and returns to its initial height ([link](#)).

- What is the superball's change of momentum during its bounce on the floor?
- What was Earth's change of momentum due to the ball colliding with the floor?
- What was Earth's change of velocity as a result of this collision?

(This example shows that you have to be careful about defining your system.)



A superball is dropped to the floor (t_0), hits the floor (t_1), bounces (t_2), and returns to its initial height (t_3).

Strategy

Since we are asked only about the ball's change of momentum, we define our system to be the ball. But this is clearly not a closed system; gravity applies a downward force on the ball while it is falling, and the normal force from the floor applies a force during the bounce. Thus, we cannot use conservation of momentum as a strategy. Instead, we simply determine the ball's momentum just before it collides with the floor and just after, and calculate the difference. We have the ball's mass, so we need its velocities.

Solution

- Since this is a one-dimensional problem, we use the scalar form of the equations. Let:

- p_0 = the magnitude of the ball's momentum at time t_0 , the moment it was released; since it was dropped from rest, this is zero.
- p_1 = the magnitude of the ball's momentum at time t_1 , the instant just before it hits the floor.
- p_2 = the magnitude of the ball's momentum at time t_2 , just after it loses contact with the floor after the bounce.

The ball's change of momentum is

Equation:

$$\begin{aligned}\Delta \vec{p} &= \vec{p}_2 - \vec{p}_1 \\ &= p_2 \hat{j} - (-p_1 \hat{j}) \\ &= (p_2 + p_1) \hat{j}.\end{aligned}$$

Its velocity just before it hits the floor can be determined from either conservation of energy or kinematics. We use kinematics here; you should re-solve it using conservation of energy and confirm you get the same result.

We want the velocity just before it hits the ground (at time t_1). We know its initial velocity $v_0 = 0$ (at time t_0), the height it falls, and its acceleration; we don't know the fall time. We could calculate that, but instead we use

Equation:

$$\vec{v}_1 = -\hat{j}\sqrt{2gy} = -5.4 \text{ m/s}\hat{j}.$$

Thus the ball has a momentum of

Equation:

$$\begin{aligned}\vec{p}_1 &= -(0.25 \text{ kg}) (-5.4 \text{ m/s}\hat{j}) \\ &= -(1.4 \text{ kg} \cdot \text{m/s})\hat{j}.\end{aligned}$$

We don't have an easy way to calculate the momentum after the bounce. Instead, we reason from the symmetry of the situation.

Before the bounce, the ball starts with zero velocity and falls 1.50 m under the influence of gravity, achieving some amount of momentum just before

it hits the ground. On the return trip (after the bounce), it starts with some amount of momentum, rises the same 1.50 m it fell, and ends with zero velocity. Thus, the motion after the bounce was the mirror image of the motion before the bounce. From this symmetry, it must be true that the ball's momentum after the bounce must be equal and opposite to its momentum before the bounce. (This is a subtle but crucial argument; make sure you understand it before you go on.)

Therefore,

Equation:

$$\vec{p}_2 = -\vec{p}_1 = + (1.4 \text{ kg} \cdot \text{m/s})\hat{j}.$$

Thus, the ball's change of momentum during the bounce is

Equation:

$$\begin{aligned}\Delta\vec{p} &= \vec{p}_2 - \vec{p}_1 \\ &= (1.4 \text{ kg} \cdot \text{m/s})\hat{j} - (-1.4 \text{ kg} \cdot \text{m/s})\hat{j} \\ &= + (2.8 \text{ kg} \cdot \text{m/s})\hat{j}.\end{aligned}$$

- b. What was Earth's change of momentum due to the ball colliding with the floor?

Your instinctive response may well have been either “zero; the Earth is just too massive for that tiny ball to have affected it” or possibly, “more than zero, but utterly negligible.” But no—if we re-define our system to be the Superball + Earth, then this system is closed (neglecting the gravitational pulls of the Sun, the Moon, and the other planets in the solar system), and therefore the total change of momentum of this new system must be zero. Therefore, Earth's change of momentum is exactly the same magnitude:

Equation:

$$\Delta\vec{p}_{\text{Earth}} = -2.8 \text{ kg} \cdot \text{m/s}\hat{j}.$$

- c. What was Earth's change of velocity as a result of this collision?

This is where your instinctive feeling is probably correct:

Equation:

$$\begin{aligned}
 \Delta \vec{v}_{\text{Earth}} &= \frac{\Delta \vec{p}_{\text{Earth}}}{M_{\text{Earth}}} \\
 &= -\frac{2.8 \text{ kg}\cdot\text{m/s}}{5.97 \times 10^{24} \text{ kg}} \hat{\mathbf{j}} \\
 &= -(4.7 \times 10^{-25} \text{ m/s}) \hat{\mathbf{j}}.
 \end{aligned}$$

This change of Earth's velocity *is* utterly negligible.

Significance

It is important to realize that the answer to part (c) is not a velocity; it is a change of velocity, which is a very different thing. Nevertheless, to give you a feel for just how small that change of velocity is, suppose you were moving with a velocity of $4.7 \times 10^{-25} \text{ m/s}$. At this speed, it would take you about 7 million years to travel a distance equal to the diameter of a hydrogen atom.

Note:

Exercise:

Problem:

Check Your Understanding Would the ball's change of momentum have been larger, smaller, or the same, if it had collided with the floor and stopped (without bouncing)?

Would the ball's change of momentum have been larger, smaller, or the same, if it had collided with the floor and stopped (without bouncing)?

Solution:

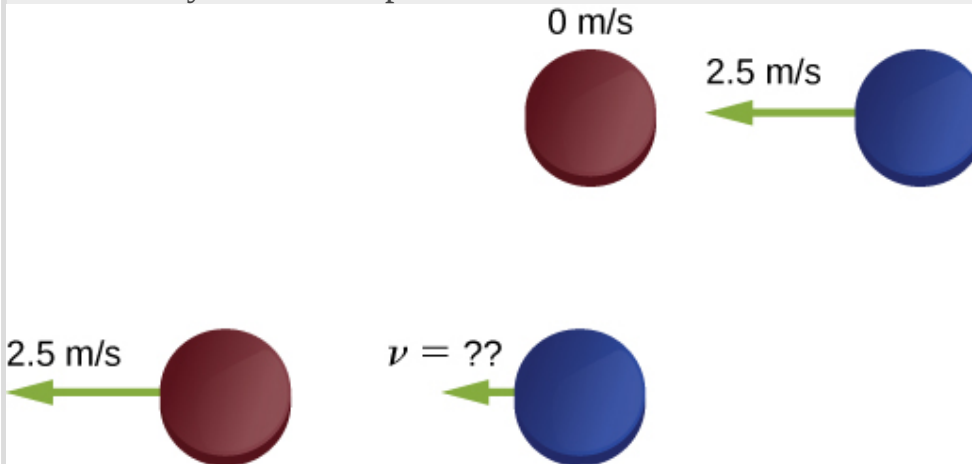
If the ball does not bounce, its final momentum \vec{p}_2 is zero, so

$$\begin{aligned}
 \Delta \vec{p} &= \vec{p}_2 - \vec{p}_1 \\
 &= (0) \hat{\mathbf{j}} - (-1.4 \text{ kg} \cdot \text{m/s}) \hat{\mathbf{j}} \\
 &= + (1.4 \text{ kg} \cdot \text{m/s}) \hat{\mathbf{j}}
 \end{aligned}$$

Example:

Ice Hockey 1

Two hockey pucks of identical mass are on a flat, horizontal ice hockey rink. The red puck is motionless; the blue puck is moving at 2.5 m/s to the left ([link](#)). It collides with the motionless red puck. The pucks have a mass of 15 g. After the collision, the red puck is moving at 2.5 m/s, to the left. What is the final velocity of the blue puck?



Two identical hockey pucks colliding. The top diagram shows the pucks the instant before the collision, and the bottom diagram show the pucks the instant after the collision. The net external force is zero.

Strategy

We're told that we have two colliding objects, we're told the masses and initial velocities, and one final velocity; we're asked for both final velocities.

Conservation of momentum seems like a good strategy. Define the system to be the two pucks; there's no friction, so we have a closed system.

Before you look at the solution, what do you think the answer will be?

The blue puck final velocity will be:

- zero
- 2.5 m/s to the left
- 2.5 m/s to the right
- 1.25 m/s to the left
- 1.25 m/s to the right
- something else

Solution

Define the $+x$ -direction to point to the right. Conservation of momentum then reads

Equation:

$$\begin{aligned}\vec{\mathbf{p}}_{\mathbf{f}} &= \vec{\mathbf{p}}_{\mathbf{i}} \\ mv_{\mathbf{r}\mathbf{f}}\hat{\mathbf{i}} + mv_{\mathbf{b}\mathbf{f}}\hat{\mathbf{i}} &= mv_{\mathbf{r}\mathbf{i}}\hat{\mathbf{i}} - mv_{\mathbf{b}\mathbf{i}}\hat{\mathbf{i}}.\end{aligned}$$

Before the collision, the momentum of the system is entirely and only in the blue puck. Thus,

Equation:

$$\begin{aligned}mv_{\mathbf{r}\mathbf{f}}\hat{\mathbf{i}} + mv_{\mathbf{b}\mathbf{f}}\hat{\mathbf{i}} &= -mv_{\mathbf{b}\mathbf{i}}\hat{\mathbf{i}} \\ v_{\mathbf{r}\mathbf{f}}\hat{\mathbf{i}} + v_{\mathbf{b}\mathbf{f}}\hat{\mathbf{i}} &= -v_{\mathbf{b}\mathbf{i}}\hat{\mathbf{i}}.\end{aligned}$$

(Remember that the masses of the pucks are equal.) Substituting numbers:

Equation:

$$\begin{aligned}-(2.5 \text{ m/s})\hat{\mathbf{i}} + \vec{\mathbf{v}}_{\mathbf{b}\mathbf{f}} &= -(2.5 \text{ m/s})\hat{\mathbf{i}} \\ \vec{\mathbf{v}}_{\mathbf{b}\mathbf{f}} &= 0.\end{aligned}$$

Significance

Evidently, the two pucks simply exchanged momentum. The blue puck transferred all of its momentum to the red puck. In fact, this is what happens in similar collision where $m_1 = m_2$.

Note:

Exercise:

Problem:

Check Your Understanding Even if there were some friction on the ice, it is still possible to use conservation of momentum to solve this problem, but you would need to impose an additional condition on the problem. What is that additional condition?

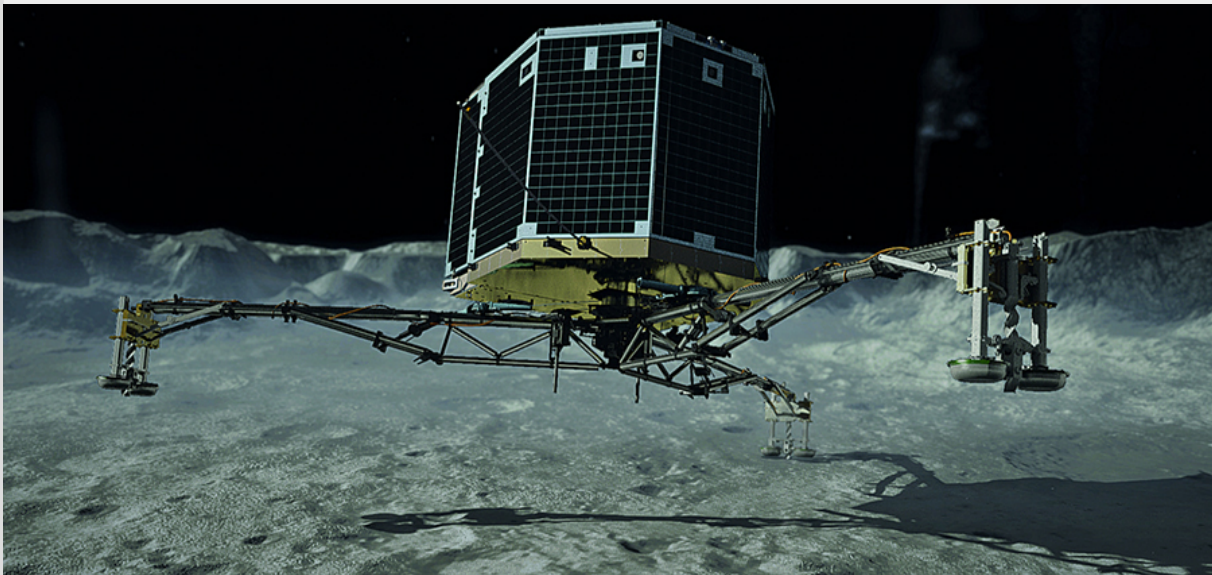
Solution:

Consider the impulse momentum theory, which is $\vec{J} = \Delta\vec{p}$. If $\vec{J} = 0$, we have the situation described in the example. If a force acts on the system, then $\vec{J} = \vec{F}_{\text{ave}}\Delta t$. Thus, instead of $\vec{p}_f = \vec{p}_i$, we have $\vec{F}_{\text{ave}}\Delta t = \Delta\vec{p} = \vec{p}_f - \vec{p}_i$ where \vec{F}_{ave} is the force due to friction.

Example:

Landing of *Philae*

On November 12, 2014, the European Space Agency successfully landed a probe named *Philae* on Comet 67P/Churyumov/Gerasimenko ([\[link\]](#)). During the landing, however, the probe actually landed three times, because it bounced twice. Let's calculate how much the comet's speed changed as a result of the first bounce.



An artist's rendering of *Philae* landing on a comet. (credit: modification of work by "DLR German Aerospace Center"/Flickr)

Let's define upward to be the +y-direction, perpendicular to the surface of the comet, and $y = 0$ to be at the surface of the comet. Here's what we know:

- The mass of Comet 67P: $M_c = 1.0 \times 10^{13} \text{ kg}$

- The acceleration due to the comet's gravity: $\vec{a} = - (5.0 \times 10^{-3} \text{ m/s}^2) \hat{j}$
- *Philae's* mass: $M_p = 96 \text{ kg}$
- Initial touchdown speed: $\vec{v}_1 = - (1.0 \text{ m/s}) \hat{j}$
- Initial upward speed due to first bounce: $\vec{v}_2 = (0.38 \text{ m/s}) \hat{j}$
- Landing impact time: $\Delta t = 1.3 \text{ s}$

Strategy

We're asked for how much the comet's speed changed, but we don't know much about the comet, beyond its mass and the acceleration its gravity causes. However, we *are* told that the *Philae* lander collides with (lands on) the comet, and bounces off of it. A collision suggests momentum as a strategy for solving this problem.

If we define a system that consists of both *Philae* and Comet 67/P, then there is no net external force on this system, and thus the momentum of this system is conserved. (We'll neglect the gravitational force of the sun.) Thus, if we calculate the change of momentum of the lander, we automatically have the change of momentum of the comet. Also, the comet's change of velocity is directly related to its change of momentum as a result of the lander "colliding" with it.

Solution

Let \vec{p}_1 be *Philae's* momentum at the moment just before touchdown, and \vec{p}_2 be its momentum just after the first bounce. Then its momentum just before landing was

Equation:

$$\vec{p}_1 = M_p \vec{v}_1 = (96 \text{ kg}) (-1.0 \text{ m/s} \hat{j}) = - (96 \text{ kg} \cdot \text{m/s}) \hat{j}$$

and just after was

Equation:

$$\vec{p}_2 = M_p \vec{v}_2 = (96 \text{ kg}) (+0.38 \text{ m/s} \hat{j}) = (36.5 \text{ kg} \cdot \text{m/s}) \hat{j}.$$

Therefore, the lander's change of momentum during the first bounce is

Equation:

$$\Delta \vec{p} = \vec{p}_2 - \vec{p}_1$$

$$= (36.5 \text{ kg} \cdot \text{m/s})\hat{j} - (-96.0 \text{ kg} \cdot \text{m/s})\hat{j} = (133 \text{ kg} \cdot \text{m/s})\hat{j}$$

Notice how important it is to include the negative sign of the initial momentum. Now for the comet. Since momentum of the system must be conserved, the *comet's* momentum changed by exactly the negative of this:

Equation:

$$\Delta \vec{p}_c = -\Delta \vec{p} = -(133 \text{ kg} \cdot \text{m/s})\hat{j}.$$

Therefore, its change of velocity is

Equation:

$$\Delta \vec{v}_c = \frac{\Delta \vec{p}_c}{M_c} = \frac{-(133 \text{ kg} \cdot \text{m/s})\hat{j}}{1.0 \times 10^{13} \text{ kg}} = -(1.33 \times 10^{-11} \text{ m/s})\hat{j}.$$

Significance

This is a very small change in velocity, about a thousandth of a billionth of a meter per second. Crucially, however, it is *not* zero.

Note:

Exercise:

Problem:

Check Your Understanding The changes of momentum for *Philae* and for Comet 67/P were equal (in magnitude). Were the impulses experienced by *Philae* and the comet equal? How about the forces? How about the changes of kinetic energies?

Solution:

The impulse is the change in momentum multiplied by the time required for the change to occur. By conservation of momentum, the changes in momentum of the probe and the comet are of the same magnitude, but in opposite directions, and the interaction time for each is also the same. Therefore, the impulse each receives is of the same magnitude, but in

opposite directions. Because they act in opposite directions, the impulses are not the same. As for the impulse, the force on each body acts in opposite directions, so the forces on each are not equal. However, the change in kinetic energy differs for each, because the collision is not elastic.

Summary

- The law of conservation of momentum says that the momentum of a closed system is constant in time (conserved).
- A closed (or isolated) system is defined to be one for which the mass remains constant, and the net external force is zero.
- The total momentum of a system is conserved *only* when the system is closed.

Conceptual Questions

Exercise:

Problem: Under what circumstances is momentum conserved?

Solution:

Momentum is conserved when the mass of the system of interest remains constant during the interaction in question and when no *net* external force acts on the system during the interaction.

Exercise:

Problem:

Can momentum be conserved for a system if there are external forces acting on the system? If so, under what conditions? If not, why not?

Exercise:

Problem:

Explain in terms of momentum and Newton's laws how a car's air resistance is due in part to the fact that it pushes air in its direction of motion.

Solution:

To accelerate air molecules in the direction of motion of the car, the car must exert a force on these molecules by Newton's second law $\vec{F} = d\vec{p}/dt$. By Newton's third law, the air molecules exert a force of equal magnitude but in the opposite direction on the car. This force acts in the direction opposite the motion of the car and constitutes the force due to air resistance.

Exercise:**Problem:**

Can objects in a system have momentum while the momentum of the system is zero? Explain your answer.

Exercise:**Problem:**

A sprinter accelerates out of the starting blocks. Can you consider him as a closed system? Explain.

Solution:

No, he is not a closed system because a net nonzero external force acts on him in the form of the starting blocks pushing on his feet.

Exercise:**Problem:**

A rocket in deep space (zero gravity) accelerates by firing hot gas out of its thrusters. Does the rocket constitute a closed system? Explain.

Problems

Exercise:**Problem:**

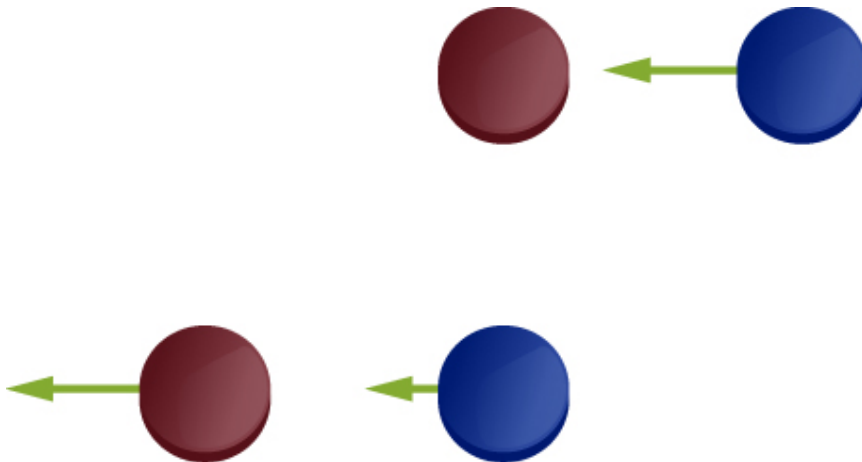
Train cars are coupled together by being bumped into one another. Suppose two loaded train cars are moving toward one another, the first having a mass of $1.50 \times 10^5 \text{ kg}$ and a velocity of $(0.30 \text{ m/s})\hat{\mathbf{i}}$, and the second having a mass of $1.10 \times 10^5 \text{ kg}$ and a velocity of $-(0.12 \text{ m/s})\hat{\mathbf{i}}$. What is their final velocity?

**Solution:**

$$(0.122 \text{ m/s})\hat{\mathbf{i}}$$

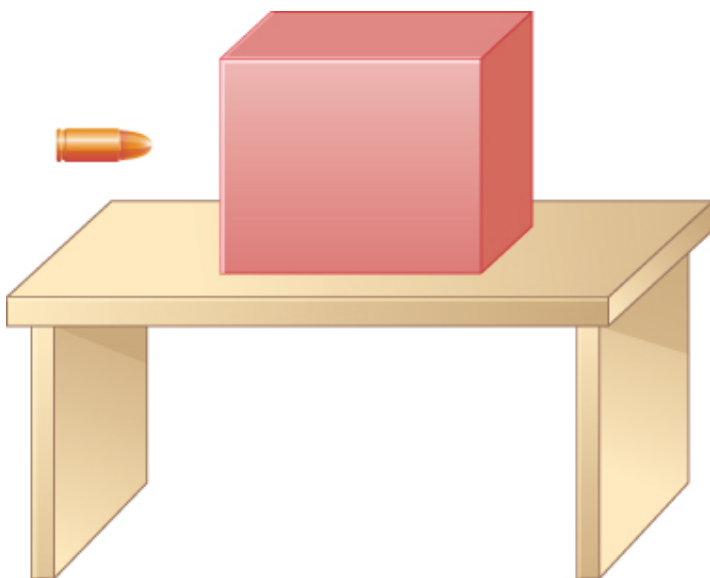
Exercise:**Problem:**

Two identical pucks collide elastically on an air hockey table. Puck 1 was originally at rest; puck 2 has an incoming speed of 6.00 m/s and scatters at an angle of 30° with respect to its incoming direction. What is the velocity (magnitude and direction) of puck 1 after the collision?



Exercise:**Problem:**

The figure below shows a bullet of mass 200 g traveling horizontally towards the east with speed 400 m/s, which strikes a block of mass 1.5 kg that is initially at rest on a frictionless table.



After striking the block, the bullet is embedded in the block and the block and the bullet move together as one unit.

- What is the magnitude and direction of the velocity of the block/bullet combination immediately after the impact?
- What is the magnitude and direction of the impulse by the block on the bullet?
- What is the magnitude and direction of the impulse from the bullet on the block?
- If it took 3 ms for the bullet to change the speed from 400 m/s to the final speed after impact, what is the average force between the block and the bullet during this time?

Solution:

- a. 47 m/s in the bullet to block direction; b. $70.6 \text{ N} \cdot \text{s}$, toward the bullet; c. $70.6 \text{ N} \cdot \text{s}$, toward the block; d. magnitude is $2.35 \times 10^4 \text{ N}$

Exercise:**Problem:**

A 20-kg child is coasting at 3.3 m/s over flat ground in a 4.0-kg wagon. The child drops a 1.0-kg ball out the back of the wagon. What is the final speed of the child and wagon?

Exercise:**Problem:**

A 4.5 kg puffer fish expands to 40% of its mass by taking in water. When the puffer fish is threatened, it releases the water toward the threat to move quickly forward. What is the ratio of the speed of the puffer fish forward to the speed of the expelled water backwards?

Solution:

2:5

Exercise:

Problem: Explain why a cannon recoils when it fires a shell.

Exercise:**Problem:**

Two figure skaters are coasting in the same direction, with the leading skater moving at 5.5 m/s and the trailing skating moving at 6.2 m/s. When the trailing skater catches up with the leading skater, he picks her up without applying any horizontal forces on his skates. If the trailing skater is 50% heavier than the 50-kg leading skater, what is their speed after he picks her up?

Solution:

5.9 m/s

Exercise:

Problem:

A 2000-kg railway freight car coasts at 4.4 m/s underneath a grain terminal, which dumps grain directly down into the freight car. If the speed of the loaded freight car must not go below 3.0 m/s, what is the maximum mass of grain that it can accept?

Glossary

closed system

system for which the mass is constant and the net external force on the system is zero

Law of Conservation of Momentum

total momentum of a closed system cannot change

system

object or collection of objects whose motion is currently under investigation; however, your system is defined at the start of the problem, you must keep that definition for the entire problem

Types of Collisions

By the end of this section, you will be able to:

- Identify the type of collision
- Correctly label a collision as elastic or inelastic
- Use kinetic energy along with momentum and impulse to analyze a collision

Although momentum is conserved in all interactions, not all interactions (collisions or explosions) are the same. The possibilities include:

- A single object can explode into multiple objects (explosions).
- Multiple objects can collide and stick together, forming a single object (inelastic).
- Multiple objects can collide and bounce off of each other, remaining as multiple objects (elastic). If they do bounce off each other, then they may recoil at the same speeds with which they approached each other before the collision, or they may move off more slowly.

It's useful, therefore, to categorize different types of interactions, according to how the interacting objects move before and after the interaction.

Explosions

The first possibility is that a single object may break apart into two or more pieces. An example of this is a firecracker, or a bow and arrow, or a rocket rising through the air toward space. These can be difficult to analyze if the number of fragments after the collision is more than about three or four; but nevertheless, the total momentum of the system before and after the explosion is identical.

Note that if the object is initially motionless, then the system (which is just the object) has no momentum and no kinetic energy. After the explosion, the net momentum of all the pieces of the object must sum to zero (since the momentum of this closed system cannot change). However, the system *will* have a great deal of kinetic energy after the explosion, although it had none before. Thus, we see that, although the momentum of the system is

conserved in an explosion, the kinetic energy of the system most definitely is not; it increases. This interaction—one object becoming many, with an increase of kinetic energy of the system—is called an **explosion**.

Where does the energy come from? Does conservation of energy still hold? Yes; some form of potential energy is converted to kinetic energy. In the case of gunpowder burning and pushing out a bullet, chemical potential energy is converted to kinetic energy of the bullet, and of the recoiling gun. For a bow and arrow, it is elastic potential energy in the bowstring.

Inelastic

The second possibility is the reverse: that two or more objects collide with each other and stick together, thus (after the collision) forming one single composite object. The total mass of this composite object is the sum of the masses of the original objects, and the new single object moves with a velocity dictated by the conservation of momentum. However, it turns out again that, although the total momentum of the system of objects remains constant, the kinetic energy doesn't; but this time, the kinetic energy decreases. This type of collision is called **inelastic**.

Any collision where the objects stick together will result in the maximum loss of kinetic energy (i.e., K_f will be a minimum).

Such a collision is called **perfectly inelastic**. In the extreme case, multiple objects collide, stick together, and remain motionless after the collision. Since the objects are all motionless after the collision, the final kinetic energy is also zero; therefore, the loss of kinetic energy is a maximum.

- If $0 < K_f < K_i$, the collision is inelastic.
- If K_f is the lowest energy, or the energy lost by both objects is the most, the collision is perfectly inelastic (objects stick together).
- If $K_f = K_i$, the collision is elastic.

Elastic

The extreme case on the other end is if two or more objects approach each other, collide, and bounce off each other, moving away from each other at the same relative speed at which they approached each other. In this case, the total kinetic energy of the system is conserved. Such an interaction is called **elastic**.

In any interaction of a closed system of objects, the total momentum of the system is conserved ($\vec{p}_f = \vec{p}_i$) but the kinetic energy may not be:

- If $0 < K_f < K_i$, the collision is inelastic.
- If $K_f = 0$, the collision is perfectly inelastic.
- If $K_f = K_i$, the collision is elastic.
- If $K_f > K_i$, the interaction is an explosion.

The point of all this is that, in analyzing a collision or explosion, you can use both momentum and kinetic energy.

Note:

Collisions

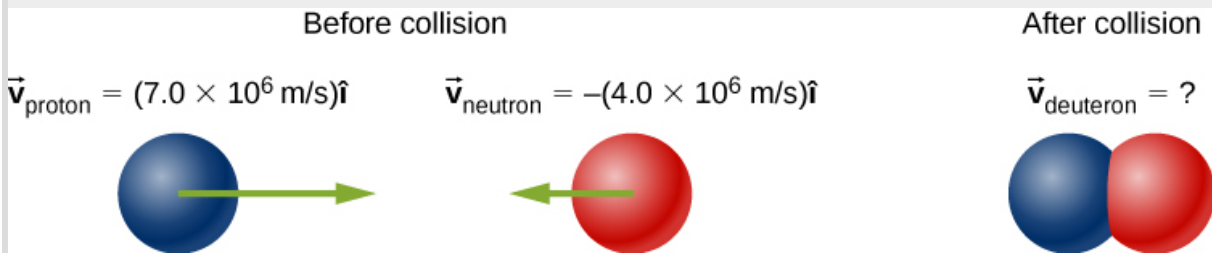
A closed system always conserves momentum; it might also conserve kinetic energy, but very often it doesn't. Energy-momentum problems confined to a plane (as ours are) usually have two unknowns. Generally, this approach works well:

1. Define a closed system.
2. Write down the expression for conservation of momentum.
3. If kinetic energy is conserved, write down the expression for conservation of kinetic energy; if not, write down the expression for the change of kinetic energy.
4. You now have two equations in two unknowns, which you solve by standard methods.

Example:

Formation of a Deuteron

A proton (mass 1.67×10^{-27} kg) collides with a neutron (with essentially the same mass as the proton) to form a particle called a *deuteron*. What is the velocity of the deuteron if it is formed from a proton moving with velocity 7.0×10^6 m/s to the left and a neutron moving with velocity 4.0×10^6 m/s to the right?



Strategy

Define the system to be the two particles. This is a collision, so we should first identify what kind. Since we are told the two particles form a single particle after the collision, this means that the collision is perfectly inelastic. Thus, kinetic energy is not conserved, but momentum is. Thus, we use conservation of momentum to determine the final velocity of the system.

Solution

Treat the two particles as having identical masses M . Use the subscripts p, n, and d for proton, neutron, and deuteron, respectively. This is a one-dimensional problem, so we have

Equation:

$$Mv_p - Mv_n = 2Mv_d.$$

The masses divide out:

Equation:

$$\begin{aligned}v_p - v_n &= 2v_d \\7.0 \times 10^6 \text{ m/s} - 4.0 \times 10^6 \text{ m/s} &= 2v_d \\v_d &= 1.5 \times 10^6 \text{ m/s}.\end{aligned}$$

The velocity is thus $\vec{v}_d = (1.5 \times 10^6 \text{ m/s})\hat{i}$.

Significance

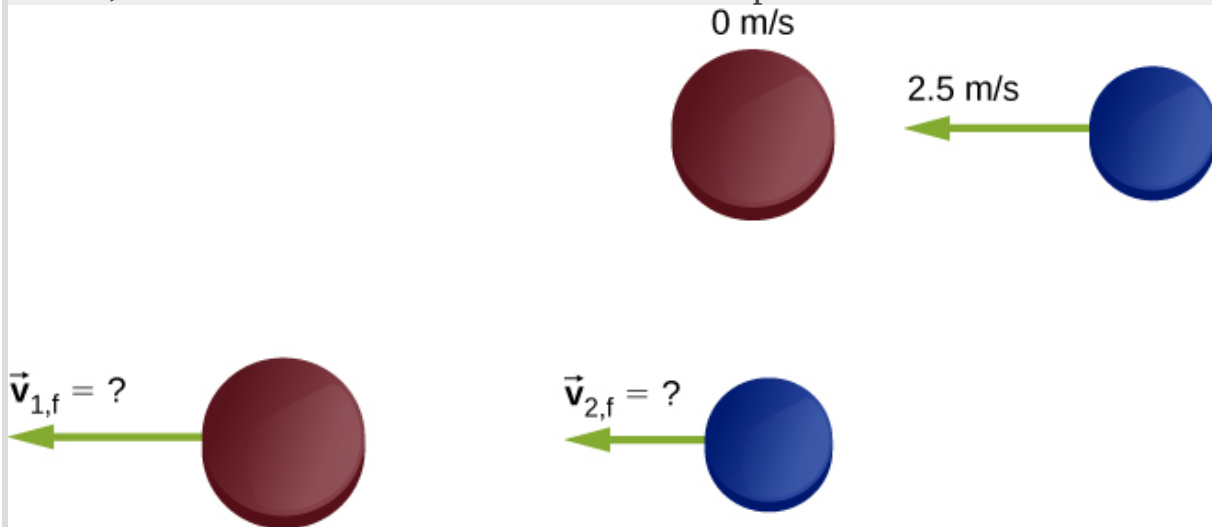
This is essentially how particle colliders like the Large Hadron Collider work: They accelerate particles up to very high speeds (large momenta), but in opposite directions. This maximizes the creation of so-called “daughter particles.”

Example:

Ice Hockey 2

(This is a variation of an earlier example.)

Two ice hockey pucks of different masses are on a flat, horizontal hockey rink. The red puck has a mass of 15 grams, and is motionless; the blue puck has a mass of 12 grams, and is moving at 2.5 m/s to the left. It collides with the motionless red puck ([link](#)). If the collision is perfectly elastic, what are the final velocities of the two pucks?



Two different hockey pucks colliding. The top diagram shows the pucks the instant before the collision, and the bottom diagram show the pucks the instant after the collision. The net external force is zero.

Strategy

We're told that we have two colliding objects, and we're told their masses and initial velocities; we're asked for both final velocities. Conservation of momentum seems like a good strategy; define the system to be the two

pucks. There is no friction, so we have a closed system. We have two unknowns (the two final velocities), but only one equation. The comment about the collision being perfectly elastic is the clue; it suggests that kinetic energy is also conserved in this collision. That gives us our second equation.

The initial momentum and initial kinetic energy of the system resides entirely and only in the second puck (the blue one); the collision transfers some of this momentum and energy to the first puck.

Solution

Conservation of momentum, in this case, reads

Equation:

$$\begin{aligned}p_i &= p_f \\m_2 v_{2,i} &= m_1 v_{1,f} + m_2 v_{2,f}.\end{aligned}$$

Conservation of kinetic energy reads

Equation:

$$\begin{aligned}K_i &= K_f \\ \frac{1}{2} m_2 v_{2,i}^2 &= \frac{1}{2} m_1 v_{1,f}^2 + \frac{1}{2} m_2 v_{2,f}^2.\end{aligned}$$

There are our two equations in two unknowns. The algebra is tedious but not terribly difficult; you definitely should work it through. The solution is

Equation:

$$\begin{aligned}v_{1,f} &= \frac{(m_1 - m_2)v_{1,i} + 2m_2 v_{2,i}}{m_1 + m_2} \\ v_{2,f} &= \frac{(m_2 - m_1)v_{2,i} + 2m_1 v_{1,i}}{m_1 + m_2}.\end{aligned}$$

Substituting the given numbers, we obtain

Equation:

$$\begin{aligned}v_{1,f} &= 2.22 \frac{\text{m}}{\text{s}} \\ v_{2,f} &= -0.28 \frac{\text{m}}{\text{s}}.\end{aligned}$$

Significance

Notice that after the collision, the blue puck is moving to the right; its direction of motion was reversed. The red puck is now moving to the left.

Note:

Exercise:

Problem:

Check Your Understanding There is a second solution to the system of equations solved in this example (because the energy equation is quadratic): $v_{1,f} = -2.5 \text{ m/s}$, $v_{2,f} = 0$. This solution is unacceptable on physical grounds; what's wrong with it?

Solution:

This solution represents the case in which no interaction takes place: the first puck misses the second puck and continues on with a velocity of 2.5 m/s to the left. This case offers no meaningful physical insights.

Example:

Thor vs. Iron Man

The 2012 movie “The Avengers” has a scene where Iron Man and Thor fight. At the beginning of the fight, Thor throws his hammer at Iron Man, hitting him and throwing him slightly up into the air and against a small tree, which breaks. From the video, Iron Man is standing still when the hammer hits him. The distance between Thor and Iron Man is approximately 10 m, and the hammer takes about 1 s to reach Iron Man after Thor releases it. The tree is about 2 m behind Iron Man, which he hits in about 0.75 s. Also from the video, Iron Man's trajectory to the tree is very close to horizontal. Assuming Iron Man's total mass is 200 kg:

- a. Estimate the mass of Thor's hammer
- b. Estimate how much kinetic energy was lost in this collision

Strategy

After the collision, Thor's hammer is in contact with Iron Man for the entire time, so this is a perfectly inelastic collision. Thus, with the correct choice of a closed system, we expect momentum is conserved, but not kinetic energy. We use the given numbers to estimate the initial momentum, the initial kinetic energy, and the final kinetic energy. Because this is a one-dimensional problem, we can go directly to the scalar form of the equations.

Solution

- a. First, we posit conservation of momentum. For that, we need a closed system. The choice here is the system (hammer + Iron Man), from the time of collision to the moment just before Iron Man and the hammer hit the tree. Let:

- M_H = mass of the hammer
- M_I = mass of Iron Man
- v_H = velocity of the hammer before hitting Iron Man
- v = combined velocity of Iron Man + hammer after the collision

Again, Iron Man's initial velocity was zero. Conservation of momentum here reads:

Equation:

$$M_H v_H = (M_H + M_I)v.$$

We are asked to find the mass of the hammer, so we have

Equation:

$$\begin{aligned} M_H v_H &= M_H v + M_I v \\ M_H (v_H - v) &= M_I v \\ M_H &= \frac{M_I v}{v_H - v} \\ &= \frac{(200 \text{ kg}) \left(\frac{2 \text{ m}}{0.75 \text{ s}} \right)}{10 \frac{\text{m}}{\text{s}} - \left(\frac{2 \text{ m}}{0.75 \text{ s}} \right)} \\ &= 73 \text{ kg}. \end{aligned}$$

Considering the uncertainties in our estimates, this should be expressed with just one significant figure; thus, $M_H = 7 \times 10^1 \text{ kg}$.

- b. The initial kinetic energy of the system, like the initial momentum, is all in the hammer:

Equation:

$$\begin{aligned} K_i &= \frac{1}{2} M_H v_H^2 \\ &= \frac{1}{2} (70 \text{ kg})(10 \text{ m/s})^2 \\ &= 3500 \text{ J.} \end{aligned}$$

After the collision,

Equation:

$$\begin{aligned} K_f &= \frac{1}{2} (M_H + M_I) v^2 \\ &= \frac{1}{2} (70 \text{ kg} + 200 \text{ kg})(2.67 \text{ m/s})^2 \\ &= 960 \text{ J.} \end{aligned}$$

Thus, there was a loss of $3500 \text{ J} - 960 \text{ J} = 2540 \text{ J}$.

Significance

From other scenes in the movie, Thor apparently can control the hammer's velocity with his mind. It is possible, therefore, that he mentally causes the hammer to maintain its initial velocity of 10 m/s while Iron Man is being driven backward toward the tree. If so, this would represent an external force on our system, so it would not be closed. Thor's mental control of his hammer is beyond the scope of this book, however.

Example:

Analyzing a Car Crash

At a stoplight, a large truck (3000 kg) collides with a motionless small car (1200 kg). The truck comes to an instantaneous stop; the car slides straight ahead, coming to a stop after sliding 10 meters. The measured coefficient of friction between the car's tires and the road was 0.62. How fast was the truck moving at the moment of impact?

Strategy

At first it may seem we don't have enough information to solve this problem. Although we know the initial speed of the car, we don't know the speed of the truck (indeed, that's what we're asked to find), so we don't know the initial momentum of the system. Similarly, we know the final speed of the truck, but not the speed of the car immediately after impact. The fact that the car eventually slid to a speed of zero doesn't help with the final momentum, since an external friction force caused that. Nor can we calculate an impulse, since we don't know the collision time, or the amount of time the car slid before stopping. A useful strategy is to impose a restriction on the analysis.

Suppose we define a system consisting of just the truck and the car. The momentum of this system isn't conserved, because of the friction between the car and the road. But if we *could* find the speed of the car the instant after impact—before friction had any measurable effect on the car—then we could consider the momentum of the system to be conserved, with that restriction.

Can we find the final speed of the car? Yes; we invoke the work-kinetic energy theorem.

Solution

First, define some variables. Let:

- M_c and M_T be the masses of the car and truck, respectively
- $v_{T,i}$ and $v_{T,f}$ be the velocities of the truck before and after the collision, respectively
- $v_{c,i}$ and $v_{c,f}$ be the velocities of the car before and after the collision, respectively
- K_i and K_f be the kinetic energies of the car immediately after the collision, and after the car has stopped sliding (so $K_f = 0$).
- d be the distance the car slides after the collision before eventually coming to a stop.

Since we actually want the initial speed of the truck, and since the truck is not part of the work-energy calculation, let's start with conservation of momentum. For the car + truck system, conservation of momentum reads

Equation:

$$p_i = p_f$$

$$M_c v_{c,i} + M_T v_{T,i} = M_c v_{c,f} + M_T v_{T,f}.$$

Since the car's initial velocity was zero, as was the truck's final velocity, this simplifies to

Equation:

$$v_{T,i} = \frac{M_c}{M_T} v_{c,f}.$$

So now we need the car's speed immediately after impact. Recall that

Equation:

$$W = \Delta K$$

where

Equation:

$$\begin{aligned} \Delta K &= K_f - K_i \\ &= 0 - \frac{1}{2} M_c v_{c,f}^2. \end{aligned}$$

Also,

Equation:

$$W = \vec{\mathbf{F}} \cdot \vec{\mathbf{d}} = F d \cos \theta.$$

The work is done over the distance the car slides, which we've called d .

Equating:

Equation:

$$F d \cos \theta = -\frac{1}{2} M_c v_{c,f}^2.$$

Friction is the force on the car that does the work to stop the sliding. With a level road, the friction force is

Equation:

$$F = \mu_k M_c g.$$

Since the angle between the directions of the friction force vector and the displacement d is 180° , and $\cos(180^\circ) = -1$, we have

Equation:

$$-(\mu_k M_c g)d = -\frac{1}{2} M_c v_{c,f}^2$$

(Notice that the car's mass divides out; evidently the mass of the car doesn't matter.)

Solving for the car's speed immediately after the collision gives

Equation:

$$v_{c,f} = \sqrt{2\mu_k g d}.$$

Substituting the given numbers:

Equation:

$$\begin{aligned} v_{c,f} &= \sqrt{2(0.62) \left(9.81 \frac{\text{m}}{\text{s}^2}\right) (10 \text{ m})} \\ &= 11.0 \text{ m/s.} \end{aligned}$$

Now we can calculate the initial speed of the truck:

Equation:

$$v_{T,i} = \left(\frac{1200 \text{ kg}}{3000 \text{ kg}} \right) \left(11.0 \frac{\text{m}}{\text{s}} \right) = 4.4 \text{ m/s.}$$

Significance

This is an example of the type of analysis done by investigators of major car accidents. A great deal of legal and financial consequences depend on an accurate analysis and calculation of momentum and energy.

Note:

Exercise:

Problem:

Check Your Understanding Suppose there had been no friction (the collision happened on ice); that would make μ_k zero, and thus $v_{c,f} = \sqrt{2\mu_k g d} = 0$, which is obviously wrong. What is the mistake in this conclusion?

Solution:

If zero friction acts on the car, then it will continue to slide indefinitely ($d \rightarrow \infty$), so we cannot use the work-kinetic-energy theorem as is done in the example. Thus, we could not solve the problem from the information given.

Subatomic Collisions and Momentum

Conservation of momentum is crucial to our understanding of atomic and subatomic particles because much of what we know about these particles comes from collision experiments.

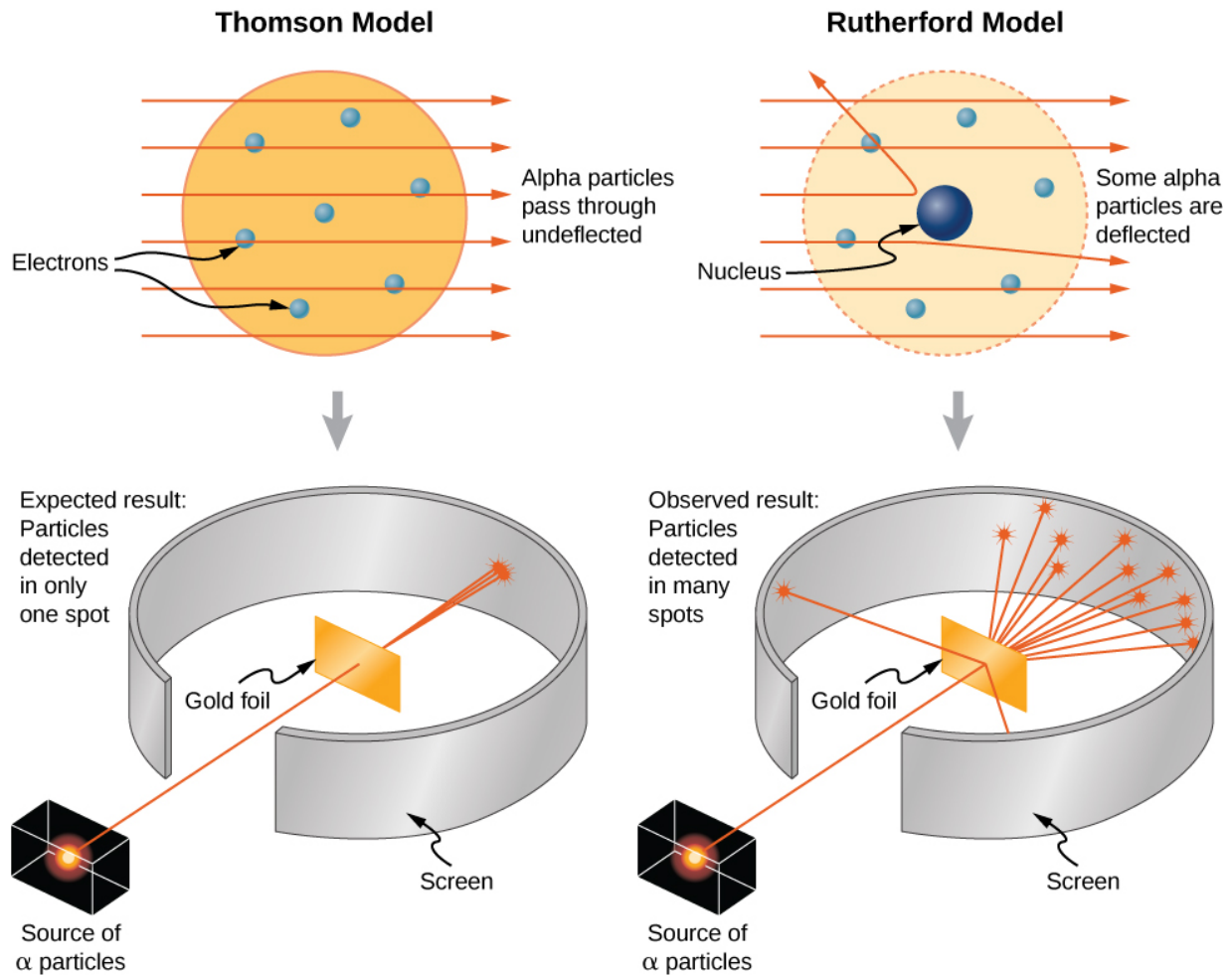
At the beginning of the twentieth century, there was considerable interest in, and debate about, the structure of the atom. It was known that atoms contain two types of electrically charged particles: negatively charged electrons and positively charged protons. (The existence of an electrically neutral particle was suspected, but would not be confirmed until 1932.) The question was, how were these particles arranged in the atom? Were they distributed uniformly throughout the volume of the atom (as J.J. Thomson proposed), or arranged at the corners of regular polygons (which was Gilbert Lewis' model), or rings of negative charge that surround the positively charged nucleus—rather like the planetary rings surrounding Saturn (as suggested by Hantaro Nagaoka), or something else?

The New Zealand physicist Ernest Rutherford (along with the German physicist Hans Geiger and the British physicist Ernest Marsden) performed the crucial experiment in 1909. They bombarded a thin sheet of gold foil

with a beam of high-energy (that is, high-speed) alpha-particles (the nucleus of a helium atom). The alpha-particles collided with the gold atoms, and their subsequent velocities were detected and analyzed, using conservation of momentum and conservation of energy.

If the charges of the gold atoms were distributed uniformly (per Thomson), then the alpha-particles should collide with them and nearly all would be deflected through many angles, all small; the Nagaoka model would produce a similar result. If the atoms were arranged as regular polygons (Lewis), the alpha-particles would deflect at a relatively small number of angles.

What *actually* happened is that nearly *none* of the alpha-particles were deflected. Those that were, were deflected at large angles, some close to 180° —those alpha-particles reversed direction completely ([\[link\]](#)). None of the existing atomic models could explain this. Eventually, Rutherford developed a model of the atom that was much closer to what we now have—again, using conservation of momentum and energy as his starting point.



The Thomson and Rutherford models of the atom. The Thomson model predicted that nearly all of the incident alpha-particles would be scattered and at small angles. Rutherford and Geiger found that nearly none of the alpha particles were scattered, but those few that were deflected did so through very large angles. The results of Rutherford's experiments were inconsistent with the Thomson model. Rutherford used conservation of momentum and energy to develop a new, and better model of the atom—the nuclear model.

Summary

- An elastic collision is one that conserves kinetic energy.

- An inelastic collision does not conserve kinetic energy.
- Momentum is conserved regardless of whether or not kinetic energy is conserved.
- Analysis of kinetic energy changes and conservation of momentum together allow the final velocities to be calculated in terms of initial velocities and masses in one-dimensional, two-body collisions.

Conceptual Questions

Exercise:

Problem:

Two objects of equal mass are moving with equal and opposite velocities when they collide. Can all the kinetic energy be lost in the collision?

Solution:

Yes, all the kinetic energy can be lost if the two masses come to rest due to the collision (i.e., they stick together).

Exercise:

Problem:

Describe a system for which momentum is conserved but mechanical energy is not. Now the reverse: Describe a system for which kinetic energy is conserved but momentum is not.

Problems

Exercise:

Problem:

A 5.50-kg bowling ball moving at 9.00 m/s collides with a 0.850-kg bowling pin, which is scattered at an angle of 15.8° to the initial direction of the bowling ball and with a speed of 15.0 m/s.

- a. Calculate the final velocity (magnitude and direction) of the bowling ball.
- b. Is the collision elastic?

Solution:

- a. 6.80 m/s, 5.33°; b. yes (calculate the ratio of the initial and final kinetic energies)

Exercise:

Problem:

Ernest Rutherford (the first New Zealander to be awarded the Nobel Prize in Chemistry) demonstrated that nuclei were very small and dense by scattering helium-4 nuclei from gold-197 nuclei. The energy of the incoming helium nucleus was 8.00×10^{-13} J, and the masses of the helium and gold nuclei were 6.68×10^{-27} kg and 3.29×10^{-25} kg, respectively (note that their mass ratio is 4 to 197).

- a. If a helium nucleus scatters to an angle of 120° during an elastic collision with a gold nucleus, calculate the helium nucleus's final speed and the final velocity (magnitude and direction) of the gold nucleus.



- b. What is the final kinetic energy of the helium nucleus?

Exercise:

Problem:

A 90.0-kg ice hockey player hits a 0.150-kg puck, giving the puck a velocity of 45.0 m/s. If both are initially at rest and if the ice is frictionless, how far does the player recoil in the time it takes the puck to reach the goal 15.0 m away?

Solution:

2.5 cm

Exercise:**Problem:**

A 100-g firecracker is launched vertically into the air and explodes into two pieces at the peak of its trajectory. If a 72-g piece is projected horizontally to the left at 20 m/s, what is the speed and direction of the other piece?

Exercise:**Problem:**

In an elastic collision, a 400-kg bumper car collides directly from behind with a second, identical bumper car that is traveling in the same direction. The initial speed of the leading bumper car is 5.60 m/s and that of the trailing car is 6.00 m/s. Assuming that the mass of the drivers is much, much less than that of the bumper cars, what are their final speeds?

Solution:

the speed of the leading bumper car is 6.00 m/s and that of the trailing bumper car is 5.60 m/s

Exercise:

Problem:

Repeat the preceding problem if the mass of the leading bumper car is 30.0% greater than that of the trailing bumper car.

Exercise:**Problem:**

An alpha particle (${}^4\text{He}$) undergoes an elastic collision with a stationary uranium nucleus (${}^{235}\text{U}$). What percent of the kinetic energy of the alpha particle is transferred to the uranium nucleus? Assume the collision is one-dimensional.

Solution:

6.6%

Exercise:**Problem:**

You are standing on a very slippery icy surface and throw a 1-kg football horizontally at a speed of 6.7 m/s. What is your velocity when you release the football? Assume your mass is 65 kg.

Exercise:**Problem:**

A 35-kg child rides a relatively massless sled down a hill and then coasts along the flat section at the bottom, where a second 35-kg child jumps on the sled as it passes by her. If the speed of the sled is 3.5 m/s before the second child jumps on, what is its speed after she jumps on?

Solution:

1.8 m/s

Exercise:

Problem:

A boy sleds down a hill and onto a frictionless ice-covered lake at 10.0 m/s. In the middle of the lake is a 1000-kg boulder. When the sled crashes into the boulder, he is propelled backwards from the boulder. The collision is an elastic collision. If the boy's mass is 40.0 kg and the sled's mass is 2.50 kg, what is the speed of the sled and the boulder after the collision?

Glossary

elastic

collision that conserves kinetic energy

explosion

single object breaks up into multiple objects; kinetic energy is not conserved in explosions

inelastic

collision that does not conserve kinetic energy

perfectly inelastic

collision after which all objects are motionless, the final kinetic energy is zero, and the loss of kinetic energy is a maximum

Collisions in Multiple Dimensions

By the end of this section, you will be able to:

- Express momentum as a two-dimensional vector
- Write equations for momentum conservation in component form
- Calculate momentum in two dimensions, as a vector quantity

It is far more common for collisions to occur in two dimensions; that is, the angle between the initial velocity vectors is neither zero nor 180° . Let's see what complications arise from this.

The first idea we need is that momentum is a vector; like all vectors, it can be expressed as a sum of perpendicular components (usually, though not always, an x -component and a y -component, and a z -component if necessary). Thus, when we write down the statement of conservation of momentum for a problem, our momentum vectors can be, and usually will be, expressed in component form.

The second idea we need comes from the fact that momentum is related to force:

Equation:

$$\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}}{dt}.$$

Expressing both the force and the momentum in component form,

Equation:

$$F_x = \frac{dp_x}{dt}, \quad F_y = \frac{dp_y}{dt}, \quad F_z = \frac{dp_z}{dt}.$$

Remember, these equations are simply Newton's second law, in vector form and in component form. We know that Newton's second law is true in each direction, independently of the others. It follows therefore (via Newton's

third law) that conservation of momentum is also true in each direction independently.

These two ideas motivate the solution to two-dimensional problems: We write down the expression for conservation of momentum twice: once in the x -direction and once in the y -direction.

Note:

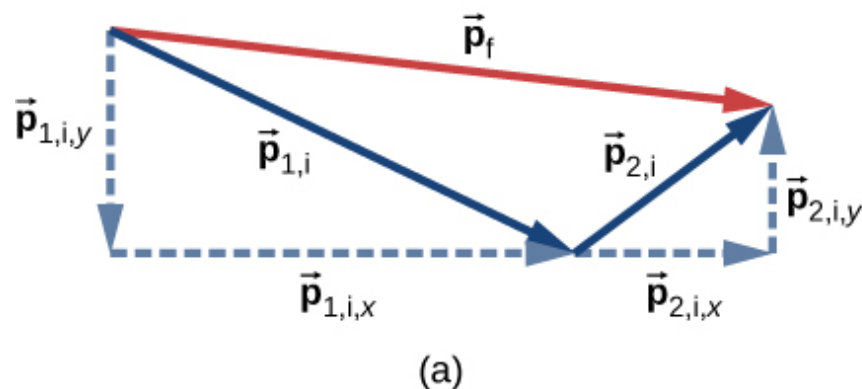
Equation:

$$p_{f,x} = p_{1,i,x} + p_{2,i,x}$$

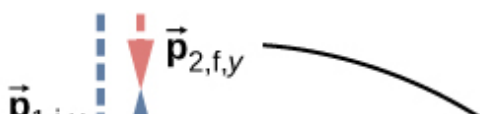
$$p_{f,y} = p_{1,i,y} + p_{2,i,y}$$

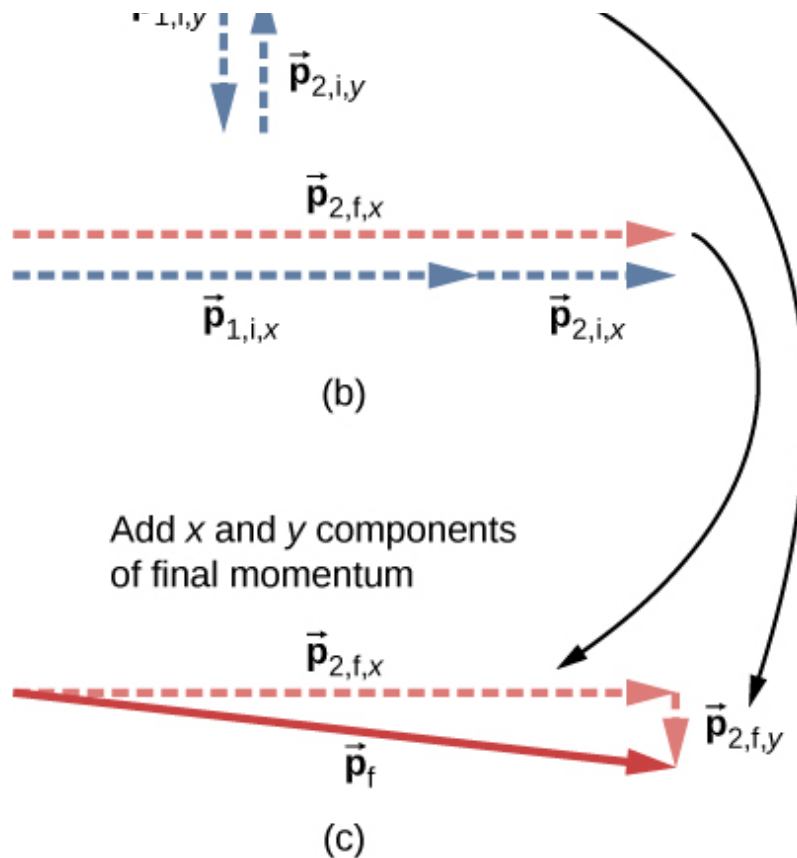
This procedure is shown graphically in [\[link\]](#).

Break initial momentum
into x and y components



Add x and y components
to obtain x and y components
of final momentum





(a) For two-dimensional momentum problems, break the initial momentum vectors into their x- and y-components. (b) Add the x- and y-components together separately. This gives you the x- and y-components of the final momentum, which are shown as red dashed vectors. (c) Adding these components together gives the final momentum.

We solve each of these two component equations independently to obtain the x- and y-components of the desired velocity vector:

Equation:

$$v_{f,x} = \frac{m_1 v_{1,i,x} + m_2 v_{2,i,x}}{m}$$

$$v_{f,y} = \frac{m_1 v_{1,i,y} + m_2 v_{2,i,y}}{m}.$$

(Here, m represents the total mass of the system.) Finally, combine these components using the Pythagorean theorem,

Equation:

$$v_f = |\vec{\mathbf{v}}_f| = \sqrt{v_{f,x}^2 + v_{f,y}^2}.$$

Note:

Conservation of Momentum in Two Dimensions

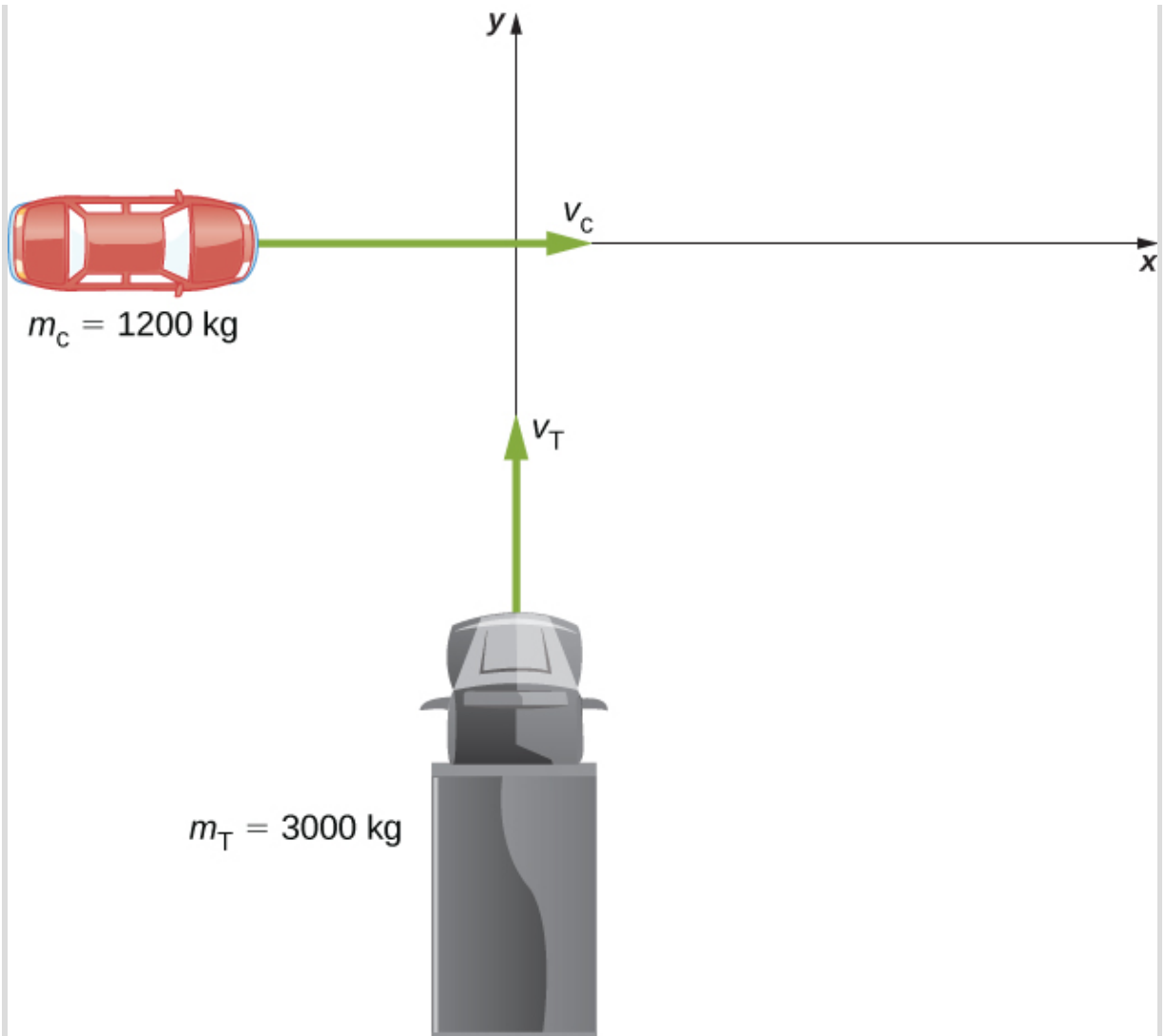
The method for solving a two-dimensional (or even three-dimensional) conservation of momentum problem is generally the same as the method for solving a one-dimensional problem, except that you have to conserve momentum in both (or all three) dimensions simultaneously:

1. Identify a closed system.
2. Write down the equation that represents conservation of momentum in the x -direction, and solve it for the desired quantity. If you are calculating a vector quantity (velocity, usually), this will give you the x -component of the vector.
3. Write down the equation that represents conservation of momentum in the y -direction, and solve. This will give you the y -component of your vector quantity.
4. Assuming you are calculating a vector quantity, use the Pythagorean theorem to calculate its magnitude, using the results of steps 3 and 4.

Example:

Traffic Collision

A small car of mass 1200 kg traveling east at 60 km/hr collides at an intersection with a truck of mass 3000 kg that is traveling due north at 40 km/hr ([link](#)). The two vehicles are locked together. What is the velocity of the combined wreckage?



A large truck moving north is about to collide with a small car moving east. The final momentum vector has both x - and y -components.

Strategy

First off, we need a closed system. The natural system to choose is the (car + truck), but this system is not closed; friction from the road acts on both vehicles. We avoid this problem by restricting the question to finding the velocity at the instant just after the collision, so that friction has not yet had any effect on the system. With that restriction, momentum is conserved for this system.

Since there are two directions involved, we do conservation of momentum twice: once in the x-direction and once in the y-direction.

Solution

Before the collision the total momentum is

Equation:

$$\vec{\mathbf{p}} = m_c \vec{\mathbf{v}}_c + m_T \vec{\mathbf{v}}_T.$$

After the collision, the wreckage has momentum

Equation:

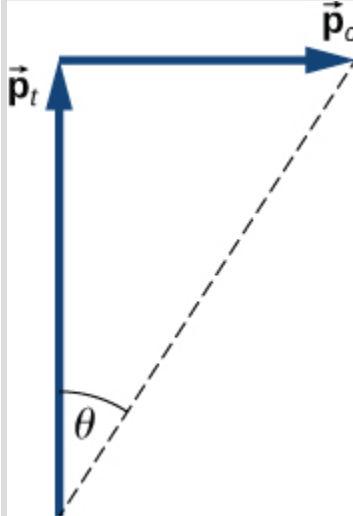
$$\vec{\mathbf{p}} = (m_c + m_T) \vec{\mathbf{v}}_w.$$

Since the system is closed, momentum must be conserved, so we have

Equation:

$$m_c \vec{\mathbf{v}}_c + m_T \vec{\mathbf{v}}_T = (m_c + m_T) \vec{\mathbf{v}}_w.$$

We have to be careful; the two initial momenta are not parallel. We must add vectorially ([\[link\]](#)).



Graphical
addition of
momentum
vectors. Notice
that, although

the car's
velocity is larger
than the truck's,
its momentum is
smaller.

If we define the $+x$ -direction to point east and the $+y$ -direction to point north, as in the figure, then (conveniently),

Equation:

$$\begin{aligned}\vec{\mathbf{p}}_{\text{c}} &= p_{\text{c}}\hat{\mathbf{i}} = m_{\text{c}}v_{\text{c}}\hat{\mathbf{i}} \\ \vec{\mathbf{p}}_{\text{T}} &= p_{\text{T}}\hat{\mathbf{j}} = m_{\text{T}}v_{\text{T}}\hat{\mathbf{j}}.\end{aligned}$$

Therefore, in the x -direction:

Equation:

$$\begin{aligned}m_{\text{c}}v_{\text{c}} &= (m_{\text{c}} + m_{\text{T}})v_{\text{w},x} \\ v_{\text{w},x} &= \left(\frac{m_{\text{c}}}{m_{\text{c}} + m_{\text{T}}}\right)v_{\text{c}}\end{aligned}$$

and in the y -direction:

Equation:

$$\begin{aligned}m_{\text{T}}v_{\text{T}} &= (m_{\text{c}} + m_{\text{T}})v_{\text{w},y} \\ v_{\text{w},y} &= \left(\frac{m_{\text{T}}}{m_{\text{c}} + m_{\text{T}}}\right)v_{\text{T}}.\end{aligned}$$

Applying the Pythagorean theorem gives

Equation:

$$\begin{aligned}
|\vec{v}_w| &= \sqrt{\left[\left(\frac{m_c}{m_c+m_t}\right)v_c\right]^2 + \left[\left(\frac{m_t}{m_c+m_t}\right)v_t\right]^2} \\
&= \sqrt{\left[\left(\frac{1200 \text{ kg}}{4200 \text{ kg}}\right)(16.67 \frac{\text{m}}{\text{s}})\right]^2 + \left[\left(\frac{3000 \text{ kg}}{4200 \text{ kg}}\right)(11.1 \frac{\text{m}}{\text{s}})\right]^2} \\
&= \sqrt{\left(4.76 \frac{\text{m}}{\text{s}}\right)^2 + \left(7.93 \frac{\text{m}}{\text{s}}\right)^2} \\
&= 9.25 \frac{\text{m}}{\text{s}} \approx 33.3 \frac{\text{km}}{\text{hr}}.
\end{aligned}$$

As for its direction, using the angle shown in the figure,

Equation:

$$\theta = \tan^{-1} \left(\frac{v_{w,x}}{v_{w,y}} \right) = \tan^{-1} \left(\frac{7.93 \text{ m/s}}{4.76 \text{ m/s}} \right) = 59^\circ.$$

This angle is east of north, or 31° counterclockwise from the +x-direction.

Significance

As a practical matter, accident investigators usually work in the “opposite direction”; they measure the distance of skid marks on the road (which gives the stopping distance) and use the work-energy theorem along with conservation of momentum to determine the speeds and directions of the cars prior to the collision. We saw that analysis in an earlier section.

Note:

Exercise:

Problem:

Check Your Understanding Suppose the initial velocities were *not* at right angles to each other. How would this change both the physical result and the mathematical analysis of the collision?

Solution:

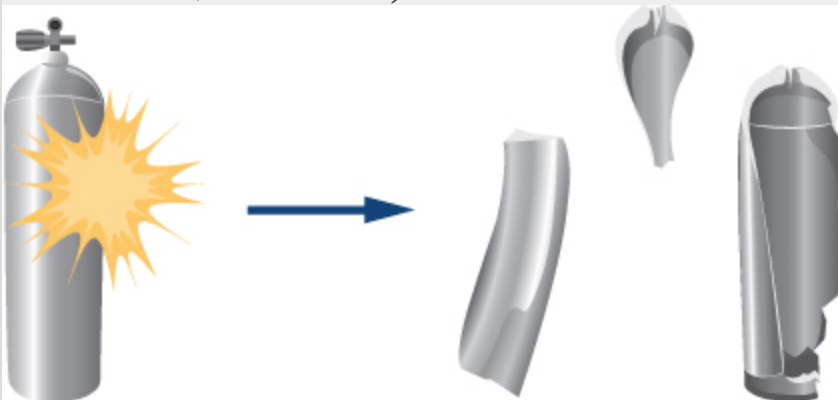
Were the initial velocities not at right angles, then one or both of the velocities would have to be expressed in component form. The

mathematical analysis of the problem would be slightly more involved, but the physical result would not change.

Example:

Exploding Scuba Tank

A common scuba tank is an aluminum cylinder that weighs 31.7 pounds empty ([\[link\]](#)). When full of compressed air, the internal pressure is between 2500 and 3000 psi (pounds per square inch). Suppose such a tank, which had been sitting motionless, suddenly explodes into three pieces. The first piece, weighing 10 pounds, shoots off horizontally at 235 miles per hour; the second piece (7 pounds) shoots off at 172 miles per hour, also in the horizontal plane, but at a 19° angle to the first piece. What is the mass and initial velocity of the third piece? (Do all work, and express your final answer, in SI units.)



A scuba tank explodes into three pieces.

Strategy

To use conservation of momentum, we need a closed system. If we define the system to be the scuba tank, this is not a closed system, since gravity is an external force. However, the problem asks for just the initial velocity of the third piece, so we can neglect the effect of gravity and consider the tank by itself as a closed system. Notice that, for this system, the initial momentum vector is zero.

We choose a coordinate system where all the motion happens in the xy -plane. We then write down the equations for conservation of momentum in each direction, thus obtaining the x - and y -components of the momentum of the third piece, from which we obtain its magnitude (via the Pythagorean theorem) and its direction. Finally, dividing this momentum by the mass of the third piece gives us the velocity.

Solution

First, let's get all the conversions to SI units out of the way:

Equation:

$$31.7 \text{ lb} \times \frac{1 \text{ kg}}{2.2 \text{ lb}} \rightarrow 14.4 \text{ kg}$$

$$10 \text{ lb} \rightarrow 4.5 \text{ kg}$$

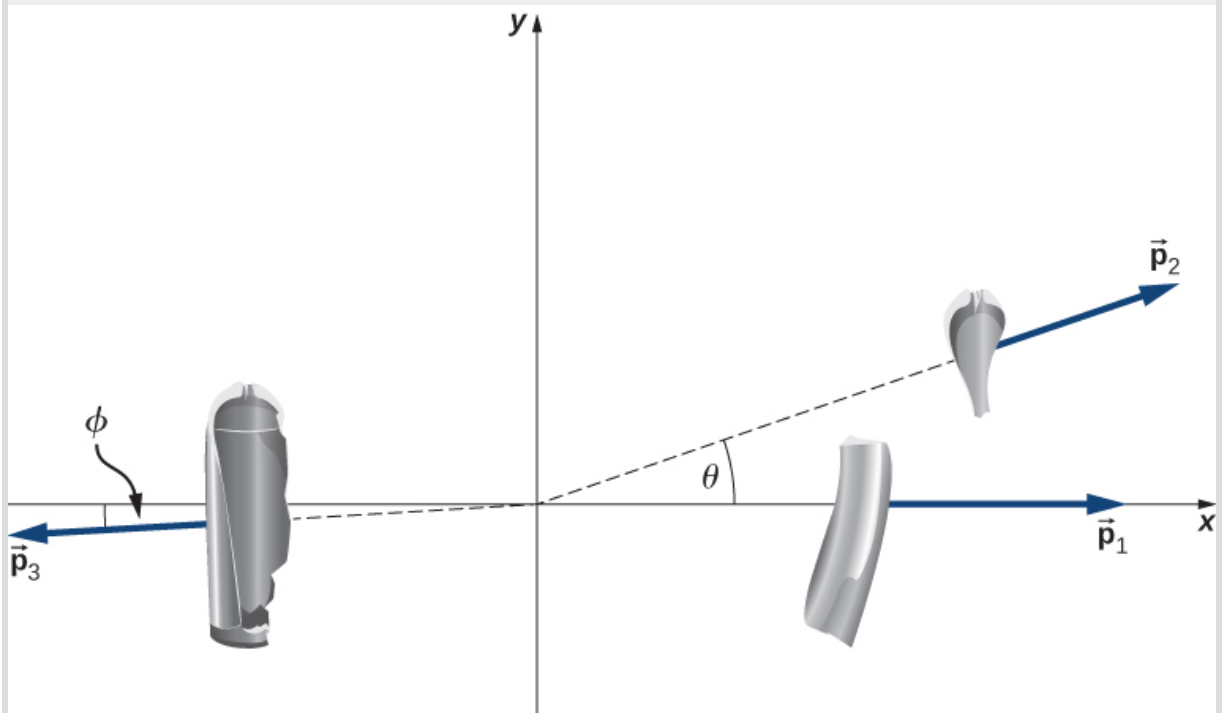
$$235 \frac{\text{miles}}{\text{hour}} \times \frac{1 \text{ hour}}{3600 \text{ s}} \times \frac{1609 \text{ m}}{\text{mile}} = 105 \frac{\text{m}}{\text{s}}$$

$$7 \text{ lb} \rightarrow 3.2 \text{ kg}$$

$$172 \frac{\text{mile}}{\text{hour}} = 77 \frac{\text{m}}{\text{s}}$$

$$m_3 = 14.4 \text{ kg} - (4.5 \text{ kg} + 3.2 \text{ kg}) = 6.7 \text{ kg}.$$

Now apply conservation of momentum in each direction.



x-direction:

Equation:

$$\begin{aligned}p_{f,x} &= p_{0,x} \\p_{1,x} + p_{2,x} + p_{3,x} &= 0 \\m_1 v_{1,x} + m_2 v_{2,x} + p_{3,x} &= 0 \\p_{3,x} &= -m_1 v_{1,x} - m_2 v_{2,x}\end{aligned}$$

y-direction:

Equation:

$$\begin{aligned}p_{f,y} &= p_{0,y} \\p_{1,y} + p_{2,y} + p_{3,y} &= 0 \\m_1 v_{1,y} + m_2 v_{2,y} + p_{3,y} &= 0 \\p_{3,y} &= -m_1 v_{1,y} - m_2 v_{2,y}\end{aligned}$$

From our chosen coordinate system, we write the x-components as

Equation:

$$\begin{aligned}p_{3,x} &= -m_1 v_1 - m_2 v_2 \cos \theta \\&= - (4.5 \text{ kg}) \left(105 \frac{\text{m}}{\text{s}} \right) - (3.2 \text{ kg}) \left(77 \frac{\text{m}}{\text{s}} \right) \cos (19^\circ) \\&= -705 \frac{\text{kg}\cdot\text{m}}{\text{s}}.\end{aligned}$$

For the y-direction, we have

Equation:

$$\begin{aligned}p_{3y} &= 0 - m_2 v_2 \sin \theta \\&= - (3.2 \text{ kg}) \left(77 \frac{\text{m}}{\text{s}} \right) \sin (19^\circ) \\&= -80.2 \frac{\text{kg}\cdot\text{m}}{\text{s}}.\end{aligned}$$

This gives the magnitude of p_3 :

Equation:

$$\begin{aligned}
 p_3 &= \sqrt{p_{3,x}^2 + p_{3,y}^2} \\
 &= \sqrt{\left(-705 \frac{\text{kg}\cdot\text{m}}{\text{s}}\right)^2 + \left(-80.2 \frac{\text{kg}\cdot\text{m}}{\text{s}}\right)^2} \\
 &= 710 \frac{\text{kg}\cdot\text{m}}{\text{s}}.
 \end{aligned}$$

The velocity of the third piece is therefore

Equation:

$$v_3 = \frac{p_3}{m_3} = \frac{710 \frac{\text{kg}\cdot\text{m}}{\text{s}}}{6.7 \text{ kg}} = 106 \frac{\text{m}}{\text{s}}.$$

The direction of its velocity vector is the same as the direction of its momentum vector:

Equation:

$$\phi = \tan^{-1} \left(\frac{p_{3,y}}{p_{3,x}} \right) = \tan^{-1} \left(\frac{80.2 \frac{\text{kg}\cdot\text{m}}{\text{s}}}{705 \frac{\text{kg}\cdot\text{m}}{\text{s}}} \right) = 6.49^\circ.$$

Because ϕ is below the $-x$ -axis, the actual angle is 183.5° from the $+x$ -direction.

Significance

The enormous velocities here are typical; an exploding tank of any compressed gas can easily punch through the wall of a house and cause significant injury, or death. Fortunately, such explosions are extremely rare, on a percentage basis.

Note:

Exercise:

Problem:

Check Your Understanding Notice that the mass of the air in the tank was neglected in the analysis and solution. How would the solution method change if the air was included? How large a difference do you think it would make in the final answer?

Solution:

The volume of a scuba tank is about 11 L. Assuming air is an ideal gas, the number of gas molecules in the tank is

$$PV = NRT$$

$$N = \frac{PV}{RT} = \frac{(2500 \text{ psi})(0.011 \text{ m}^3)}{(8.31 \text{ J/mol}\cdot\text{K})(300 \text{ K})} \left(\frac{6894.8 \text{ Pa}}{1 \text{ psi}} \right)$$
$$= 7.59 \times 10^1 \text{ mol}$$

The average molecular mass of air is 29 g/mol, so the mass of air contained in the tank is about 2.2 kg. This is about 10 times less than the mass of the tank, so it is safe to neglect it. Also, the initial force of the air pressure is roughly proportional to the surface area of each piece, which is in turn proportional to the mass of each piece (assuming uniform thickness). Thus, the initial acceleration of each piece would change very little if we explicitly consider the air.

Summary

- The approach to two-dimensional collisions is to choose a convenient coordinate system and break the motion into components along perpendicular axes.
- Momentum is conserved in both directions simultaneously and independently.
- The Pythagorean theorem gives the magnitude of the momentum vector using the x- and y-components, calculated using conservation of momentum in each direction.

Conceptual Questions

Exercise:

Problem:

Momentum for a system can be conserved in one direction while not being conserved in another. What is the angle between the directions? Give an example.

Solution:

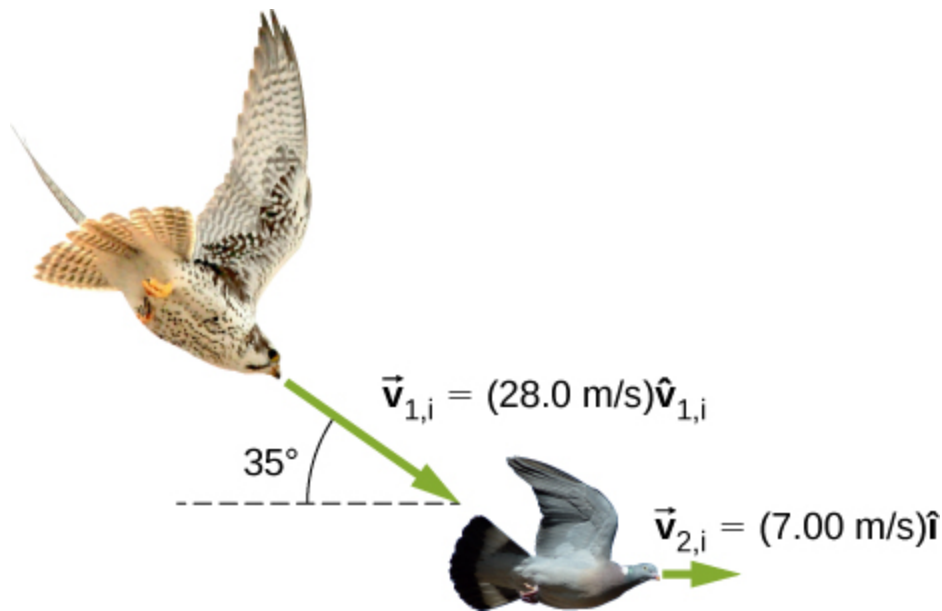
The angle between the directions must be 90° . Any system that has zero net external force in one direction and nonzero net external force in a perpendicular direction will satisfy these conditions.

Problems

Exercise:

Problem:

A 0.90-kg falcon is diving at 28.0 m/s at a downward angle of 35° . It catches a 0.325-kg pigeon from behind in midair. What is their combined velocity after impact if the pigeon's initial velocity was 7.00 m/s directed horizontally? Note that $\hat{\mathbf{v}}_{1,i}$ is a unit vector pointing in the direction in which the falcon is initially flying.



(credit “falcon”: modification of work by “USFWS Mountain-Prairie”/Flickr; credit “pigeon”: modification of work by Jacob Spinks)

Solution:

22.1 m/s at 32.2° below the horizontal

Exercise:

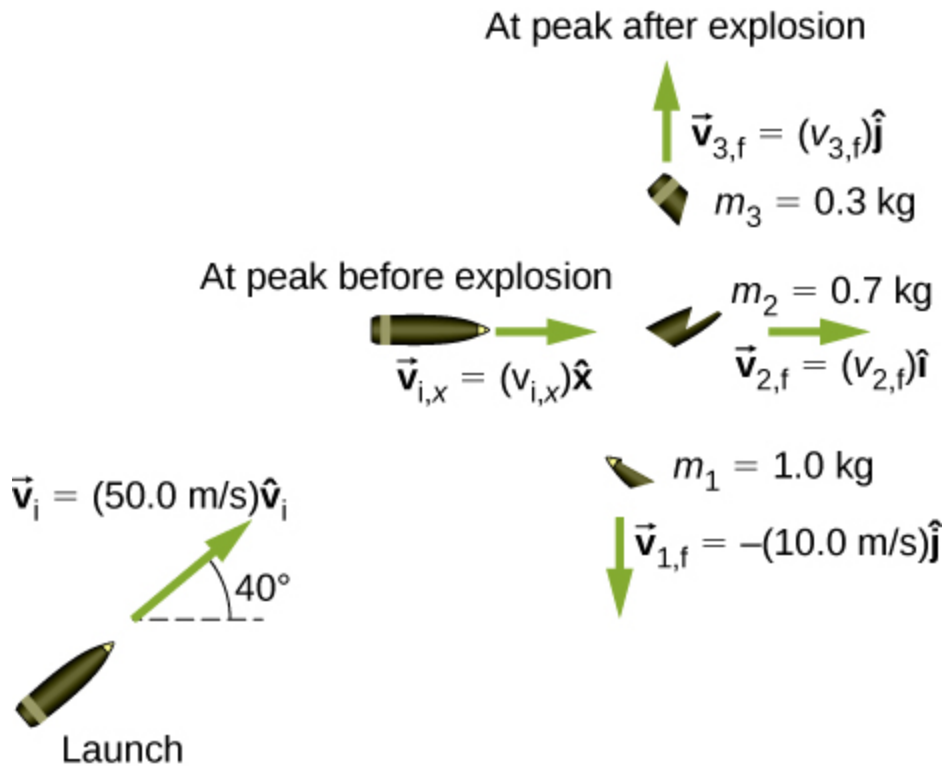
Problem:

A billiard ball, labeled 1, moving horizontally strikes another billiard ball, labeled 2, at rest. Before impact, ball 1 was moving at a speed of 3.00 m/s, and after impact it is moving at 0.50 m/s at 50° from the original direction. If the two balls have equal masses of 300 g, what is the velocity of the ball 2 after the impact?

Exercise:

Problem:

A projectile of mass 2.0 kg is fired in the air at an angle of 40.0° to the horizon at a speed of 50.0 m/s. At the highest point in its flight, the projectile breaks into three parts of mass 1.0 kg, 0.7 kg, and 0.3 kg. The 1.0-kg part falls straight down after breakup with an initial speed of 10.0 m/s, the 0.7-kg part moves in the original forward direction, and the 0.3-kg part goes straight up.



- Find the speeds of the 0.3-kg and 0.7-kg pieces immediately after the break-up.
- How high from the break-up point does the 0.3-kg piece go before coming to rest?
- Where does the 0.7-kg piece land relative to where it was fired from?

Solution:

a. 33 m/s and 110 m/s; b. 57 m; c. 480 m

Exercise:**Problem:**

Two asteroids collide and stick together. The first asteroid has mass of 15×10^3 kg and is initially moving at 770 m/s. The second asteroid has mass of 20×10^3 kg and is moving at 1020 m/s. Their initial velocities made an angle of 20° with respect to each other. What is the final speed and direction with respect to the velocity of the first asteroid?

Exercise:**Problem:**

A 200-kg rocket in deep space moves with a velocity of $(121 \text{ m/s})\hat{\mathbf{i}} + (38.0 \text{ m/s})\hat{\mathbf{j}}$. Suddenly, it explodes into three pieces, with the first (78 kg) moving at $-(321 \text{ m/s})\hat{\mathbf{i}} + (228 \text{ m/s})\hat{\mathbf{j}}$ and the second (56 kg) moving at $(16.0 \text{ m/s})\hat{\mathbf{i}} - (88.0 \text{ m/s})\hat{\mathbf{j}}$. Find the velocity of the third piece.

Solution:

$$(732 \text{ m/s})\hat{\mathbf{i}} + (-79.6 \text{ m/s})\hat{\mathbf{j}}$$

Exercise:**Problem:**

A proton traveling at 3.0×10^6 m/s scatters elastically from an initially stationary alpha particle and is deflected at an angle of 85° with respect to its initial velocity. Given that the alpha particle has four times the mass of the proton, what percent of its initial kinetic energy does the proton retain after the collision?

Exercise:

Problem:

Three 70-kg deer are standing on a flat 200-kg rock that is on an ice-covered pond. A gunshot goes off and the deer scatter, with deer A running at $(15 \text{ m/s})\hat{\mathbf{i}} + (5.0 \text{ m/s})\hat{\mathbf{j}}$, deer B running at $(-12 \text{ m/s})\hat{\mathbf{i}} + (8.0 \text{ m/s})\hat{\mathbf{j}}$, and deer C running at $(1.2 \text{ m/s})\hat{\mathbf{i}} - (18.0 \text{ m/s})\hat{\mathbf{j}}$. What is the velocity of the rock on which they were standing?

Solution:

$$-(0.21 \text{ m/s})\hat{\mathbf{i}} + (0.25 \text{ m/s})\hat{\mathbf{j}}$$

Exercise:**Problem:**

A family is skating. The father (75 kg) skates at 8.2 m/s and collides and sticks to the mother (50 kg), who was initially moving at 3.3 m/s and at 45° with respect to the father's velocity. The pair then collides with their daughter (30 kg), who was stationary, and the three slide off together. What is their final velocity?

Exercise:**Problem:**

An oxygen atom (mass 16 u) moving at 733 m/s at 15.0° with respect to the $\hat{\mathbf{i}}$ direction collides and sticks to an oxygen molecule (mass 32 u) moving at 528 m/s at 128° with respect to the $\hat{\mathbf{i}}$ direction. The two stick together to form ozone. What is the final velocity of the ozone molecule?

Solution:

341 m/s at 86.8° with respect to the $\hat{\mathbf{i}}$ axis.

Exercise:

Problem:

Two cars of the same mass approach an extremely icy four-way perpendicular intersection. Car A travels northward at 30 m/s and car B is travelling eastward. They collide and stick together, traveling at 28° north of east. What was the initial velocity of car B?

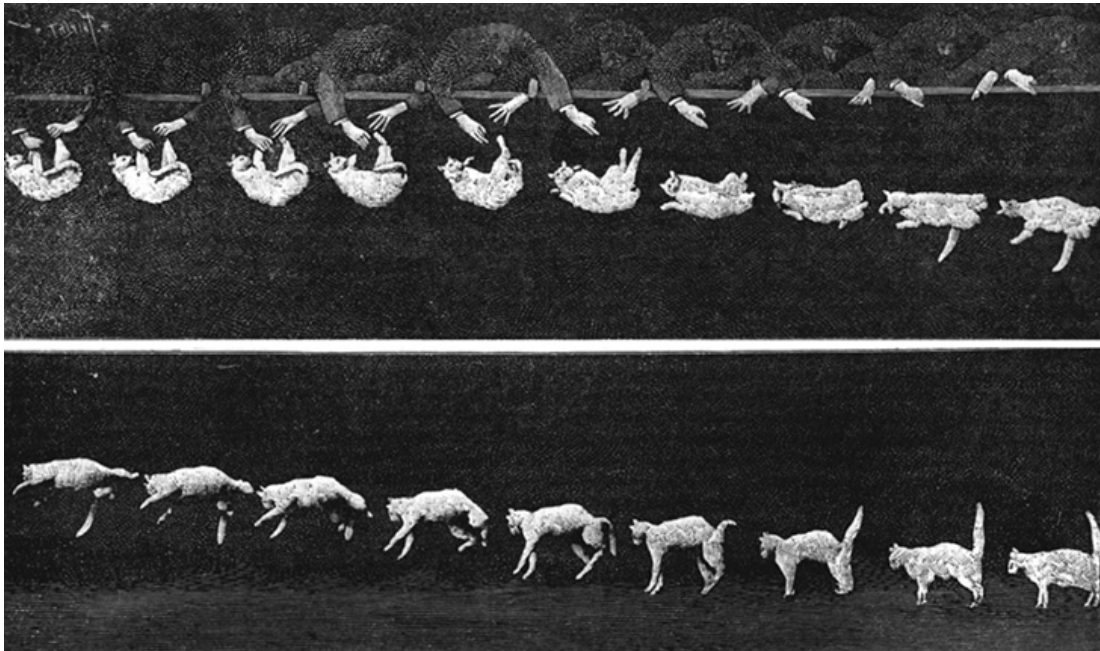
Center of Mass

By the end of this section, you will be able to:

- Explain the meaning and usefulness of the concept of center of mass
- Calculate the center of mass of a given system
- Apply the center of mass concept in two and three dimensions
- Calculate the velocity and acceleration of the center of mass

We have been avoiding an important issue up to now: When we say that an object moves (more correctly, accelerates) in a way that obeys Newton's second law, we have been ignoring the fact that all objects are actually made of many constituent particles. A car has an engine, steering wheel, seats, passengers; a football is leather and rubber surrounding air; a brick is made of atoms. There are many different types of particles, and they are generally not distributed uniformly in the object. How do we include these facts into our calculations?

Then too, an extended object might change shape as it moves, such as a water balloon or a cat falling ([link](#)). This implies that the constituent particles are applying internal forces on each other, in addition to the external force that is acting on the object as a whole. We want to be able to handle this, as well.



As the cat falls, its body performs complicated motions so it can land on its feet, but one point in the system moves with the simple uniform acceleration of gravity.

The problem before us, then, is to determine what part of an extended object is obeying Newton's second law when an external force is applied and to determine how the motion of the object as a whole is affected by both the internal and external forces.

Be warned: To treat this new situation correctly, we must be rigorous and completely general. We won't make any assumptions about the nature of the object, or of its constituent particles, or either the internal or external forces. Thus, the arguments will be complex.

Internal and External Forces

Suppose we have an extended object of mass M , made of N interacting particles. Let's label their masses as m_j , where $j = 1, 2, 3, \dots, N$. Note that

Equation:

$$M = \sum_{j=1}^N m_j.$$

If we apply some net **external force** \vec{F}_{ext} on the object, every particle experiences some “share” or some fraction of that external force. Let:

Equation:

$$\vec{f}_j^{\text{ext}} = \text{the fraction of the external force that the } j\text{th particle experiences.}$$

Notice that these fractions of the total force are not necessarily equal; indeed, they virtually never are. (They *can* be, but they usually aren't.) In general, therefore,

Equation:

$$\vec{f}_1^{\text{ext}} \neq \vec{f}_2^{\text{ext}} \neq \dots \neq \vec{f}_N^{\text{ext}}.$$

Next, we assume that each of the particles making up our object can interact (apply forces on) every other particle of the object. We won't try to guess what kind of forces they are; but since these forces are the result of particles of the object acting on other particles of the same object, we refer to them as **internal forces** \vec{f}_j^{int} ; thus:

\vec{f}_j^{int} = the net internal force that the j th particle experiences from all the other particles that make up the object.

Now, the *net* force, internal plus external, on the j th particle is the vector sum of these:

Equation:

$$\vec{f}_j = \vec{f}_j^{\text{int}} + \vec{f}_j^{\text{ext}}.$$

where again, this is for all N particles; $j = 1, 2, 3, \dots, N$.

As a result of this fractional force, the momentum of each particle gets changed:

Equation:

$$\begin{aligned}\vec{\mathbf{f}}_j &= \frac{d\vec{\mathbf{p}}_j}{dt} \\ \vec{\mathbf{f}}_j^{\text{int}} + \vec{\mathbf{f}}_j^{\text{ext}} &= \frac{d\vec{\mathbf{p}}_j}{dt}.\end{aligned}$$

The net force $\vec{\mathbf{F}}$ on the *object* is the vector sum of these forces:

Equation:

$$\begin{aligned}\vec{\mathbf{F}}_{\text{net}} &= \sum_{j=1}^N \left(\vec{\mathbf{f}}_j^{\text{int}} + \vec{\mathbf{f}}_j^{\text{ext}} \right) \\ &= \sum_{j=1}^N \vec{\mathbf{f}}_j^{\text{int}} + \sum_{j=1}^N \vec{\mathbf{f}}_j^{\text{ext}}.\end{aligned}$$

This net force changes the momentum of the object as a whole, and the net change of momentum of the object must be the vector sum of all the individual changes of momentum of all of the particles:

Equation:

$$\vec{\mathbf{F}}_{\text{net}} = \sum_{j=1}^N \frac{d\vec{\mathbf{p}}_j}{dt}.$$

Combining [\[link\]](#) and [\[link\]](#) gives

Equation:

$$\sum_{j=1}^N \vec{\mathbf{f}}_j^{\text{int}} + \sum_{j=1}^N \vec{\mathbf{f}}_j^{\text{ext}} = \sum_{j=1}^N \frac{d\vec{\mathbf{p}}_j}{dt}.$$

Let's now think about these summations. First consider the internal forces term; remember that each $\vec{\mathbf{f}}_j^{\text{int}}$ is the force on the j th particle from the other particles in the object. But by Newton's third law, for every one of these forces, there must be another force that has the same magnitude, but the opposite sign (points in the opposite direction). These forces do not cancel; however, that's not what we're doing in the summation. Rather, we're simply *mathematically adding up* all the internal force vectors. That is, in general, the internal forces for any individual part of the object won't cancel, but when all the internal forces are added up, the internal forces must cancel in pairs. It follows, therefore, that the sum of all the internal forces must be zero:

Equation:

$$\sum_{j=1}^N \vec{\mathbf{f}}_j^{\text{int}} = 0.$$

(This argument is subtle, but crucial; take plenty of time to completely understand it.)

For the external forces, this summation is simply the total external force that was applied to the whole object:

Equation:

$$\sum_{j=1}^N \vec{\mathbf{f}}_j^{\text{ext}} = \vec{\mathbf{F}}_{\text{ext}}.$$

As a result,

Note:

Equation:

$$\vec{\mathbf{F}}_{\text{ext}} = \sum_{j=1}^N \frac{d\vec{\mathbf{p}}_j}{dt}.$$

This is an important result. [\[link\]](#) tells us that the total change of momentum of the entire object (all N particles) is due only to the external forces; the internal forces do not change the momentum of the object as a whole. This is why you can't lift yourself in the air by standing in a basket and pulling up on the handles: For the system of you + basket, your upward pulling force is an internal force.

Force and Momentum

Remember that our actual goal is to determine the equation of motion for the entire object (the entire system of particles). To that end, let's define:

$\vec{\mathbf{p}}_{\text{CM}}$ = the total momentum of the system of N particles (the reason for the subscript will become clear shortly)

Then we have

Equation:

$$\vec{\mathbf{p}}_{\text{CM}} \equiv \sum_{j=1}^N \vec{\mathbf{p}}_j,$$

and therefore [\[link\]](#) can be written simply as

Note:

Equation:

$$\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}_{\text{CM}}}{dt}.$$

Since this change of momentum is caused by only the net external force, we have dropped the “ext” subscript.

This is Newton’s second law, but now for the entire extended object. If this feels a bit anticlimactic, remember what is hiding inside it: $\vec{\mathbf{p}}_{\text{CM}}$ is the vector sum of the momentum of (in principle) hundreds of thousands of billions of billions of particles (6.02×10^{23}), all caused by one simple net external force—a force that you can calculate.

Center of Mass

Our next task is to determine what part of the extended object, if any, is obeying [\[link\]](#).

It’s tempting to take the next step; does the following equation mean anything?

Equation:

$$\vec{\mathbf{F}} = M\vec{\mathbf{a}}$$

If it *does* mean something (acceleration of what, exactly?), then we could write

Equation:

$$M\vec{\mathbf{a}} = \frac{d\vec{\mathbf{p}}_{\text{CM}}}{dt}$$

and thus

Equation:

$$M\vec{\mathbf{a}} = \sum_{j=1}^N \frac{d\vec{\mathbf{p}}_j}{dt} = \frac{d}{dt} \sum_{j=1}^N \vec{\mathbf{p}}_j.$$

which follows because the derivative of a sum is equal to the sum of the derivatives.

Now, $\vec{\mathbf{p}}_j$ is the momentum of the j th particle. Defining the positions of the constituent particles (relative to some coordinate system) as $\vec{\mathbf{r}}_j = (x_j, y_j, z_j)$, we thus have

Equation:

$$\vec{\mathbf{p}}_j = m_j \vec{\mathbf{v}}_j = m_j \frac{d\vec{\mathbf{r}}_j}{dt}.$$

Substituting back, we obtain

Equation:

$$\begin{aligned}
 M\vec{\mathbf{a}} &= \frac{d}{dt} \sum_{j=1}^N m_j \frac{d\vec{\mathbf{r}}_j}{dt} \\
 &= \frac{d^2}{dt^2} \sum_{j=1}^N m_j \vec{\mathbf{r}}_j.
 \end{aligned}$$

Dividing both sides by M (the total mass of the extended object) gives us

Note:
Equation:

$$\vec{\mathbf{a}} = \frac{d^2}{dt^2} \left(\frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{r}}_j \right).$$

Thus, the point in the object that traces out the trajectory dictated by the applied force in [\[link\]](#) is inside the parentheses in [\[link\]](#).

Looking at this calculation, notice that (inside the parentheses) we are calculating the product of each particle's mass with its position, adding all N of these up, and dividing this sum by the total mass of particles we summed. This is reminiscent of an average; inspired by this, we'll (loosely) interpret it to be the weighted average position of the mass of the extended object. It's actually called the **center of mass** of the object. Notice that the position of the center of mass has units of meters; that suggests a definition:

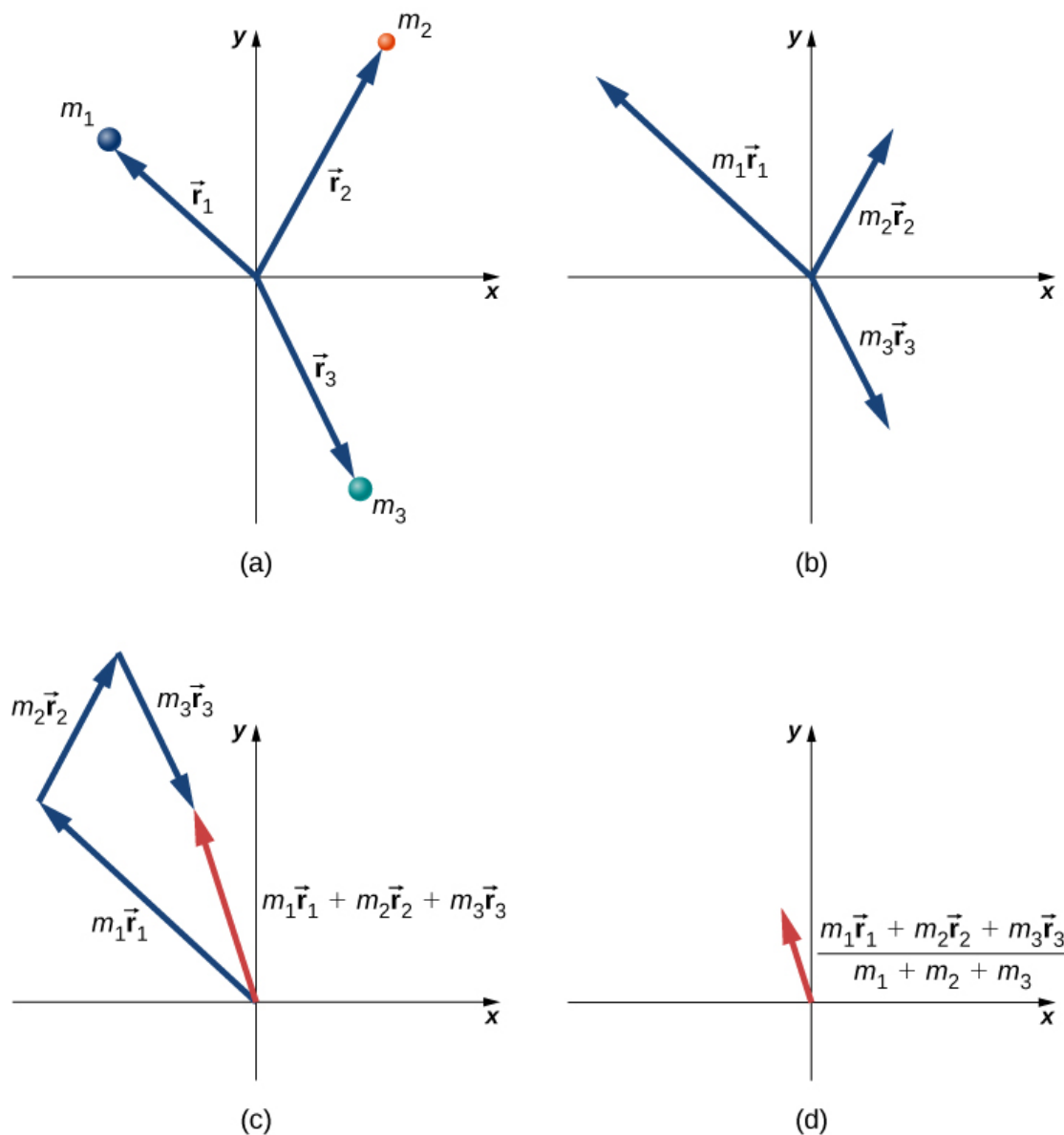
Note:
Equation:

$$\vec{\mathbf{r}}_{\text{CM}} \equiv \frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{r}}_j.$$

So, the point that obeys [\[link\]](#) (and therefore [\[link\]](#) as well) is the center of mass of the object, which is located at the position vector $\vec{\mathbf{r}}_{\text{CM}}$.

It may surprise you to learn that there does not have to be any actual mass at the center of mass of an object. For example, a hollow steel sphere with a vacuum inside it is spherically symmetrical (meaning its mass is uniformly distributed about the center of the sphere); all of the sphere's mass is out on its surface, with no mass inside. But it can be shown that the center of mass of the sphere is at its geometric center, which seems reasonable. Thus, there is no mass at the position of the center of

mass of the sphere. (Another example is a doughnut.) The procedure to find the center of mass is illustrated in [\[link\]](#).



Finding the center of mass of a system of three different particles. (a) Position vectors are created for each object. (b) The position vectors are multiplied by the mass of the corresponding object. (c) The scaled vectors from part (b) are added together. (d) The final vector is divided by the total mass. This vector points to the center of mass of the system. Note that no mass is actually present at the center of mass of this system.

Since $\vec{r}_j = x_j\hat{i} + y_j\hat{j} + z_j\hat{k}$, it follows that:

Equation:

$$r_{\text{CM},x} = \frac{1}{M} \sum_{j=1}^N m_j x_j$$

Equation:

$$r_{\text{CM},y} = \frac{1}{M} \sum_{j=1}^N m_j y_j$$

Equation:

$$r_{\text{CM},z} = \frac{1}{M} \sum_{j=1}^N m_j z_j$$

and thus

Equation:

$$\begin{aligned} \vec{\mathbf{r}}_{\text{CM}} &= r_{\text{CM},x} \hat{\mathbf{i}} + r_{\text{CM},y} \hat{\mathbf{j}} + r_{\text{CM},z} \hat{\mathbf{k}} \\ r_{\text{CM}} &= |\vec{\mathbf{r}}_{\text{CM}}| = \left(r_{\text{CM},x}^2 + r_{\text{CM},y}^2 + r_{\text{CM},z}^2 \right)^{1/2}. \end{aligned}$$

Therefore, you can calculate the components of the center of mass vector individually.

Finally, to complete the kinematics, the instantaneous velocity of the center of mass is calculated exactly as you might suspect:

Note:

Equation:

$$\vec{\mathbf{v}}_{\text{CM}} = \frac{d}{dt} \left(\frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{r}}_j \right) = \frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{v}}_j$$

and this, like the position, has x -, y -, and z -components.

To calculate the center of mass in actual situations, we recommend the following procedure:

Note:

Calculating the Center of Mass

The center of mass of an object is a position vector. Thus, to calculate it, do these steps:

1. Define your coordinate system. Typically, the origin is placed at the location of one of the particles. This is not required, however.
2. Determine the x , y , z -coordinates of each particle that makes up the object.
3. Determine the mass of each particle, and sum them to obtain the total mass of the object. Note that the mass of the object at the origin *must* be included in the total mass.
4. Calculate the x -, y -, and z -components of the center of mass vector, using [\[link\]](#), [\[link\]](#), and [\[link\]](#).
5. If required, use the Pythagorean theorem to determine its magnitude.

Here are two examples that will give you a feel for what the center of mass is.

Example:

Center of Mass of the Earth-Moon System

Using data from text appendix, determine how far the center of mass of the Earth-moon system is from the center of Earth. Compare this distance to the radius of Earth, and comment on the result. Ignore the other objects in the solar system.

Strategy

We get the masses and separation distance of the Earth and moon, impose a coordinate system, and use [\[link\]](#) with just $N = 2$ objects. We use a subscript “e” to refer to Earth, and subscript “m” to refer to the moon.

Solution

Define the origin of the coordinate system as the center of Earth. Then, with just two objects, [\[link\]](#) becomes

Equation:

$$R = \frac{m_e r_e + m_m r_m}{m_e + m_m}.$$

From [Appendix D](#),

Equation:

$$m_e = 5.97 \times 10^{24} \text{ kg}$$

Equation:

$$m_m = 7.36 \times 10^{22} \text{ kg}$$

Equation:

$$r_m = 3.82 \times 10^8 \text{ m}.$$

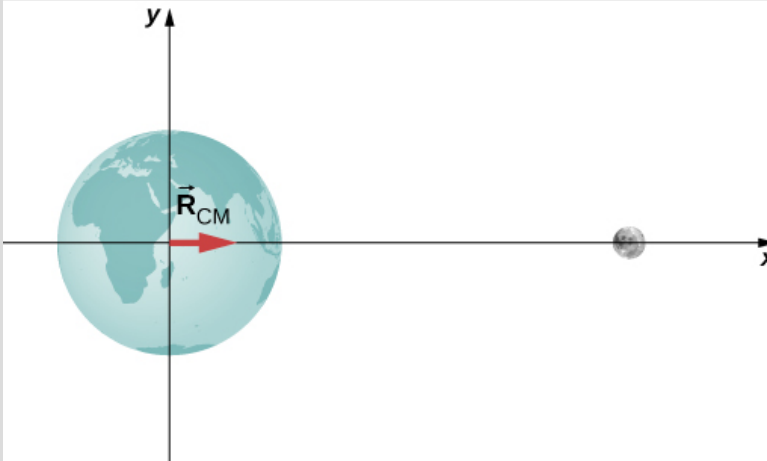
We defined the center of Earth as the origin, so $r_e = 0$ m. Inserting these into the equation for R gives

Equation:

$$\begin{aligned}
 R &= \frac{(5.97 \times 10^{24} \text{ kg})(0 \text{ m}) + (7.36 \times 10^{22} \text{ kg})(3.82 \times 10^8 \text{ m})}{5.97 \times 10^{24} \text{ kg} + 7.36 \times 10^{22} \text{ kg}} \\
 &= 4.64 \times 10^6 \text{ m}.
 \end{aligned}$$

Significance

The radius of Earth is $6.37 \times 10^6 \text{ m}$, so the center of mass of the Earth-moon system is $(6.37 - 4.64) \times 10^6 \text{ m} = 1.73 \times 10^6 \text{ m} = 1730 \text{ km}$ (roughly 1080 miles) *below* the surface of Earth. The location of the center of mass is shown (not to scale).

**Note:****Exercise:****Problem:**

Check Your Understanding Suppose we included the sun in the system. Approximately where would the center of mass of the Earth-moon-sun system be located? (Feel free to actually calculate it.)

Solution:

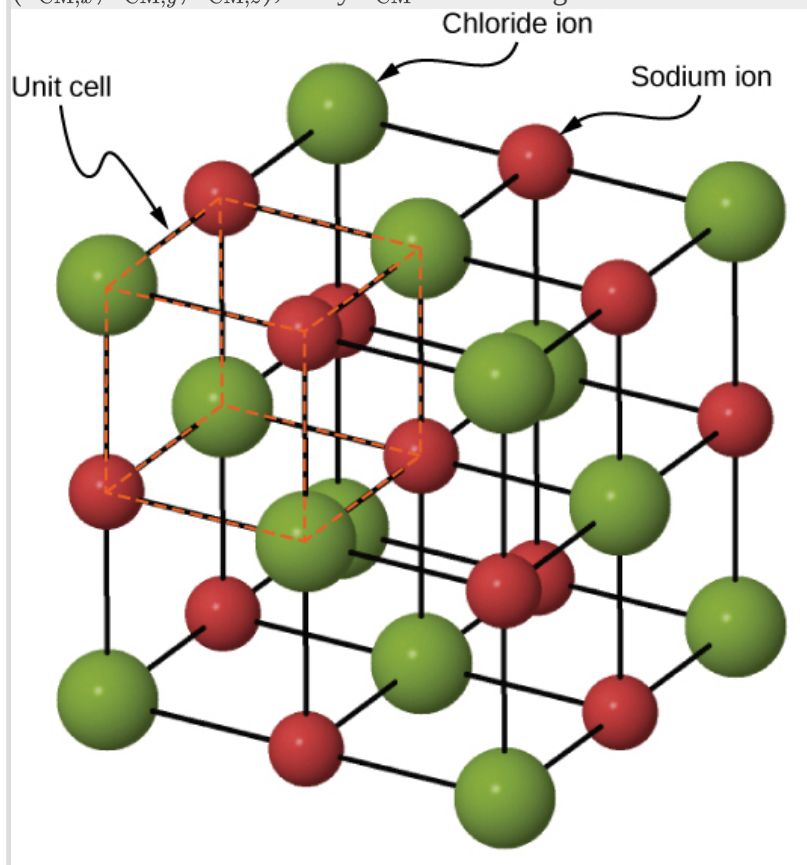
The average radius of Earth's orbit around the Sun is $1.496 \times 10^9 \text{ m}$. Taking the Sun to be the origin, and noting that the mass of the Sun is approximately the same as the masses of the Sun, Earth, and Moon combined, the center of mass of the Earth + Moon system and the Sun is

$$\begin{aligned}
 R_{\text{CM}} &= \frac{m_{\text{Sun}} R_{\text{Sun}} + m_{\text{em}} R_{\text{em}}}{m_{\text{Sun}}} \\
 &= \frac{(1.989 \times 10^{30} \text{ kg})(0) + (5.97 \times 10^{24} \text{ kg} + 7.36 \times 10^{22} \text{ kg})(1.496 \times 10^9 \text{ m})}{1.989 \times 10^{30} \text{ kg}} \\
 &= 4.6 \text{ km}
 \end{aligned}$$

Thus, the center of mass of the Sun, Earth, Moon system is 4.6 km from the center of the Sun.

Example:**Center of Mass of a Salt Crystal**

[\[link\]](#) shows a single crystal of sodium chloride—ordinary table salt. The sodium and chloride ions form a single unit, NaCl. When multiple NaCl units group together, they form a cubic lattice. The smallest possible cube (called the *unit cell*) consists of four sodium ions and four chloride ions, alternating. The length of one edge of this cube (i.e., the bond length) is 2.36×10^{-10} m. Find the location of the center of mass of the unit cell. Specify it either by its coordinates ($r_{CM,x}$, $r_{CM,y}$, $r_{CM,z}$), or by r_{CM} and two angles.



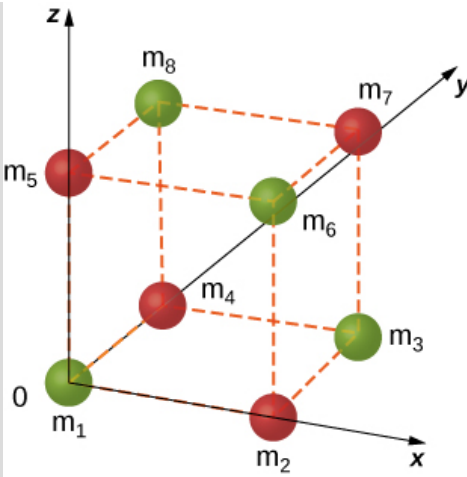
A drawing of a sodium chloride (NaCl) crystal.

Strategy

We can look up all the ion masses. If we impose a coordinate system on the unit cell, this will give us the positions of the ions. We can then apply [\[link\]](#), [\[link\]](#), and [\[link\]](#) (along with the Pythagorean theorem).

Solution

Define the origin to be at the location of the chloride ion at the bottom left of the unit cell. [\[link\]](#) shows the coordinate system.



A single unit cell of a NaCl crystal.

There are eight ions in this crystal, so $N = 8$:

Equation:

$$\vec{r}_{\text{CM}} = \frac{1}{M} \sum_{j=1}^8 m_j \vec{r}_j.$$

The mass of each of the chloride ions is

Equation:

$$35.453\text{u} \times \frac{1.660 \times 10^{-27} \text{ kg}}{\text{u}} = 5.885 \times 10^{-26} \text{ kg}$$

so we have

Equation:

$$m_1 = m_3 = m_6 = m_8 = 5.885 \times 10^{-26} \text{ kg}.$$

For the sodium ions,

Equation:

$$m_2 = m_4 = m_5 = m_7 = 3.816 \times 10^{-26} \text{ kg}.$$

The total mass of the unit cell is therefore

Equation:

$$M = (4) (5.885 \times 10^{-26} \text{ kg}) + (4) (3.816 \times 10^{-26} \text{ kg}) = 3.880 \times 10^{-25} \text{ kg}.$$

From the geometry, the locations are

Equation:

$$\vec{r}_1 = 0$$

$$\vec{r}_2 = (2.36 \times 10^{-10} \text{ m})\hat{i}$$

$$\vec{r}_3 = r_{3x}\hat{i} + r_{3y}\hat{j} = (2.36 \times 10^{-10} \text{ m})\hat{i} + (2.36 \times 10^{-10} \text{ m})\hat{j}$$

$$\vec{r}_4 = (2.36 \times 10^{-10} \text{ m})\hat{j}$$

$$\vec{r}_5 = (2.36 \times 10^{-10} \text{ m})\vec{k}$$

$$\vec{r}_6 = r_{6x}\hat{i} + r_{6z}\hat{k} = (2.36 \times 10^{-10} \text{ m})\hat{i} + (2.36 \times 10^{-10} \text{ m})\hat{k}$$

$$\vec{r}_7 = r_{7x}\hat{i} + r_{7y}\hat{j} + r_{7z}\hat{k} = (2.36 \times 10^{-10} \text{ m})\hat{i} + (2.36 \times 10^{-10} \text{ m})\hat{j} + (2.36 \times 10^{-10} \text{ m})\hat{k}$$

$$\vec{r}_8 = r_{8y}\hat{j} + r_{8z}\hat{k} = (2.36 \times 10^{-10} \text{ m})\hat{j} + (2.36 \times 10^{-10} \text{ m})\hat{k}.$$

Substituting:

Equation:

$$\begin{aligned} |\vec{r}_{\text{CM},x}| &= \sqrt{r_{\text{CM},x}^2 + r_{\text{CM},y}^2 + r_{\text{CM},z}^2} \\ &= \frac{1}{M} \sum_{j=1}^8 m_j (r_x)_j \\ &= \frac{1}{M} (m_1 r_{1x} + m_2 r_{2x} + m_3 r_{3x} + m_4 r_{4x} + m_5 r_{5x} + m_6 r_{6x} + m_7 r_{7x} + m_8 r_{8x}) \\ &= \frac{1}{3.8804 \times 10^{-25} \text{ kg}} [(5.885 \times 10^{-26} \text{ kg})(0 \text{ m}) + (3.816 \times 10^{-26} \text{ kg})(2.36 \times 10^{-10} \text{ m}) \\ &\quad + (5.885 \times 10^{-26} \text{ kg})(2.36 \times 10^{-10} \text{ m}) \\ &\quad + (3.816 \times 10^{-26} \text{ kg})(2.36 \times 10^{-10} \text{ m}) + 0 + 0 \\ &\quad + (3.816 \times 10^{-26} \text{ kg})(2.36 \times 10^{-10} \text{ m}) + 0] \\ &= 1.18 \times 10^{-10} \text{ m}. \end{aligned}$$

Similar calculations give $r_{\text{CM},y} = r_{\text{CM},z} = 1.18 \times 10^{-10} \text{ m}$ (you could argue that this must be true, by symmetry, but it's a good idea to check).

Significance

Although this is a great exercise to determine the center of mass given a Chloride ion at the origin, in fact the origin could be chosen at any location. Therefore, there is no meaningful application of the center of mass of a unit cell beyond as an exercise.

Note:

Exercise:

Problem:

Check Your Understanding Suppose you have a macroscopic salt crystal (that is, a crystal that is large enough to be visible with your unaided eye). It is made up of a *huge* number of unit cells. Is the center of mass of this crystal necessarily at the geometric center of the crystal?

Solution:

On a macroscopic scale, the size of a unit cell is negligible and the crystal mass may be considered to be distributed homogeneously throughout the crystal. Thus,

$$\vec{r}_{\text{CM}} = \frac{1}{M} \sum_{j=1}^N m_j \vec{r}_j = \frac{1}{M} \sum_{j=1}^N m \vec{r}_j = \frac{m}{M} \sum_{j=1}^N \vec{r}_j = \frac{Nm}{M} \frac{\sum_{j=1}^N \vec{r}_j}{N}$$

where we sum over the number N of unit cells in the crystal and m is the mass of a unit cell. Because $Nm = M$, we can write

$$\vec{r}_{\text{CM}} = \frac{m}{M} \sum_{j=1}^N \vec{r}_j = \frac{Nm}{M} \frac{\sum_{j=1}^N \vec{r}_j}{N} = \frac{1}{N} \sum_{j=1}^N \vec{r}_j.$$

This is the definition of the geometric center of the crystal, so the center of mass is at the same point as the geometric center.

Two crucial concepts come out of these examples:

1. As with all problems, you must define your coordinate system and origin. For center-of-mass calculations, it often makes sense to choose your origin to be located at one of the masses of your system. That choice automatically defines its distance in [\[link\]](#) to be zero. However, you must still include the mass of the object at your origin in your calculation of M , the total mass [\[link\]](#). In the Earth-moon system example, this means including the mass of Earth. If you hadn't, you'd have ended up with the center of mass of the system being at the center of the moon, which is clearly wrong.
2. In the second example (the salt crystal), notice that there is no mass at all at the location of the center of mass. This is an example of what we stated above, that there does not have to be any actual mass at the center of mass of an object.

Center of Mass of Continuous Objects

If the object in question has its mass distributed uniformly in space, rather than as a collection of discrete particles, then $m_j \rightarrow dm$, and the summation becomes an integral:

Note:

Equation:

$$\vec{r}_{\text{CM}} = \frac{1}{M} \int \vec{r} dm.$$

In this context, r is a characteristic dimension of the object (the radius of a sphere, the length of a long rod). To generate an integrand that can actually be calculated, you need to express the differential mass element dm as a function of the mass density of the continuous object, and the dimension r . An example will clarify this.

Example:**CM of a Uniform Thin Hoop**

Find the center of mass of a uniform thin hoop (or ring) of mass M and radius r .

Strategy

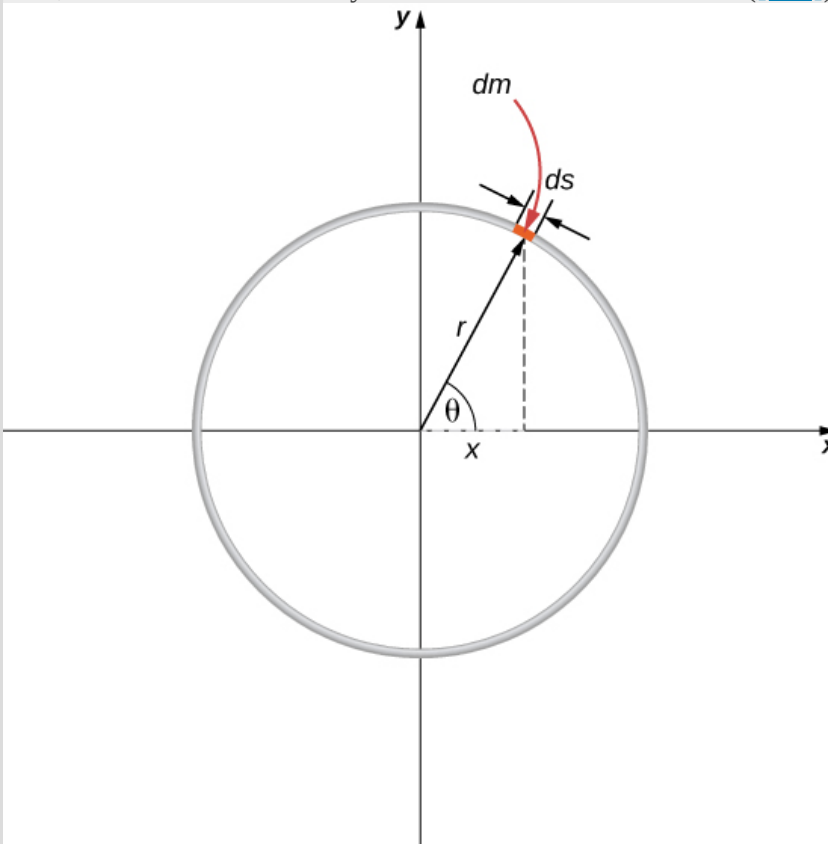
First, the hoop's symmetry suggests the center of mass should be at its geometric center. If we define our coordinate system such that the origin is located at the center of the hoop, the integral should evaluate to zero.

We replace dm with an expression involving the density of the hoop and the radius of the hoop. We then have an expression we can actually integrate. Since the hoop is described as “thin,” we treat it as a one-dimensional object, neglecting the thickness of the hoop. Therefore, its density is expressed as the number of kilograms of material per meter. Such a density is called a **linear mass density**, and is given the symbol λ ; this is the Greek letter “lambda,” which is the equivalent of the English letter “l” (for “linear”).

Since the hoop is described as uniform, this means that the linear mass density λ is constant. Thus, to get our expression for the differential mass element dm , we multiply λ by a differential length of the hoop, substitute, and integrate (with appropriate limits for the definite integral).

Solution

First, define our coordinate system and the relevant variables ([\[link\]](#)).



Finding the center of mass of a uniform hoop. We express the coordinates of a differential piece of the hoop, and then integrate around the hoop.

The center of mass is calculated with [\[link\]](#):

Equation:

$$\vec{\mathbf{r}}_{\text{CM}} = \frac{1}{M} \int_a^b \vec{\mathbf{r}} dm.$$

We have to determine the limits of integration a and b . Expressing $\vec{\mathbf{r}}$ in component form gives us

Equation:

$$\vec{\mathbf{r}}_{\text{CM}} = \frac{1}{M} \int_a^b \left[(r \cos \theta) \hat{\mathbf{i}} + (r \sin \theta) \hat{\mathbf{j}} \right] dm.$$

In the diagram, we highlighted a piece of the hoop that is of differential length ds ; it therefore has a differential mass $dm = \lambda ds$. Substituting:

Equation:

$$\vec{\mathbf{r}}_{\text{CM}} = \frac{1}{M} \int_a^b \left[(r \cos \theta) \hat{\mathbf{i}} + (r \sin \theta) \hat{\mathbf{j}} \right] \lambda ds.$$

However, the arc length ds subtends a differential angle $d\theta$, so we have

Equation:

$$ds = r d\theta$$

and thus

Equation:

$$\vec{\mathbf{r}}_{\text{CM}} = \frac{1}{M} \int_a^b \left[(r \cos \theta) \hat{\mathbf{i}} + (r \sin \theta) \hat{\mathbf{j}} \right] \lambda r d\theta.$$

One more step: Since λ is the linear mass density, it is computed by dividing the total mass by the length of the hoop:

Equation:

$$\lambda = \frac{M}{2\pi r}$$

giving us

Equation:

$$\begin{aligned} \vec{\mathbf{r}}_{\text{CM}} &= \frac{1}{M} \int_a^b \left[(r \cos \theta) \hat{\mathbf{i}} + (r \sin \theta) \hat{\mathbf{j}} \right] \left(\frac{M}{2\pi r} \right) r d\theta \\ &= \frac{1}{2\pi} \int_a^b \left[(r \cos \theta) \hat{\mathbf{i}} + (r \sin \theta) \hat{\mathbf{j}} \right] d\theta. \end{aligned}$$

Notice that the variable of integration is now the angle θ . This tells us that the limits of integration (around the circular hoop) are $\theta = 0$ to $\theta = 2\pi$, so $a = 0$ and $b = 2\pi$. Also, for convenience, we separate the integral into the x- and y-components of $\vec{\mathbf{r}}_{\text{CM}}$. The final integral expression is

Equation:

$$\begin{aligned}
\vec{r}_{\text{CM}} &= r_{\text{CM},x}\hat{\mathbf{i}} + r_{\text{CM},y}\hat{\mathbf{j}} \\
&= \left[\frac{1}{2\pi} \int_0^{2\pi} (r\cos\theta)d\theta \right] \hat{\mathbf{i}} + \left[\frac{1}{2\pi} \int_0^{2\pi} (r\sin\theta)d\theta \right] \hat{\mathbf{j}} \\
&= 0\hat{\mathbf{i}} + 0\hat{\mathbf{j}} = \vec{0}
\end{aligned}$$

as expected.

Center of Mass and Conservation of Momentum

How does all this connect to conservation of momentum?

Suppose you have N objects with masses $m_1, m_2, m_3, \dots, m_N$ and initial velocities $\vec{v}_1, \vec{v}_2, \vec{v}_3, \dots, \vec{v}_N$. The center of mass of the objects is

Equation:

$$\vec{r}_{\text{CM}} = \frac{1}{M} \sum_{j=1}^N m_j \vec{r}_j.$$

Its velocity is

Equation:

$$\vec{v}_{\text{CM}} = \frac{d\vec{r}_{\text{CM}}}{dt} = \frac{1}{M} \sum_{j=1}^N m_j \frac{d\vec{r}_j}{dt}$$

and thus the initial momentum of the center of mass is

Equation:

$$\begin{aligned}
\left[M \frac{d\vec{r}_{\text{CM}}}{dt} \right]_{\text{i}} &= \sum_{j=1}^N m_j \frac{d\vec{r}_{j,\text{i}}}{dt} \\
M \vec{v}_{\text{CM},\text{i}} &= \sum_{j=1}^N m_j \vec{v}_{j,\text{i}}.
\end{aligned}$$

After these masses move and interact with each other, the momentum of the center of mass is

Equation:

$$M \vec{v}_{\text{CM},\text{f}} = \sum_{j=1}^N m_j \vec{v}_{j,\text{f}}.$$

But conservation of momentum tells us that the right-hand side of both equations must be equal, which says

Note:
Equation:

$$M\vec{v}_{CM,f} = M\vec{v}_{CM,i}.$$

This result implies that conservation of momentum is expressed in terms of the center of mass of the system. Notice that as an object moves through space with no net external force acting on it, an individual particle of the object may accelerate in various directions, with various magnitudes, depending on the net internal force acting on that object at any time. (Remember, it is only the vector sum of all the internal forces that vanishes, not the internal force on a single particle.) Thus, such a particle's momentum will not be constant—but the momentum of the entire extended object will be, in accord with [\[link\]](#).

[\[link\]](#) implies another important result: Since M represents the mass of the entire system of particles, it is necessarily constant. (If it isn't, we don't have a closed system, so we can't expect the system's momentum to be conserved.) As a result, [\[link\]](#) implies that, for a closed system,

Note:
Equation:

$$\vec{v}_{CM,f} = \vec{v}_{CM,i}.$$

That is to say, *in the absence of an external force, the velocity of the center of mass never changes.*

You might be tempted to shrug and say, “Well yes, that’s just Newton’s first law,” but remember that Newton’s first law discusses the constant velocity of a particle, whereas [\[link\]](#) applies to the center of mass of a (possibly vast) collection of interacting particles, and that there may not be any particle at the center of mass at all! So, this really is a remarkable result.

Example:

Fireworks Display

When a fireworks rocket explodes, thousands of glowing fragments fly outward in all directions, and fall to Earth in an elegant and beautiful display ([\[link\]](#)). Describe what happens, in terms of conservation of momentum and center of mass.



These exploding fireworks are a vivid example of conservation of momentum and the motion of the center of mass.

The picture shows radial symmetry about the central points of the explosions; this suggests the idea of center of mass. We can also see the parabolic motion of the glowing particles; this brings to mind projectile motion ideas.

Solution

Initially, the fireworks rocket is launched and flies more or less straight upward; this is the cause of the more-or-less-straight, white trail going high into the sky below the explosion in the upper-right of the picture (the yellow explosion). This trail is not parabolic because the explosive shell, during its launch phase, is actually a rocket; the impulse applied to it by the ejection of the burning fuel applies a force on the shell during the rise-time interval. (This is a phenomenon we will study in the next section.) The shell has multiple forces on it; thus, it is not in free-fall prior to the explosion. At the instant of the explosion, the thousands of glowing fragments fly outward in a radially symmetrical pattern. The symmetry of the explosion is the result of all the internal forces summing

to zero $\left(\sum_j \vec{f}_j^{\text{int}} = 0 \right)$; for every internal force, there is another that is equal in magnitude and opposite in direction.

However, as we learned above, these internal forces cannot change the momentum of the center of mass of the (now exploded) shell. Since the rocket force has now vanished, the center of mass of the shell is now a projectile (the only force on it is gravity), so its trajectory does become parabolic. The

two red explosions on the left show the path of their centers of mass at a slightly longer time after explosion compared to the yellow explosion on the upper right.

In fact, if you look carefully at all three explosions, you can see that the glowing trails are not truly radially symmetric; rather, they are somewhat denser on one side than the other. Specifically, the yellow explosion and the lower middle explosion are slightly denser on their right sides, and the upper-left explosion is denser on its left side. This is because of the momentum of their centers of mass; the differing trail densities are due to the momentum each piece of the shell had at the moment of its explosion. The fragment for the explosion on the upper left of the picture had a momentum that pointed upward and to the left; the middle fragment's momentum pointed upward and slightly to the right; and the right-side explosion clearly upward and to the right (as evidenced by the white rocket exhaust trail visible below the yellow explosion).

Finally, each fragment is a projectile on its own, thus tracing out thousands of glowing parabolas.

Significance

In the discussion above, we said, "...the center of mass of the shell is now a projectile (the only force on it is gravity)...." This is not quite accurate, for there may not be any mass at all at the center of mass; in which case, there could not be a force acting on it. This is actually just verbal shorthand for describing the fact that the gravitational forces on all the particles act so that the center of mass changes position exactly as if all the mass of the shell were always located at the position of the center of mass.

Note:

Exercise:

Problem:

Check Your Understanding How would the firework display change in deep space, far away from any source of gravity?

Solution:

The explosions would essentially be spherically symmetric, because gravity would not act to distort the trajectories of the expanding projectiles.

You may sometimes hear someone describe an explosion by saying something like, "the fragments of the exploded object always move in a way that makes sure that the center of mass continues to move on its original trajectory." This makes it sound as if the process is somewhat magical: how can it be that, in *every* explosion, it *always* works out that the fragments move in just the right way so that the center of mass' motion is unchanged? Phrased this way, it would be hard to believe no explosion ever does anything differently.

The explanation of this apparently astonishing coincidence is: We defined the center of mass precisely so this is exactly what we would get. Recall that first we defined the momentum of the system:

Equation:

$$\vec{\mathbf{p}}_{\text{CM}} = \sum_{j=1}^N \frac{d\vec{\mathbf{p}}_j}{dt}.$$

We then concluded that the net external force on the system (if any) changed this momentum:

Equation:

$$\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}_{\text{CM}}}{dt}$$

and then—and here’s the point—we defined an acceleration that would obey Newton’s second law. That is, we demanded that we should be able to write

Equation:

$$\vec{\mathbf{a}} = \frac{\vec{\mathbf{F}}}{M}$$

which requires that

Equation:

$$\vec{\mathbf{a}} = \frac{d^2}{dt^2} \left(\frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{r}}_j \right).$$

where the quantity inside the parentheses is the center of mass of our system. So, it’s not astonishing that the center of mass obeys Newton’s second law; we defined it so that it would.

Summary

- An extended object (made up of many objects) has a defined position vector called the center of mass.
- The center of mass can be thought of, loosely, as the average location of the total mass of the object.
- The center of mass of an object traces out the trajectory dictated by Newton’s second law, due to the net external force.
- The internal forces within an extended object cannot alter the momentum of the extended object as a whole.

Conceptual Questions

Exercise:

Problem:

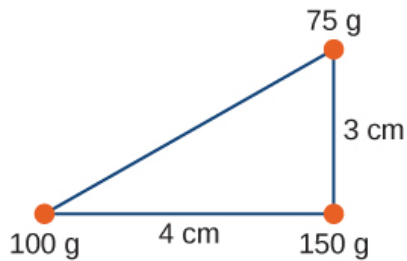
Suppose a fireworks shell explodes, breaking into three large pieces for which air resistance is negligible. How does the explosion affect the motion of the center of mass? How would it be affected if the pieces experienced significantly more air resistance than the intact shell?

Problems

Exercise:

Problem:

Three point masses are placed at the corners of a triangle as shown in the figure below.



Find the center of mass of the three-mass system.

Solution:

With the origin defined to be at the position of the 150-g mass, $x_{\text{CM}} = -1.23\text{cm}$ and $y_{\text{CM}} = 0.69\text{cm}$

Exercise:

Problem:

Two particles of masses m_1 and m_2 separated by a horizontal distance D are released from the same height h at the same time. Find the vertical position of the center of mass of these two particles at a time before the two particles strike the ground. Assume no air resistance.

Exercise:

Problem:

Two particles of masses m_1 and m_2 separated by a horizontal distance D are let go from the same height h at different times. Particle 1 starts at $t = 0$, and particle 2 is let go at $t = T$. Find the vertical position of the center of mass at a time before the first particle strikes the ground. Assume no air resistance.

Solution:

$$y_{\text{CM}} = \begin{cases} \frac{h}{2} - \frac{1}{4}gt^2, & t < T \\ h - \frac{1}{2}gt^2 - \frac{1}{4}gT^2 + \frac{1}{2}gtT, & t \geq T \end{cases}$$

Exercise:

Problem:

Two particles of masses m_1 and m_2 move uniformly in different circles of radii R_1 and R_2 about origin in the x, y -plane. The x - and y -coordinates of the center of mass and that of particle 1 are given as follows (where length is in meters and t in seconds):

$$x_1(t) = 4\cos(2t), y_1(t) = 4\sin(2t)$$

and:

$$x_{\text{CM}}(t) = 3\cos(2t), y_{\text{CM}}(t) = 3\sin(2t).$$

- Find the radius of the circle in which particle 1 moves.
- Find the x - and y -coordinates of particle 2 and the radius of the circle this particle moves.

Exercise:**Problem:**

Two particles of masses m_1 and m_2 move uniformly in different circles of radii R_1 and R_2 about the origin in the x, y -plane. The coordinates of the two particles in meters are given as follows ($z = 0$ for both). Here t is in seconds:

$$x_1(t) = 4 \cos(2t)$$

$$y_1(t) = 4 \sin(2t)$$

$$x_2(t) = 2 \cos\left(3t - \frac{\pi}{2}\right)$$

$$y_2(t) = 2 \sin\left(3t - \frac{\pi}{2}\right)$$

- Find the radii of the circles of motion of both particles.
- Find the x - and y -coordinates of the center of mass.
- Decide if the center of mass moves in a circle by plotting its trajectory.

Solution:

a. $R_1 = 4$ m, $R_2 = 2$ m; b. $X_{\text{CM}} = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}$, $Y_{\text{CM}} = \frac{m_1 y_1 + m_2 y_2}{m_1 + m_2}$; c. yes, with

$$R = \frac{1}{m_1 + m_2} \sqrt{16m_1^2 + 4m_2^2}$$

Exercise:**Problem:**

Find the center of mass of a one-meter long rod, made of 50 cm of iron (density $8 \frac{\text{g}}{\text{cm}^3}$) and 50 cm of aluminum (density $2.7 \frac{\text{g}}{\text{cm}^3}$).

Exercise:

Problem:

Find the center of mass of a rod of length L whose mass density changes from one end to the other quadratically. That is, if the rod is laid out along the x -axis with one end at the origin and the other end at $x = L$, the density is given by $\rho(x) = \rho_0 + (\rho_1 - \rho_0)\left(\frac{x}{L}\right)^2$, where ρ_0 and ρ_1 are constant values.

Solution:

$$x_{cm} = \frac{3}{4} L \left(\frac{\rho_1 + \rho_0}{\rho_1 + 2\rho_0} \right)$$

Exercise:**Problem:**

Find the center of mass of a rectangular block of length a and width b that has a nonuniform density such that when the rectangle is placed in the x,y -plane with one corner at the origin and the block placed in the first quadrant with the two edges along the x - and y -axes, the density is given by $\rho(x, y) = \rho_0 x$, where ρ_0 is a constant.

Exercise:**Problem:**

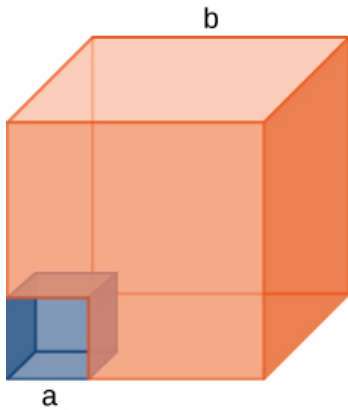
Find the center of mass of a rectangular material of length a and width b made up of a material of nonuniform density. The density is such that when the rectangle is placed in the xy -plane, the density is given by $\rho(x, y) = \rho_0 xy$.

Solution:

$$\left(\frac{2a}{3}, \frac{2b}{3} \right)$$

Exercise:

Problem: A cube of side a is cut out of another cube of side b as shown in the figure below.



Find the location of the center of mass of the structure. (*Hint:* Think of the missing part as a negative mass overlapping a positive mass.)

Exercise:**Problem:**

Find the center of mass of a cone of uniform density that has a radius R at the base, height h , and mass M . Let the origin be at the center of the base of the cone and have $+z$ going through the cone vertex.

Solution:

$$(x_{\text{CM}}, y_{\text{CM}}, z_{\text{CM}}) = (0, 0, h/4)$$

Exercise:**Problem:**

Find the center of mass of a thin wire of mass m and length L bent in a semicircular shape. Let the origin be at the center of the semicircle and have the wire arc from the $+x$ axis, cross the $+y$ axis, and terminate at the $-x$ axis.

Exercise:**Problem:**

Find the center of mass of a uniform thin semicircular plate of radius R . Let the origin be at the center of the semicircle, the plate arc from the $+x$ axis to the $-x$ axis, and the z axis be perpendicular to the plate.

Solution:

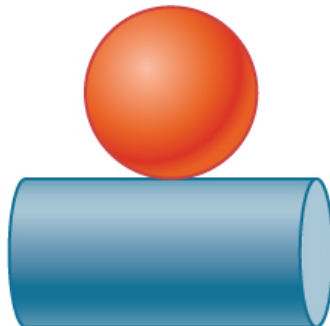
$$(x_{\text{CM}}, y_{\text{CM}}, z_{\text{CM}}) = (0, 4R/(3\pi), 0)$$

Exercise:**Problem:**

Find the center of mass of a sphere of mass M and radius R and a cylinder of mass m , radius r , and height h arranged as shown below.



(a)



(b)

Express your answers in a coordinate system that has the origin at the center of the cylinder.

Glossary

center of mass

weighted average position of the mass

external force

force applied to an extended object that changes the momentum of the extended object as a whole

internal force

force that the simple particles that make up an extended object exert on each other. Internal forces can be attractive or repulsive

linear mass density

λ , expressed as the number of kilograms of material per meter

Rocket Propulsion

By the end of this section, you will be able to:

- Describe the application of conservation of momentum when the mass changes with time, as well as the velocity
- Calculate the speed of a rocket in empty space, at some time, given initial conditions
- Calculate the speed of a rocket in Earth's gravity field, at some time, given initial conditions

Now we deal with the case where the mass of an object is changing. We analyze the motion of a rocket, which changes its velocity (and hence its momentum) by ejecting burned fuel gases, thus causing it to accelerate in the opposite direction of the velocity of the ejected fuel (see [\[link\]](#)).

Specifically: A fully fueled rocket ship in deep space has a total mass m_0 (this mass includes the initial mass of the fuel). At some moment in time, the rocket has a velocity \vec{v} and mass m ; this mass is a combination of the mass of the empty rocket and the mass of the remaining unburned fuel it contains. (We refer to m as the “instantaneous mass” and \vec{v} as the “instantaneous velocity.”) The rocket accelerates by burning the fuel it carries and ejecting the burned exhaust gases. If the burn rate of the fuel is constant, and the velocity at which the exhaust is ejected is also constant, what is the change of velocity of the rocket as a result of burning all of its fuel?



The space shuttle had a number of reusable parts. Solid fuel boosters on either side were recovered and refueled after each flight, and the entire orbiter returned to Earth for use in subsequent flights. The large liquid fuel tank was expended. The space shuttle was a complex assemblage of technologies, employing both solid and liquid fuel, and pioneering ceramic tiles as reentry heat shields. As a result, it permitted multiple launches as

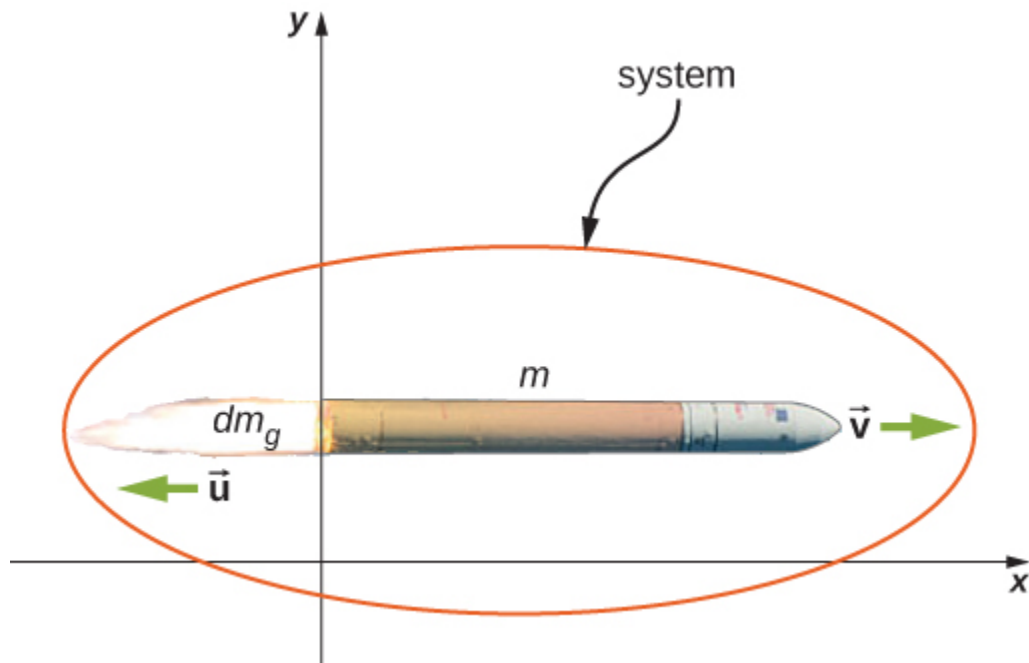
opposed to single-use rockets. (credit: modification of work by NASA)

Physical Analysis

Here's a description of what happens, so that you get a feel for the physics involved.

- As the rocket engines operate, they are continuously ejecting burned fuel gases, which have both mass and velocity, and therefore some momentum. By conservation of momentum, the rocket's momentum changes by this same amount (with the opposite sign). We will assume the burned fuel is being ejected at a constant rate, which means the rate of change of the rocket's momentum is also constant. By [\[link\]](#), this represents a constant force on the rocket.
- However, as time goes on, the mass of the rocket (which includes the mass of the remaining fuel) continuously decreases. Thus, even though the force on the rocket is constant, the resulting acceleration is not; it is continuously increasing.
- So, the total change of the rocket's velocity will depend on the amount of mass of fuel that is burned, and that dependence is not linear.

The problem has the mass and velocity of the rocket changing; also, the total mass of ejected gases is changing. If we define our system to be the rocket + fuel, then this is a closed system (since the rocket is in deep space, there are no external forces acting on this system); as a result, momentum is conserved for this system. Thus, we can apply conservation of momentum to answer the question ([\[link\]](#)).



The rocket accelerates to the right due to the expulsion of some of its fuel mass to the left. Conservation of momentum enables us to determine the resulting change of velocity. The mass m is the instantaneous total mass of the rocket (i.e., mass of rocket body plus mass of fuel at that point in time). (credit: modification of work by NASA/Bill Ingalls)

At the same moment that the total instantaneous rocket mass is m (i.e., m is the mass of the rocket body plus the mass of the fuel at that point in time), we define the rocket's instantaneous velocity to be $\vec{v} = v\hat{i}$ (in the $+x$ -direction); this velocity is measured relative to an inertial reference system (the Earth, for example). Thus, the initial momentum of the system is

$$\vec{p}_i = mv\hat{i}.$$

The rocket's engines are burning fuel at a constant rate and ejecting the exhaust gases in the $-x$ -direction. During an infinitesimal time interval dt , the engines eject a (positive) infinitesimal mass of gas dm_g at velocity

$\vec{u} = -u\hat{i}$; note that although the rocket velocity $v\hat{i}$ is measured with respect to Earth, the exhaust gas velocity is measured with respect to the (moving) rocket. Measured with respect to the Earth, therefore, the exhaust gas has velocity $(v - u)\hat{i}$.

As a consequence of the ejection of the fuel gas, the rocket's mass decreases by dm_g , and its velocity increases by $dv\hat{i}$. Therefore, including both the change for the rocket and the change for the exhaust gas, the final momentum of the system is

Equation:

$$\begin{aligned}\vec{p}_f &= \vec{p}_{\text{rocket}} + \vec{p}_{\text{gas}} \\ &= (m - dm_g)(v + dv)\hat{i} + dm_g(v - u)\hat{i}\end{aligned}$$

Since all vectors are in the x -direction, we drop the vector notation. Applying conservation of momentum, we obtain

Equation:

$$\begin{aligned}p_i &= p_f \\ mv &= (m - dm_g)(v + dv) + dm_g(v - u) \\ mv &= mv + mdv - dm_gv - dm_gdv + dm_gv - dm_gu \\ mdv &= dm_gdv + dm_gu.\end{aligned}$$

Now, dm_g and dv are each very small; thus, their product dm_gdv is very, very small, much smaller than the other two terms in this expression. We neglect this term, therefore, and obtain:

Equation:

$$mdv = dm_gu.$$

Our next step is to remember that, since dm_g represents an increase in the mass of ejected gases, it must also represent a decrease of mass of the rocket:

Equation:

$$dm_g = -dm.$$

Replacing this, we have

Equation:

$$mdv = -dmu$$

or

Equation:

$$dv = -u \frac{dm}{m}.$$

Integrating from the initial mass m_0 to the final mass m of the rocket gives us the result we are after:

Equation:

$$\begin{aligned} \int_{v_i}^v dv &= -u \int_{m_0}^m \frac{1}{m} dm \\ v - v_i &= u \ln \left(\frac{m_0}{m} \right) \end{aligned}$$

and thus our final answer is

Note:

Equation:

$$\Delta v = u \ln \left(\frac{m_0}{m} \right).$$

This result is called the **rocket equation**. It was originally derived by the Soviet physicist Konstantin Tsiolkovsky in 1897. It gives us the change of velocity that the rocket obtains from burning a mass of fuel that decreases the total rocket mass from m_0 down to m . As expected, the relationship between Δv and the change of mass of the rocket is nonlinear.

Note:

Rocket Propulsion

In rocket problems, the most common questions are finding the change of velocity due to burning some amount of fuel for some amount of time; or to determine the acceleration that results from burning fuel.

1. To determine the change of velocity, use the rocket equation [\[link\]](#).
2. To determine the acceleration, determine the force by using the impulse-momentum theorem, using the rocket equation to determine the change of velocity.

Example:

Thrust on a Spacecraft

A spacecraft is moving in gravity-free space along a straight path when its pilot decides to accelerate forward. He turns on the thrusters, and burned fuel is ejected at a constant rate of 2.0×10^2 kg/s, at a speed (relative to the rocket) of 2.5×10^2 m/s. The initial mass of the spacecraft and its unburned fuel is 2.0×10^4 kg, and the thrusters are on for 30 s.

- a. What is the thrust (the force applied to the rocket by the ejected fuel) on the spacecraft?
- b. What is the spacecraft's acceleration as a function of time?
- c. What are the spacecraft's accelerations at $t = 0, 15, 30$, and 35 s?

Strategy

- a. The force on the spacecraft is equal to the rate of change of the momentum of the fuel.
- b. Knowing the force from part (a), we can use Newton's second law to calculate the consequent acceleration. The key here is that, although the force applied to the spacecraft is constant (the fuel is being ejected at a constant rate), the mass of the spacecraft isn't; thus, the acceleration caused by the force won't be constant. We expect to get a function $a(t)$, therefore.
- c. We'll use the function we obtain in part (b), and just substitute the numbers given. Important: We expect that the acceleration will get larger as time goes on, since the mass being accelerated is continuously decreasing (fuel is being ejected from the rocket).

Solution

- a. The momentum of the ejected fuel gas is
Equation:

$$p = m_g v.$$

The ejection velocity $v = 2.5 \times 10^2 \text{ m/s}$ is constant, and therefore the force is

Equation:

$$F = \frac{dp}{dt} = v \frac{dm_g}{dt} = -v \frac{dm}{dt}.$$

Now, $\frac{dm_g}{dt}$ is the rate of change of the mass of the fuel; the problem states that this is $2.0 \times 10^2 \text{ kg/s}$. Substituting, we get

Equation:

$$\begin{aligned} F &= v \frac{dm_g}{dt} \\ &= \left(2.5 \times 10^2 \frac{\text{m}}{\text{s}} \right) \left(2.0 \times 10^2 \frac{\text{kg}}{\text{s}} \right) \\ &= 5 \times 10^4 \text{ N}. \end{aligned}$$

b. Above, we defined m to be the combined mass of the empty rocket plus however much unburned fuel it contained: $m = m_R + m_g$.

From Newton's second law,

Equation:

$$a = \frac{F}{m} = \frac{F}{m_R + m_g}.$$

The force is constant and the empty rocket mass m_R is constant, but the fuel mass m_g is decreasing at a uniform rate; specifically:

Equation:

$$m_g = m_g(t) = m_{g_0} - \left(\frac{dm_g}{dt} \right) t.$$

This gives us

Equation:

$$a(t) = \frac{F}{m_{g_i} - \left(\frac{dm_g}{dt} \right) t} = \frac{F}{M - \left(\frac{dm_g}{dt} \right) t}.$$

Notice that, as expected, the acceleration is a function of time.

Substituting the given numbers:

Equation:

$$a(t) = \frac{5 \times 10^4 \text{ N}}{2.0 \times 10^4 \text{ kg} - \left(2.0 \times 10^2 \frac{\text{kg}}{\text{s}} \right) t}.$$

c. At $t = 0 \text{ s}$:

Equation:

$$a(0 \text{ s}) = \frac{5 \times 10^4 \text{ N}}{2.0 \times 10^4 \text{ kg} - \left(2.0 \times 10^2 \frac{\text{kg}}{\text{s}} \right) (0 \text{ s})} = 2.5 \frac{\text{m}}{\text{s}^2}.$$

At $t = 15$ s, $a(15 \text{ s}) = 2.9 \text{ m/s}^2$.

At $t = 30$ s, $a(30 \text{ s}) = 3.6 \text{ m/s}^2$.

Acceleration is increasing, as we expected.

Significance

Notice that the acceleration is not constant; as a result, any dynamical quantities must be calculated either using integrals, or (more easily) conservation of total energy.

Note:

Exercise:

Problem:

Check Your Understanding What is the physical difference (or relationship) between $\frac{dm}{dt}$ and $\frac{dm_g}{dt}$ in this example?

Solution:

The notation m_g stands for the mass of the fuel and m stands for the mass of the rocket plus the initial mass of the fuel. Note that m_g changes with time, so we write it as $m_g(t)$. Using m_R as the mass of the rocket with no fuel, the total mass of the rocket plus fuel is $m = m_R + m_g(t)$. Differentiation with respect to time gives

$$\frac{dm}{dt} = \frac{dm_R}{dt} + \frac{dm_g(t)}{dt} = \frac{dm_g(t)}{dt}$$

where we used $\frac{dm_R}{dt} = 0$ because the mass of the rocket does not change. Thus, time rate of change of the mass of the rocket is the same as that of the fuel.

Rocket in a Gravitational Field

Let's now analyze the velocity change of the rocket during the launch phase, from the surface of Earth. To keep the math manageable, we'll restrict our attention to distances for which the acceleration caused by gravity can be treated as a constant g .

The analysis is similar, except that now there is an external force of $\vec{F} = -mg\hat{j}$ acting on our system. This force applies an impulse $d\vec{J} = \vec{F}dt = -mgdt\hat{j}$, which is equal to the change of momentum. This gives us

Equation:

$$\begin{aligned} d\vec{p} &= d\vec{J} \\ \vec{p}_f - \vec{p}_i &= -mgdt\hat{j} \\ [(m - dm_g)(v + dv) + dm_g(v - u) - mv]\hat{j} &= -mgdt\hat{j} \end{aligned}$$

and so

Equation:

$$mdv - dm_g u = -mgdt$$

where we have again neglected the term $dm_g dv$ and dropped the vector notation. Next we replace dm_g with $-dm$:

Equation:

$$\begin{aligned} mdv + dm u &= -mgdt \\ mdv &= -dm u - mgdt. \end{aligned}$$

Dividing through by m gives

Equation:

$$dv = -u \frac{dm}{m} - gdt$$

and integrating, we have

Note:

Equation:

$$\Delta v = u \ln \left(\frac{m_0}{m} \right) - g \Delta t.$$

Unsurprisingly, the rocket's velocity is affected by the (constant) acceleration of gravity.

Remember that Δt is the burn time of the fuel. Now, in the absence of gravity, [\[link\]](#) implies that it makes no difference how much time it takes to burn the entire mass of fuel; the change of velocity does not depend on Δt . However, in the presence of gravity, it matters a lot. The $-g\Delta t$ term in [\[link\]](#) tells us that the *longer* the burn time is, the *smaller* the rocket's change of velocity will be. This is the reason that the launch of a rocket is so spectacular at the first moment of liftoff: It's essential to burn the fuel as quickly as possible, to get as large a Δv as possible.

Summary

- A rocket is an example of conservation of momentum where the mass of the system is not constant, since the rocket ejects fuel to provide thrust.
- The rocket equation gives us the change of velocity that the rocket obtains from burning a mass of fuel that decreases the total rocket mass.

Key Equations

Definition of momentum	$\vec{\mathbf{p}} = m\vec{\mathbf{v}}$
Impulse	$\vec{\mathbf{J}} \equiv \int_{t_i}^{t_f} \vec{\mathbf{F}}(t)dt$ or $\vec{\mathbf{J}} = \vec{\mathbf{F}}_{\text{ave}}\Delta t$
Impulse-momentum theorem	$\vec{\mathbf{J}} = \Delta\vec{\mathbf{p}}$
Average force from momentum	$\vec{\mathbf{F}} = \frac{\Delta\vec{\mathbf{p}}}{\Delta t}$
Instantaneous force from momentum (Newton's second law)	$\vec{\mathbf{F}}(t) = \frac{d\vec{\mathbf{p}}}{dt}$
Conservation of momentum	$\frac{d\vec{\mathbf{p}}_1}{dt} + \frac{d\vec{\mathbf{p}}_2}{dt} = 0$ or $\vec{\mathbf{p}}_1 + \vec{\mathbf{p}}_2 = \text{constant}$
Generalized conservation of momentum	$\sum_{j=1}^N \vec{\mathbf{p}}_j = \text{constant}$
Conservation of momentum in two dimensions	$p_{f,x} = p_{1,i,x} + p_{2,i,x}$ $p_{f,y} = p_{1,i,y} + p_{2,i,y}$
External forces	$\vec{\mathbf{F}}_{\text{ext}} = \sum_{j=1}^N \frac{d\vec{\mathbf{p}}_j}{dt}$
Newton's second law for an	$\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}_{\text{CM}}}{dt}$

extended object	
Acceleration of the center of mass	$\vec{\mathbf{a}}_{\text{CM}} = \frac{d^2}{dt^2} \left(\frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{r}}_j \right) = \frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{a}}_j$
Position of the center of mass for a system of particles	$\vec{\mathbf{r}}_{\text{CM}} \equiv \frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{r}}_j$
Velocity of the center of mass	$\vec{\mathbf{v}}_{\text{CM}} = \frac{d}{dt} \left(\frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{r}}_j \right) = \frac{1}{M} \sum_{j=1}^N m_j \vec{\mathbf{v}}_j$
Position of the center of mass of a continuous object	$\vec{\mathbf{r}}_{\text{CM}} \equiv \frac{1}{M} \int \vec{\mathbf{r}} \, dm$
Rocket equation	$\Delta v = u \ln \left(\frac{m_i}{m} \right)$

Conceptual Questions

Exercise:

Problem:

It is possible for the velocity of a rocket to be greater than the exhaust velocity of the gases it ejects. When that is the case, the gas velocity and gas momentum are in the same direction as that of the rocket. How is the rocket still able to obtain thrust by ejecting the gases?

Solution:

Yes, the rocket speed can exceed the exhaust speed of the gases it ejects. The thrust of the rocket does not depend on the relative speeds

of the gases and rocket, it simply depends on conservation of momentum.

Problems

Exercise:

Problem:

(a) A 5.00-kg squid initially at rest ejects 0.250 kg of fluid with a velocity of 10.0 m/s. What is the recoil velocity of the squid if the ejection is done in 0.100 s and there is a 5.00-N frictional force opposing the squid's movement?

(b) How much energy is lost to work done against friction?

Solution:

(a) 0.413 m/s, (b) about 0.2 J

Exercise:

Problem:

A rocket takes off from Earth and reaches a speed of 100 m/s in 10.0 s. If the exhaust speed is 1500 m/s and the mass of fuel burned is 100 kg, what was the initial mass of the rocket?

Exercise:

Problem:

Repeat the preceding problem but for a rocket that takes off from a space station, where there is no gravity other than the negligible gravity due to the space station.

Solution:

1551 kg

Exercise:**Problem:**

How much fuel would be needed for a 1000-kg rocket (this is its mass with no fuel) to take off from Earth and reach 1000 m/s in 30 s? The exhaust speed is 1000 m/s.

Exercise:**Problem:**

What exhaust speed is required to accelerate a rocket in deep space from 800 m/s to 1000 m/s in 5.0 s if the total rocket mass is 1200 kg and the rocket only has 50 kg of fuel left?

Solution:

4.9 km/s

Exercise:**Problem:**

Unreasonable Results Squids have been reported to jump from the ocean and travel 30.0 m (measured horizontally) before re-entering the water.

(a) Calculate the initial speed of the squid if it leaves the water at an angle of 20.0° , assuming negligible lift from the air and negligible air resistance.

(b) The squid propels itself by squirting water. What fraction of its mass would it have to eject in order to achieve the speed found in the previous part? The water is ejected at 12.0 m/s; gravitational force and friction are neglected.

(c) What is unreasonable about the results?

(d) Which premise is unreasonable, or which premises are inconsistent?

Additional Problems

Exercise:

Problem:

Two 70-kg canoers paddle in a single, 50-kg canoe. Their paddling moves the canoe at 1.2 m/s with respect to the water, and the river they're in flows at 4 m/s with respect to the land. What is their momentum with respect to the land?

Exercise:

Problem:

Which has a larger magnitude of momentum: a 3000-kg elephant moving at 40 km/h or a 60-kg cheetah moving at 112 km/h?

Solution:

the elephant has a higher momentum

Exercise:

Problem:

A driver applies the brakes and reduces the speed of her car by 20%, without changing the direction in which the car is moving. By how much does the car's momentum change?

Exercise:

Problem:

You friend claims that momentum is mass multiplied by velocity, so things with more mass have more momentum. Do you agree? Explain.

Solution:

Answers may vary. The first clause is true, but the second clause is not true in general because the velocity of an object with small mass may

be large enough so that the momentum of the object is greater than that of a larger-mass object with a smaller velocity.

Exercise:

Problem:

Dropping a glass on a cement floor is more likely to break the glass than if it is dropped from the same height on a grass lawn. Explain in terms of the impulse.

Exercise:

Problem:

Your 1500-kg sports car accelerates from 0 to 30 m/s in 10 s. What average force is exerted on it during this acceleration?

Solution:

$$4.5 \times 10^3 \text{ N}$$

Exercise:

Problem:

A ball of mass m is dropped. What is the formula for the impulse exerted on the ball from the instant it is dropped to an arbitrary time τ later? Ignore air resistance.

Exercise:

Problem:

Repeat the preceding problem, but including a drag force due to air of $f_{\text{drag}} = -b\vec{v}$.

Solution:

$$\vec{J} = \int_0^\tau \left[m\vec{g} - m\vec{g} \left(1 - e^{-bt/m} \right) \right] dt = \frac{m^2}{b} \vec{g} \left(e^{-b\tau/m} - 1 \right)$$

Exercise:

Problem:

A 5.0-g egg falls from a 90-cm-high counter onto the floor and breaks. What impulse is exerted by the floor on the egg?

Exercise:**Problem:**

A car crashes into a large tree that does not move. The car goes from 30 m/s to 0 in 1.3 m. (a) What impulse is applied to the driver by the seatbelt, assuming he follows the same motion as the car? (b) What is the average force applied to the driver by the seatbelt?

Solution:

a. $-(2.1 \times 10^3 \text{ kg} \cdot \text{m/s})\hat{\mathbf{i}}$, b. $-(24 \times 10^3 \text{ N})\hat{\mathbf{i}}$

Exercise:**Problem:**

Two hockey players approach each other head on, each traveling at the same speed v_i . They collide and get tangled together, falling down and moving off at a speed $v_i/5$. What is the ratio of their masses?

Exercise:**Problem:**

You are coasting on your 10-kg bicycle at 15 m/s and a 5.0-g bug splatters on your helmet. The bug was initially moving at 2.0 m/s in the same direction as you. If your mass is 60 kg, (a) what is the initial momentum of you plus your bicycle? (b) What is the initial momentum of the bug? (c) What is your change in velocity due to the collision with the bug? (d) What would the change in velocity have been if the bug were traveling in the opposite direction?

Solution:

- a. $(1.1 \times 10^3 \text{ kg} \cdot \text{m/s})\hat{\mathbf{i}}$, b. $(0.010 \text{ kg} \cdot \text{m/s})\hat{\mathbf{i}}$, c. $-(0.00093 \text{ m/s})\hat{\mathbf{i}}$, d. $-(0.0012 \text{ m/s})\hat{\mathbf{i}}$

Exercise:

Problem:

A load of gravel is dumped straight down into a 30 000-kg freight car coasting at 2.2 m/s on a straight section of a railroad. If the freight car's speed after receiving the gravel is 1.5 m/s, what mass of gravel did it receive?

Exercise:

Problem:

Two carts on a straight track collide head on. The first cart was moving at 3.6 m/s in the positive x direction and the second was moving at 2.4 m/s in the opposite direction. After the collision, the second car continues moving in its initial direction of motion at 0.24 m/s. If the mass of the second car is 5.0 times that of the first, what is the final velocity of the first car?

Solution:

$$-(7.2 \text{ m/s})\hat{\mathbf{i}}$$

Exercise:

Problem:

A 100-kg astronaut finds himself separated from his spaceship by 10 m and moving away from the spaceship at 0.1 m/s. To get back to the spaceship, he throws a 10-kg tool bag away from the spaceship at 5.0 m/s. How long will he take to return to the spaceship?

Exercise:

Problem:

Derive the equations giving the final speeds for two objects that collide elastically, with the mass of the objects being m_1 and m_2 and the initial speeds being $v_{1,i}$ and $v_{2,i} = 0$ (i.e., second object is initially stationary).

Solution:

$$v_{1,f} = v_{1,i} \frac{m_1 - m_2}{m_1 + m_2}, \quad v_{2,f} = v_{1,i} \frac{2m_1}{m_1 + m_2}$$

Exercise:**Problem:**

Repeat the preceding problem for the case when the initial speed of the second object is nonzero.

Exercise:**Problem:**

A child sleds down a hill and collides at 5.6 m/s into a stationary sled that is identical to his. The child is launched forward at the same speed, leaving behind the two sleds that lock together and slide forward more slowly. What is the speed of the two sleds after this collision?

Solution:

2.8 m/s

Exercise:**Problem:**

For the preceding problem, find the final speed of each sled for the case of an elastic collision.

Exercise:

Problem:

A 90-kg football player jumps vertically into the air to catch a 0.50-kg football that is thrown essentially horizontally at him at 17 m/s. What is his horizontal speed after catching the ball?

Solution:

0.094 m/s

Exercise:**Problem:**

Three skydivers are plummeting earthward. They are initially holding onto each other, but then push apart. Two skydivers of mass 70 and 80 kg gain horizontal velocities of 1.2 m/s north and 1.4 m/s southeast, respectively. What is the horizontal velocity of the third skydiver, whose mass is 55 kg?

Exercise:**Problem:**

Two billiard balls are at rest and touching each other on a pool table. The cue ball travels at 3.8 m/s along the line of symmetry between these balls and strikes them simultaneously. If the collision is elastic, what is the velocity of the three balls after the collision?

Solution:

final velocity of cue ball is $-(0.76 \text{ m/s})\hat{\mathbf{i}}$, final velocities of the other two balls are 2.6 m/s at $\pm 30^\circ$ with respect to the initial velocity of the cue ball

Exercise:

Problem:

A billiard ball traveling at $(2.2 \text{ m/s})\hat{\mathbf{i}} - (0.4 \text{ m/s})\hat{\mathbf{j}}$ collides with a wall that is aligned in the $\hat{\mathbf{j}}$ direction. Assuming the collision is elastic, what is the final velocity of the ball?

Exercise:**Problem:**

Two identical billiard balls collide. The first one is initially traveling at $(2.2 \text{ m/s})\hat{\mathbf{i}} - (0.4 \text{ m/s})\hat{\mathbf{j}}$ and the second one at $-(1.4 \text{ m/s})\hat{\mathbf{i}} + (2.4 \text{ m/s})\hat{\mathbf{j}}$. Suppose they collide when the center of ball 1 is at the origin and the center of ball 2 is at the point $(2R, 0)$ where R is the radius of the balls. What is the final velocity of each ball?

Solution:

ball 1: $-(1.4 \text{ m/s})\hat{\mathbf{i}} - (0.4 \text{ m/s})\hat{\mathbf{j}}$, ball 2: $(2.2 \text{ m/s})\hat{\mathbf{i}} + (2.4 \text{ m/s})\hat{\mathbf{j}}$

Exercise:**Problem:**

Repeat the preceding problem if the balls collide when the center of ball 1 is at the origin and the center of ball 2 is at the point $(0, 2R)$.

Exercise:**Problem:**

Repeat the preceding problem if the balls collide when the center of ball 1 is at the origin and the center of ball 2 is at the point $(\sqrt{3}R/2, R/2)$

Solution:

ball 1: $(1.4 \text{ m/s})\hat{\mathbf{i}} - (1.7 \text{ m/s})\hat{\mathbf{j}}$, ball 2: $-(2.8 \text{ m/s})\hat{\mathbf{i}} + (0.012 \text{ m/s})\hat{\mathbf{j}}$

Exercise:

Problem:

Where is the center of mass of a semicircular wire of radius R that is centered on the origin, begins and ends on the x axis, and lies in the x,y plane?

Exercise:

Problem:

Where is the center of mass of a slice of pizza that was cut into eight equal slices? Assume the origin is at the apex of the slice and measure angles with respect to an edge of the slice. The radius of the pizza is R .

Solution:

$$(r, \theta) = (2R/3, \pi/8)$$

Exercise:

Problem:

If 1% of the Earth's mass were transferred to the Moon, how far would the center of mass of the Earth-Moon-population system move? The mass of the Earth is $5.97 \times 10^{24} \text{ kg}$ and that of the Moon is $7.34 \times 10^{22} \text{ kg}$. The radius of the Moon's orbit is about $3.84 \times 10^5 \text{ m}$.

Exercise:

Problem:

You friend wonders how a rocket continues to climb into the sky once it is sufficiently high above the surface of Earth so that its expelled gasses no longer push on the surface. How do you respond?

Solution:

Answers may vary. The rocket is propelled forward not by the gasses pushing against the surface of Earth, but by conservation of momentum. The momentum of the gas being expelled out the back of the rocket must be compensated by an increase in the forward momentum of the rocket.

Exercise:

Problem:

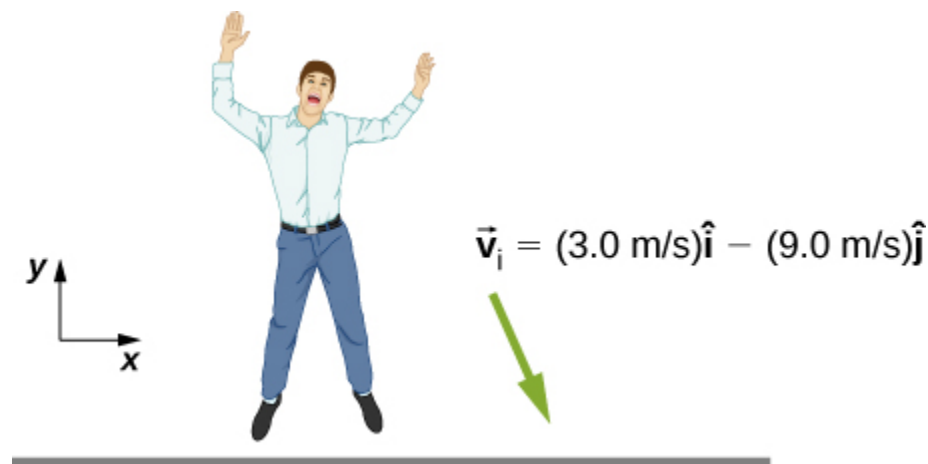
To increase the acceleration of a rocket, should you throw rocks out of the front window of the rocket or out of the back window?

Challenge

Exercise:

Problem:

A 65-kg person jumps from the first floor window of a burning building and lands almost vertically on the ground with a horizontal velocity of 3 m/s and vertical velocity of -9 m/s. Upon impact with the ground he is brought to rest in a short time. The force experienced by his feet depends on whether he keeps his knees stiff or bends them. Find the force on his feet in each case.



- First find the impulse on the person from the impact on the ground. Calculate both its magnitude and direction.

- b. Find the average force on the feet if the person keeps his leg stiff and straight and his center of mass drops by only 1 cm vertically and 1 cm horizontally during the impact.
- c. Find the average force on the feet if the person bends his legs throughout the impact so that his center of mass drops by 50 cm vertically and 5 cm horizontally during the impact.
- d. Compare the results of part (b) and (c), and draw conclusions about which way is better.

You will need to find the time the impact lasts by making reasonable assumptions about the acceleration opposite to the motion. Although the force is not constant during the impact, working with constant average force for this problem is acceptable.

Solution:

a. $617 \text{ N} \cdot \text{s}$, 108° ; b. $F_x = 2.91 \times 10^4 \text{ N}$, $F_y = 2.6 \times 10^5 \text{ N}$; c. $F_x = 5850 \text{ N}$, $F_y = 5265 \text{ N}$

Exercise:

Problem:

Two projectiles of mass m_1 and m_2 are fired at the same speed but in opposite directions from two launch sites separated by a distance D . They both reach the same spot in their highest point and strike there. As a result of the impact they stick together and move as a single body afterwards. Find the place they will land.

Exercise:

Problem:

Two identical objects (such as billiard balls) have a one-dimensional collision in which one is initially motionless. After the collision, the moving object is stationary and the other moves with the same speed as the other originally had. Show that both momentum and kinetic energy are conserved.

Solution:

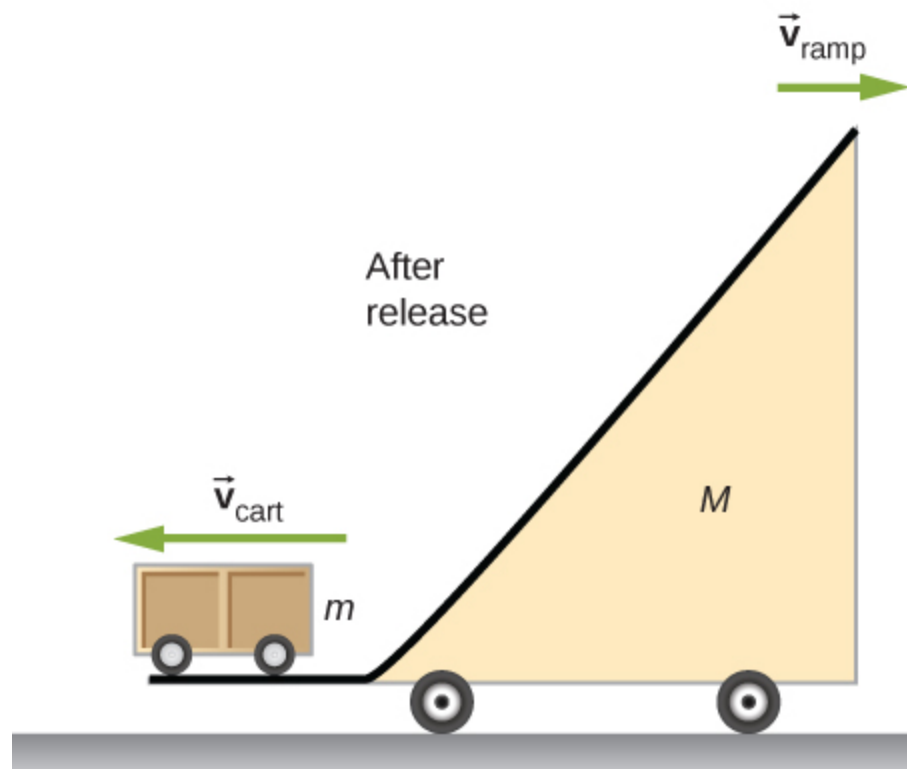
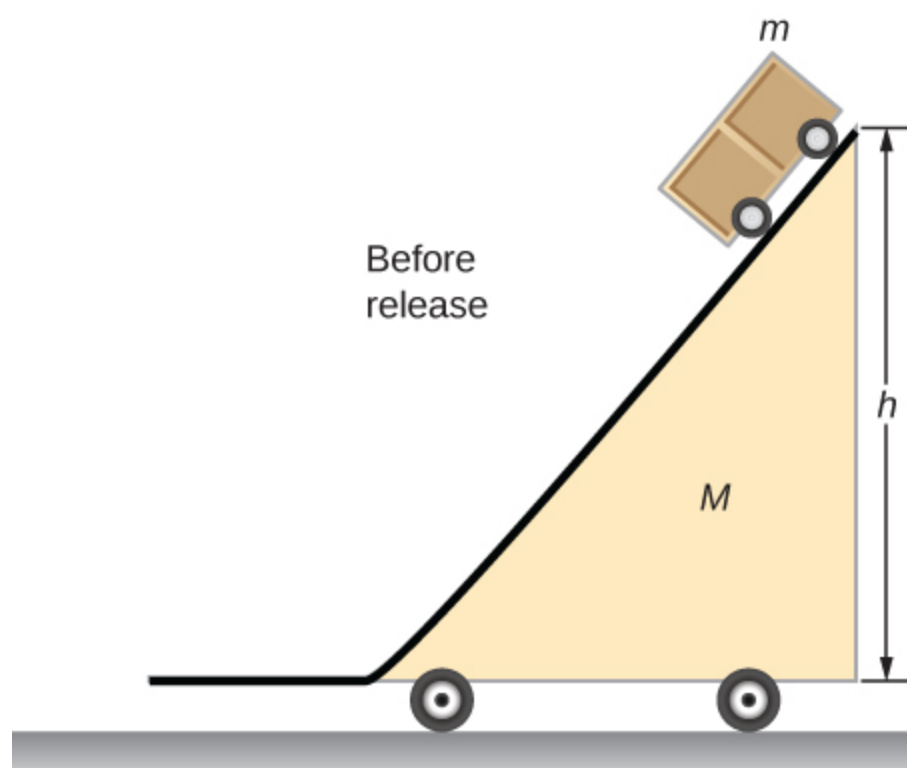
Conservation of momentum demands

$m_1 v_{1,i} + m_2 v_{2,i} = m_1 v_{1,f} + m_2 v_{2,f}$. We are given that $m_1 = m_2$, $v_{1,i} = v_{2,f}$, and $v_{2,i} = v_{1,f} = 0$. Combining these equations with the equation given by conservation of momentum gives $v_{1,i} = v_{1,i}$, which is true, so conservation of momentum is satisfied. Conservation of energy demands $\frac{1}{2} m_1 v_{1,i}^2 + \frac{1}{2} m_2 v_{2,i}^2 = \frac{1}{2} m_1 v_{1,f}^2 + \frac{1}{2} m_2 v_{2,f}^2$. Again combining this equation with the conditions given above give $v_{1,i} = v_{1,i}$, so conservation of energy is satisfied.

Exercise:

Problem:

A ramp of mass M is at rest on a horizontal surface. A small cart of mass m is placed at the top of the ramp and released.

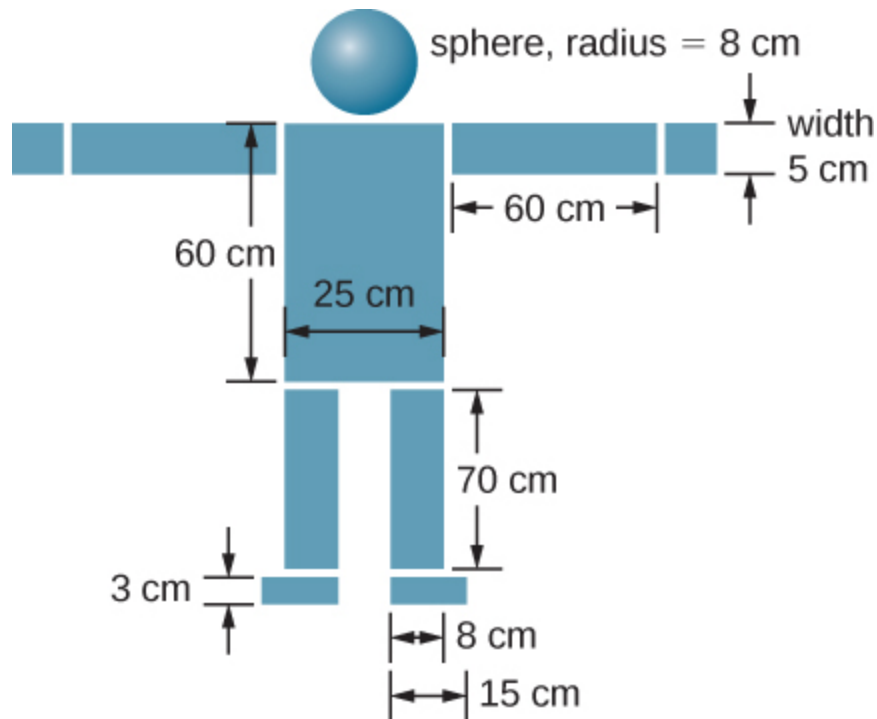


What are the velocities of the ramp and the cart relative to the ground at the instant the cart leaves the ramp?

Exercise:

Problem:

Find the center of mass of the structure given in the figure below. Assume a uniform thickness of 20 cm, and a uniform density of 1 g/cm^3 .



Solution:

Assume origin on centerline and at floor, then
 $(x_{\text{CM}}, y_{\text{CM}}) = (0, 86 \text{ cm})$

Glossary

rocket equation

derived by the Soviet physicist Konstantin Tsiolkovsky in 1897, it gives us the change of velocity that the rocket obtains from burning a mass of fuel that decreases the total rocket mass from m_i down to m

Introduction

class="introduction"

Brazos wind farm in west Texas. As of 2012, wind farms in the US had a power output of 60 gigawatts, enough capacity to power 15 million homes for a year. (credit: modification of work by U.S. Department of Energy)



In previous chapters, we described motion (kinematics) and how to change motion (dynamics), and we defined important concepts such as energy for objects that can be considered as point masses. Point masses, by definition, have no shape and so can only undergo translational motion. However, we know from everyday life that rotational motion is also very important and that many objects that move have both translation and rotation. The wind turbines in our chapter opening image are a prime example of how rotational motion impacts our daily lives, as the market for clean energy sources continues to grow.

We begin to address rotational motion in this chapter, starting with fixed-axis rotation. Fixed-axis rotation describes the rotation around a fixed axis of a rigid body; that is, an object that does not deform as it moves. We will show how to apply all the ideas we've developed up to this point about translational motion to an object rotating around a fixed axis. In the next chapter, we extend these ideas to more complex rotational motion, including objects that both rotate and translate, and objects that do not have a fixed rotational axis.

Rotational Variables

By the end of this section, you will be able to:

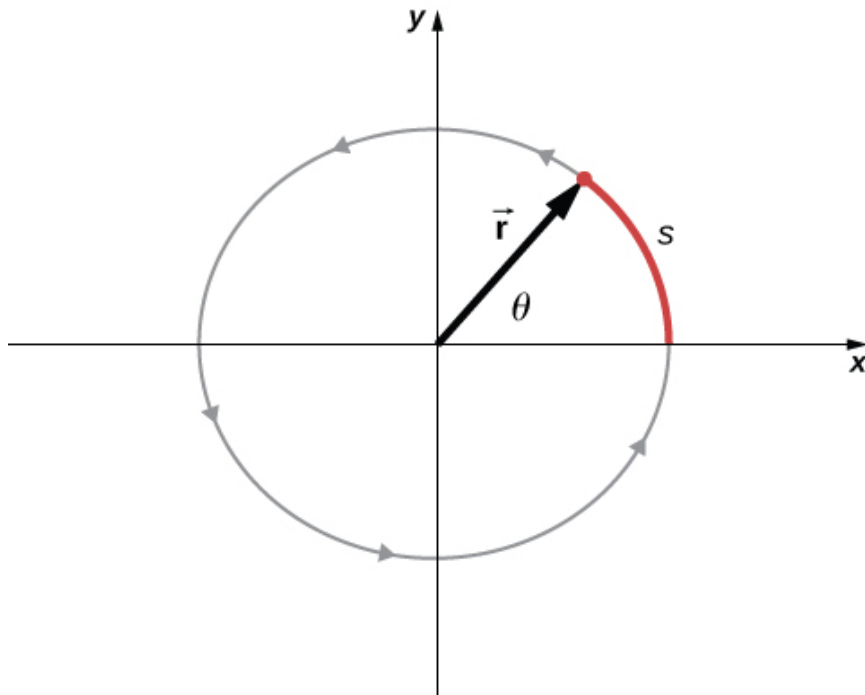
- Describe the physical meaning of rotational variables as applied to fixed-axis rotation
- Explain how angular velocity is related to tangential speed
- Calculate the instantaneous angular velocity given the angular position function
- Find the angular velocity and angular acceleration in a rotating system
- Calculate the average angular acceleration when the angular velocity is changing
- Calculate the instantaneous angular acceleration given the angular velocity function

So far in this text, we have mainly studied translational motion, including the variables that describe it: displacement, velocity, and acceleration. Now we expand our description of motion to rotation—specifically, rotational motion about a fixed axis. We will find that rotational motion is described by a set of related variables similar to those we used in translational motion.

Angular Velocity

Uniform circular motion (discussed previously in [Motion in Two and Three Dimensions](#)) is motion in a circle at constant speed. Although this is the simplest case of rotational motion, it is very useful for many situations, and we use it here to introduce rotational variables.

In [\[link\]](#), we show a particle moving in a circle. The coordinate system is fixed and serves as a frame of reference to define the particle's position. Its position vector from the origin of the circle to the particle sweeps out the angle θ , which increases in the counterclockwise direction as the particle moves along its circular path. The angle θ is called the **angular position** of the particle. As the particle moves in its circular path, it also traces an arc length s .



A particle follows a circular path. As it moves counterclockwise, it sweeps out a positive angle θ with respect to the x -axis and traces out an arc length s .

The angle is related to the radius of the circle and the arc length by

Note:
Equation:

$$\theta = \frac{s}{r}.$$

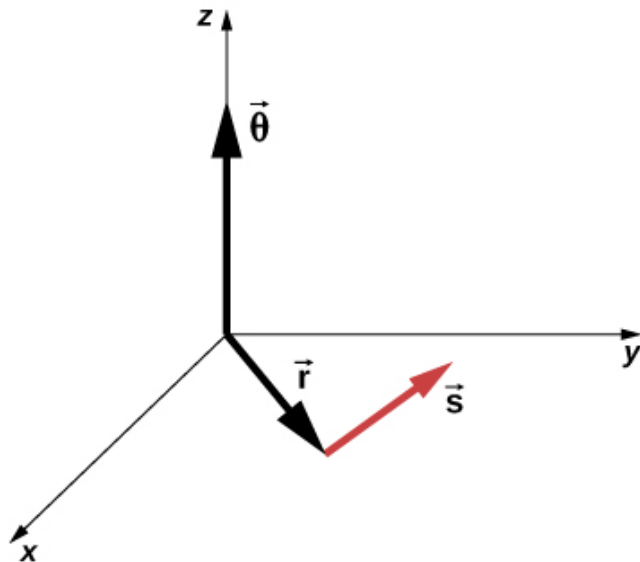
The angle θ , the angular position of the particle along its path, has units of radians (rad). There are 2π radians in 360° . Note that the radian measure is a ratio of length measurements, and therefore is a dimensionless quantity. As the particle moves along its circular path, its angular position changes and it undergoes angular displacements $\Delta\theta$.

We can assign vectors to the quantities in [\[link\]](#). The angle $\vec{\theta}$ is a vector out of the page in [\[link\]](#). The angular position vector \vec{r} and the arc length \vec{s} both lie in the plane of the page. These three vectors are related to each other by

Equation:

$$\vec{s} = \vec{\theta} \times \text{mover}.$$

That is, the arc length is the cross product of the angle vector and the position vector, as shown in [\[link\]](#).



The angle vector points along the z-axis and the position vector and arc length vector both lie in the xy-plane.

We see that $\vec{s} = \vec{\theta} \times \vec{r}$. All three vectors are perpendicular to each other.

The magnitude of the **angular velocity**, denoted by ω , is the time rate of change of the angle θ as the particle moves in its circular path. The **instantaneous angular velocity** is defined as the limit in which $\Delta t \rightarrow 0$ in the average angular velocity $\bar{\omega} = \frac{\Delta\theta}{\Delta t}$:

Note:

Equation:

$$\omega = \lim_{\Delta t \rightarrow 0} \frac{\Delta \theta}{\Delta t} = \frac{d\theta}{dt},$$

where θ is the angle of rotation ([\[link\]](#)). The units of angular velocity are radians per second (rad/s). Angular velocity can also be referred to as the rotation rate in radians per second. In many situations, we are given the rotation rate in revolutions/s or cycles/s. To find the angular velocity, we must multiply revolutions/s by 2π , since there are 2π radians in one complete revolution. Since the direction of a positive angle in a circle is counterclockwise, we take counterclockwise rotations as being positive and clockwise rotations as negative.

We can see how angular velocity is related to the tangential speed of the particle by differentiating [\[link\]](#) with respect to time. We rewrite [\[link\]](#) as

Equation:

$$s = r\theta.$$

Taking the derivative with respect to time and noting that the radius r is a constant, we have

Equation:

$$\frac{ds}{dt} = \frac{d}{dt}(r\theta) = r \frac{d\theta}{dt} = r \omega$$

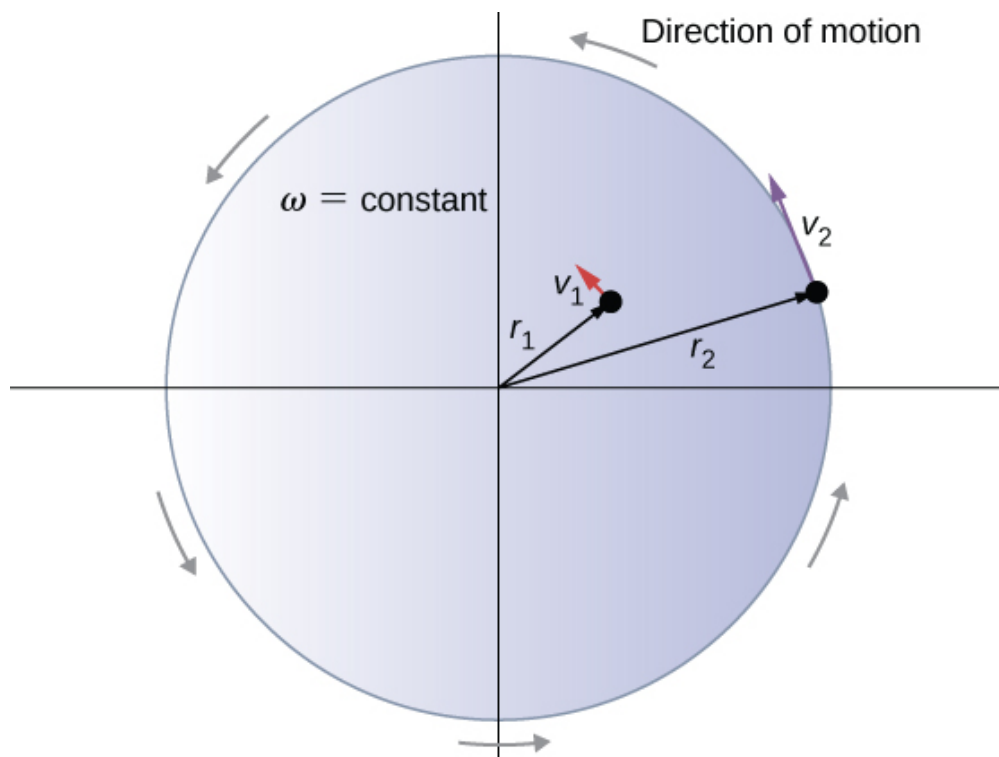
where $\theta \frac{dr}{dt} = 0$. Here $\frac{ds}{dt}$ is just the tangential speed v_t of the particle in [\[link\]](#). Thus, by using [\[link\]](#), we arrive at

Note:

Equation:

$$v_t = r\omega.$$

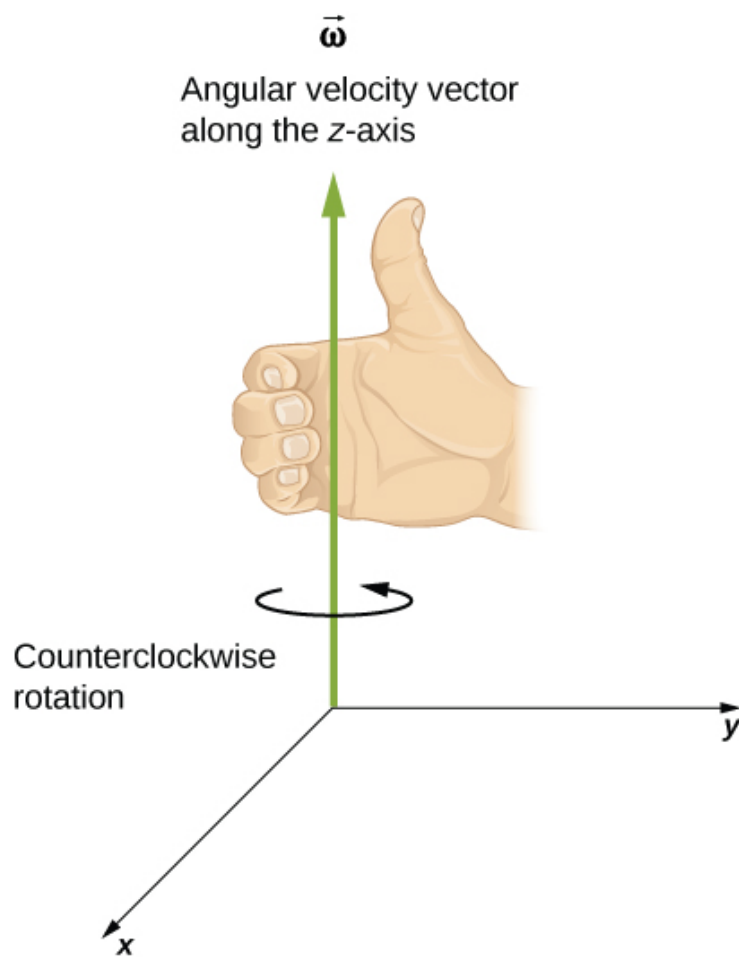
That is, the tangential speed of the particle is its angular velocity times the radius of the circle. From [\[link\]](#), we see that the tangential speed of the particle increases with its distance from the axis of rotation for a constant angular velocity. This effect is shown in [\[link\]](#). Two particles are placed at different radii on a rotating disk with a constant angular velocity. As the disk rotates, the tangential speed increases linearly with the radius from the axis of rotation. In [\[link\]](#), we see that $v_1 = r_1\omega_1$ and $v_2 = r_2\omega_2$. But the disk has a constant angular velocity, so $\omega_1 = \omega_2$. This means $\frac{v_1}{r_1} = \frac{v_2}{r_2}$ or $v_2 = \left(\frac{r_2}{r_1}\right)v_1$. Thus, since $r_2 > r_1$, $v_2 > v_1$.



Two particles on a rotating disk have different tangential speeds, depending on their distance to the axis of rotation.

Up until now, we have discussed the magnitude of the angular velocity $\omega = d\theta/dt$, which is a scalar quantity—the change in angular position with respect to time. The vector $\vec{\omega}$ is the vector associated with the angular velocity and points along the axis of rotation. This is useful because when a rigid body is rotating, we want to know both the axis of rotation and the direction that the body is rotating about the axis,

clockwise or counterclockwise. The angular velocity $\vec{\omega}$ gives us this information. The angular velocity $\vec{\omega}$ has a direction determined by what is called the right-hand rule. The right-hand rule is such that if the fingers of your right hand wrap counterclockwise from the x-axis (the direction in which θ increases) toward the y-axis, your thumb points in the direction of the positive z-axis ([link](#)). An angular velocity $\vec{\omega}$ that points along the positive z-axis therefore corresponds to a counterclockwise rotation, whereas an angular velocity $\vec{\omega}$ that points along the negative z-axis corresponds to a clockwise rotation.



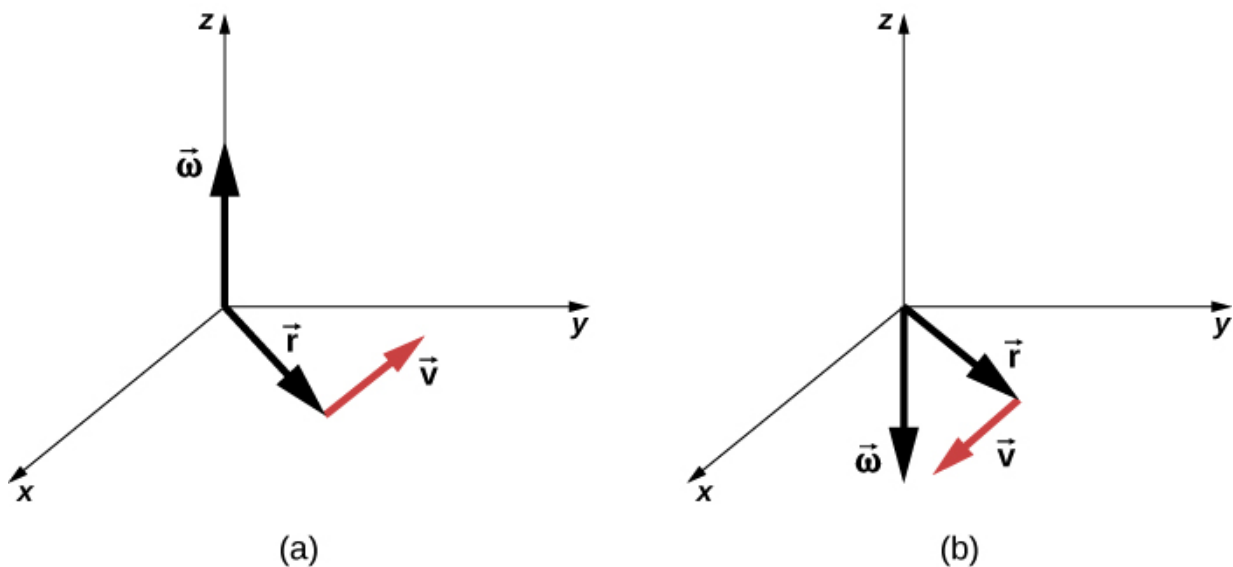
For counterclockwise rotation in the coordinate system shown, the angular velocity points in the positive z-direction by the right-hand-rule.

Similar to [\[link\]](#), one can state a cross product relation to the vector of the tangential velocity as stated in [\[link\]](#). Therefore, we have

Equation:

$$\vec{v} = \vec{\omega} \times \vec{r}.$$

That is, the tangential velocity is the cross product of the angular velocity and the position vector, as shown in [\[link\]](#). From part (a) of this figure, we see that with the angular velocity in the positive z-direction, the rotation in the xy-plane is counterclockwise. In part (b), the angular velocity is in the negative z-direction, giving a clockwise rotation in the xy-plane.



The vectors shown are the angular velocity, position, and tangential velocity. (a) The angular velocity points in the positive z-direction, giving a counterclockwise rotation in the xy-plane. (b) The angular velocity points in the negative z-direction, giving a clockwise rotation.

Example:
Rotation of a Flywheel

A flywheel rotates such that it sweeps out an angle at the rate of $\theta = \omega t = (45.0 \text{ rad/s})t$ radians. The wheel rotates counterclockwise when viewed in the plane of the page. (a) What is the angular velocity of the flywheel? (b) What direction is the angular velocity? (c) How many radians does the flywheel rotate through in 30 s? (d) What is the tangential speed of a point on the flywheel 10 cm from the axis of rotation?

Strategy

The functional form of the angular position of the flywheel is given in the problem as $\theta(t) = \omega t$, so by taking the derivative with respect to time, we can find the angular velocity. We use the right-hand rule to find the angular velocity. To find the angular displacement of the flywheel during 30 s, we seek the angular displacement $\Delta\theta$, where the change in angular position is between 0 and 30 s. To find the tangential speed of a point at a distance from the axis of rotation, we multiply its distance times the angular velocity of the flywheel.

Solution

- $\omega = \frac{d\theta}{dt} = 45 \text{ rad/s}$. We see that the angular velocity is a constant.
- By the right-hand rule, we curl the fingers in the direction of rotation, which is counterclockwise in the plane of the page, and the thumb points in the direction of the angular velocity, which is out of the page.
- $\Delta\theta = \theta(30 \text{ s}) - \theta(0 \text{ s}) = 45.0(30.0 \text{ s}) - 45.0(0 \text{ s}) = 1350.0 \text{ rad}$.
- $v_t = r\omega = (0.1 \text{ m})(45.0 \text{ rad/s}) = 4.5 \text{ m/s}$.

Significance

In 30 s, the flywheel has rotated through quite a number of revolutions, about 215 if we divide the angular displacement by 2π . A massive flywheel can be used to store energy in this way, if the losses due to friction are minimal. Recent research has considered superconducting bearings on which the flywheel rests, with zero energy loss due to friction.

Angular Acceleration

We have just discussed angular velocity for uniform circular motion, but not all motion is uniform. Envision an ice skater spinning with his arms outstretched—when he pulls his arms inward, his angular velocity increases. Or think about a computer's hard disk slowing to a halt as the angular velocity decreases. We will explore these situations later, but we can already see a need to define an **angular acceleration** for describing situations where ω changes. The faster the change in ω , the greater the

angular acceleration. We define the **instantaneous angular acceleration** α as the derivative of angular velocity with respect to time:

Note:

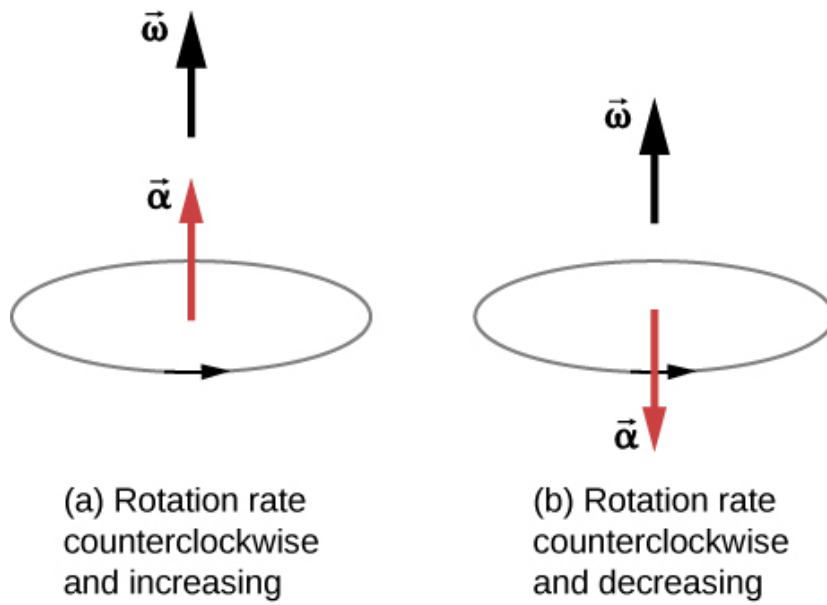
Equation:

$$\alpha = \lim_{\Delta t \rightarrow 0} \frac{\Delta \omega}{\Delta t} = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2},$$

where we have taken the limit of the average angular acceleration, $\bar{\alpha} = \frac{\Delta \omega}{\Delta t}$ as $\Delta t \rightarrow 0$.

The units of angular acceleration are (rad/s)/s, or rad/s^2 .

In the same way as we defined the vector associated with angular velocity $\vec{\omega}$, we can define $\vec{\alpha}$, the vector associated with angular acceleration ([\[link\]](#)). If the angular velocity is along the positive z-axis, as in [\[link\]](#), and $\frac{d\omega}{dt}$ is positive, then the angular acceleration $\vec{\alpha}$ is positive and points along the $+z$ -axis. Similarly, if the angular velocity $\vec{\omega}$ is along the positive z-axis and $\frac{d\omega}{dt}$ is negative, then the angular acceleration is negative and points along the $+z$ -axis.



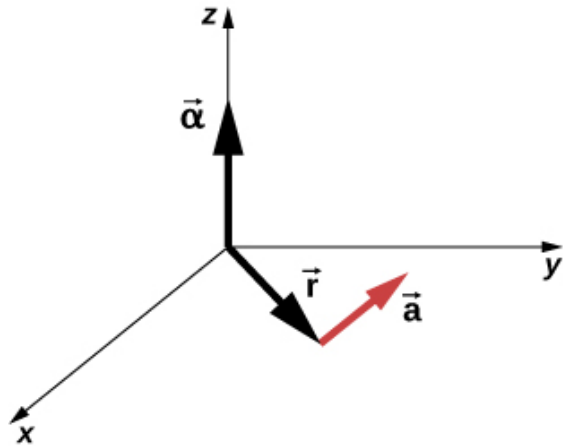
The rotation is counterclockwise in both (a) and (b) with the angular velocity in the same direction. (a) The angular acceleration is in the same direction as the angular velocity, which increases the rotation rate. (b) The angular acceleration is in the opposite direction to the angular velocity, which decreases the rotation rate.

We can express the tangential acceleration vector as a cross product of the angular acceleration and the position vector. This expression can be found by taking the time derivative of $\vec{v} = \vec{\omega} \times \vec{r}$ and is left as an exercise:

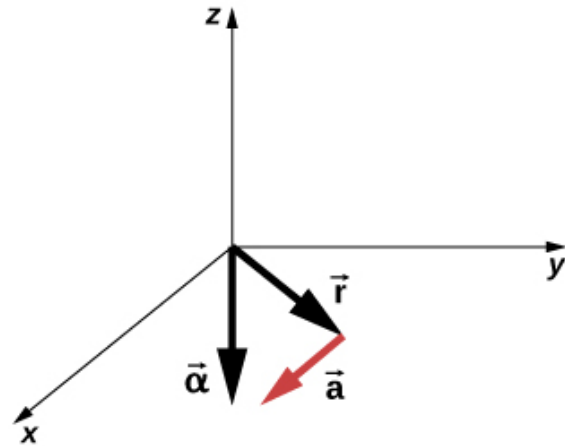
Equation:

$$\vec{a} = \vec{\alpha} \times \vec{r}.$$

The vector relationships for the angular acceleration and tangential acceleration are shown in [\[link\]](#).



(a)



(b)

(a) The angular acceleration is the positive z -direction and produces a tangential acceleration in a counterclockwise sense. (b) The angular acceleration is in the negative z -direction and produces a tangential acceleration in the clockwise sense.

We can relate the tangential acceleration of a point on a rotating body at a distance from the axis of rotation in the same way that we related the tangential speed to the angular velocity. If we differentiate [\[link\]](#) with respect to time, noting that the radius r is constant, we obtain

Note:

Equation:

$$a_t = r\alpha.$$

Thus, the tangential acceleration a_t is the radius times the angular acceleration. [\[link\]](#) and [\[link\]](#) are important for the discussion of rolling motion (see [Angular Momentum](#)).

Let's apply these ideas to the analysis of a few simple fixed-axis rotation scenarios. Before doing so, we present a problem-solving strategy that can be applied to rotational kinematics: the description of rotational motion.

Note:**Rotational Kinematics**

1. Examine the situation to determine that rotational kinematics (rotational motion) is involved.
2. Identify exactly what needs to be determined in the problem (identify the unknowns). A sketch of the situation is useful.
3. Make a complete list of what is given or can be inferred from the problem as stated (identify the knowns).
4. Solve the appropriate equation or equations for the quantity to be determined (the unknown). It can be useful to think in terms of a translational analog, because by now you are familiar with the equations of translational motion.
5. Substitute the known values along with their units into the appropriate equation and obtain numerical solutions complete with units. Be sure to use units of radians for angles.
6. Check your answer to see if it is reasonable: Does your answer make sense?

Now let's apply this problem-solving strategy to a few specific examples.

Example:**A Spinning Bicycle Wheel**

A bicycle mechanic mounts a bicycle on the repair stand and starts the rear wheel spinning from rest to a final angular velocity of 250 rpm in 5.00 s. (a) Calculate the average angular acceleration in rad/s^2 . (b) If she now hits the brakes, causing an angular acceleration of -87.3 rad/s^2 , how long does it take the wheel to stop?

Strategy

The average angular acceleration can be found directly from its definition $\bar{\alpha} = \frac{\Delta\omega}{\Delta t}$ because the final angular velocity and time are given. We see that

$\Delta\omega = \omega_{\text{final}} - \omega_{\text{initial}} = 250 \text{ rev/min}$ and Δt is 5.00 s. For part (b), we know the angular acceleration and the initial angular velocity. We can find the stopping time by using the definition of average angular acceleration and solving for Δt , yielding

Equation:

$$\Delta t = \frac{\Delta\omega}{\alpha}.$$

Solution

- a. Entering known information into the definition of angular acceleration, we get
Equation:

$$\bar{\alpha} = \frac{\Delta\omega}{\Delta t} = \frac{250 \text{ rpm}}{5.00 \text{ s}}.$$

Because $\Delta\omega$ is in revolutions per minute (rpm) and we want the standard units of rad/s^2 for angular acceleration, we need to convert from rpm to rad/s:

Equation:

$$\Delta\omega = 250 \frac{\text{rev}}{\text{min}} \cdot \frac{2\pi \text{ rad}}{\text{rev}} \cdot \frac{1 \text{ min}}{60 \text{ s}} = 26.2 \frac{\text{rad}}{\text{s}}.$$

Entering this quantity into the expression for α , we get

Equation:

$$\alpha = \frac{\Delta\omega}{\Delta t} = \frac{26.2 \text{ rad/s}}{5.00 \text{ s}} = 5.24 \text{ rad/s}^2.$$

- b. Here the angular velocity decreases from 26.2 rad/s (250 rpm) to zero, so that $\Delta\omega$ is -26.2 rad/s , and α is given to be -87.3 rad/s^2 . Thus,

Equation:

$$\Delta t = \frac{-26.2 \text{ rad/s}}{-87.3 \text{ rad/s}^2} = 0.300 \text{ s}.$$

Significance

Note that the angular acceleration as the mechanic spins the wheel is small and positive; it takes 5 s to produce an appreciable angular velocity. When she hits the brake, the angular acceleration is large and negative. The angular velocity quickly goes to zero.

Note:

Exercise:

Problem:

Check Your Understanding The fan blades on a turbofan jet engine (shown below) accelerate from rest up to a rotation rate of 40.0 rev/s in 20 s. The increase in angular velocity of the fan is constant in time. (The GE90-110B1 turbofan engine mounted on a Boeing 777, as shown, is currently the largest turbofan engine in the world, capable of thrusts of 330–510 kN.)

- (a) What is the average angular acceleration?
- (b) What is the instantaneous angular acceleration at any time during the first 20 s?



(credit: “Bubinator”/ Wikimedia Commons)

Solution:

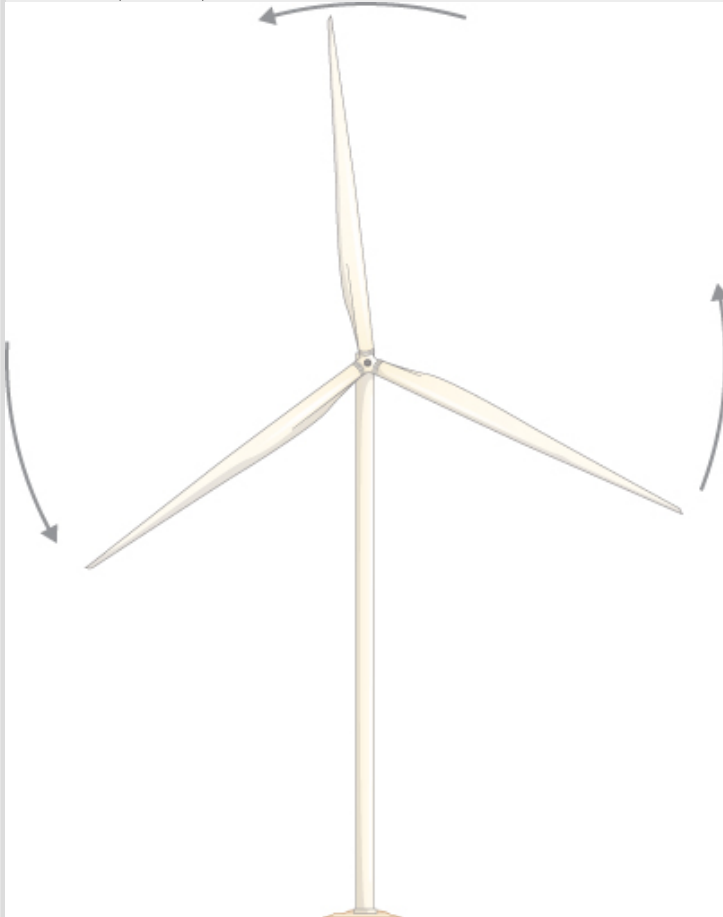
a. $40.0 \text{ rev/s} = 2\pi(40.0) \text{ rad/s}$,
 $\bar{\alpha} = \frac{\Delta\omega}{\Delta t} = \frac{2\pi(40.0) - 0 \text{ rad/s}}{20.0 \text{ s}} = 2\pi(2.0) = 4.0\pi \text{ rad/s}^2$; b. Since the angular velocity increases linearly, there has to be a constant acceleration throughout

the indicated time. Therefore, the instantaneous angular acceleration at any time is the solution to $4.0\pi \text{ rad/s}^2$.

Example:

Wind Turbine

A wind turbine ([link](#)) in a wind farm is being shut down for maintenance. It takes 30 s for the turbine to go from its operating angular velocity to a complete stop in which the angular velocity function is $\omega(t) = [(ts^{-1} - 30.0)^2 / 100.0] \text{ rad/s}$. If the turbine is rotating counterclockwise looking into the page, (a) what are the directions of the angular velocity and acceleration vectors? (b) What is the average angular acceleration? (c) What is the instantaneous angular acceleration at $t = 0.0, 15.0, 30.0 \text{ s}$?



A wind turbine that is rotating counterclockwise, as seen head on.

Strategy

- We are given the rotational sense of the turbine, which is counterclockwise in the plane of the page. Using the right hand rule ([\[link\]](#)), we can establish the directions of the angular velocity and acceleration vectors.
- We calculate the initial and final angular velocities to get the average angular acceleration. We establish the sign of the angular acceleration from the results in (a).
- We are given the functional form of the angular velocity, so we can find the functional form of the angular acceleration function by taking its derivative with respect to time.

Solution

- Since the turbine is rotating counterclockwise, angular velocity $\vec{\omega}$ points out of the page. But since the angular velocity is decreasing, the angular acceleration $\vec{\alpha}$ points into the page, in the opposite sense to the angular velocity.
- The initial angular velocity of the turbine, setting $t = 0$, is $\omega = 9.0 \text{ rad/s}$. The final angular velocity is zero, so the average angular acceleration is

Equation:

$$\bar{\alpha} = \frac{\Delta\omega}{\Delta t} = \frac{\omega - \omega_0}{t - t_0} = \frac{0 - 9.0 \text{ rad/s}}{30.0 - 0 \text{ s}} = -0.3 \text{ rad/s}^2.$$

- Taking the derivative of the angular velocity with respect to time gives

$$\alpha = \frac{d\omega}{dt} = (t - 30.0)/50.0 \text{ rad/s}^2$$

Equation:

$$\alpha(0.0 \text{ s}) = -0.6 \text{ rad/s}^2, \alpha(15.0 \text{ s}) = -0.3 \text{ rad/s}^2, \text{ and } \alpha(30.0 \text{ s}) = 0 \text{ rad/s}^2.$$

Significance

We found from the calculations in (a) and (b) that the angular acceleration α and the average angular acceleration $\bar{\alpha}$ are negative. The turbine has an angular acceleration in the opposite sense to its angular velocity.

We now have a basic vocabulary for discussing fixed-axis rotational kinematics and relationships between rotational variables. We discuss more definitions and connections in the next section.

Summary

- The angular position θ of a rotating body is the angle the body has rotated through in a fixed coordinate system, which serves as a frame of reference.
- The angular velocity of a rotating body about a fixed axis is defined as ω (rad/s), the rotational rate of the body in radians per second. The instantaneous angular velocity of a rotating body $\omega = \lim_{\Delta t \rightarrow 0} \frac{\Delta \theta}{\Delta t} = \frac{d\theta}{dt}$ is the derivative with respect to time of the angular position θ , found by taking the limit $\Delta t \rightarrow 0$ in the average angular velocity $\bar{\omega} = \frac{\Delta \theta}{\Delta t}$. The angular velocity relates v_t to the tangential speed of a point on the rotating body through the relation $v_t = r\omega$, where r is the radius to the point and v_t is the tangential speed at the given point.
- The angular velocity $\vec{\omega}$ is found using the right-hand rule. If the fingers curl in the direction of rotation about a fixed axis, the thumb points in the direction of $\vec{\omega}$ (see [link](#)).
- If the system's angular velocity is not constant, then the system has an angular acceleration. The average angular acceleration over a given time interval is the change in angular velocity over this time interval, $\bar{\alpha} = \frac{\Delta \omega}{\Delta t}$. The instantaneous angular acceleration is the time derivative of angular velocity, $\alpha = \lim_{\Delta t \rightarrow 0} \frac{\Delta \omega}{\Delta t} = \frac{d\omega}{dt}$. The angular acceleration $\vec{\alpha}$ is found by locating the angular velocity. If a rotation rate of a rotating body is decreasing, the angular acceleration is in the opposite direction to $\vec{\omega}$. If the rotation rate is increasing, the angular acceleration is in the same direction as $\vec{\omega}$.
- The tangential acceleration of a point at a radius from the axis of rotation is the angular acceleration times the radius to the point.

Conceptual Questions

Exercise:

Problem:

A clock is mounted on the wall. As you look at it, what is the direction of the angular velocity vector of the second hand?

Solution:

The second hand rotates clockwise, so by the right-hand rule, the angular velocity vector is into the wall.

Exercise:**Problem:**

What is the value of the angular acceleration of the second hand of the clock on the wall?

Exercise:**Problem:**

A baseball bat is swung. Do all points on the bat have the same angular velocity? The same tangential speed?

Solution:

They have the same angular velocity. Points further out on the bat have greater tangential speeds.

Exercise:**Problem:**

The blades of a blender on a counter are rotating clockwise as you look into it from the top. If the blender is put to a greater speed what direction is the angular acceleration of the blades?

Problems**Exercise:**

Problem: Calculate the angular velocity of Earth.

Exercise:**Problem:**

A track star runs a 400-m race on a 400-m circular track in 45 s. What is his angular velocity assuming a constant speed?

Solution:

$$\omega = \frac{2\pi \text{ rad}}{45.0 \text{ s}} = 0.14 \text{ rad/s}$$

Exercise:

Problem:

A wheel rotates at a constant rate of $2.0 \times 10^3 \text{ rev/min}$. (a) What is its angular velocity in radians per second? (b) Through what angle does it turn in 10 s? Express the solution in radians and degrees.

Exercise:**Problem:**

A particle moves 3.0 m along a circle of radius 1.5 m. (a) Through what angle does it rotate? (b) If the particle makes this trip in 1.0 s at a constant speed, what is its angular velocity? (c) What is its acceleration?

Solution:

$$\text{a. } \theta = \frac{s}{r} = \frac{3.0 \text{ m}}{1.5 \text{ m}} = 2.0 \text{ rad}; \text{ b. } \omega = \frac{2.0 \text{ rad}}{1.0 \text{ s}} = 2.0 \text{ rad/s}; \text{ c. } \frac{v^2}{r} = \frac{(3.0 \text{ m/s})^2}{1.5 \text{ m}} = 6.0 \text{ m/s}^2.$$

Exercise:**Problem:**

A compact disc rotates at 500 rev/min. If the diameter of the disc is 120 mm, (a) what is the tangential speed of a point at the edge of the disc? (b) At a point halfway to the center of the disc?

Exercise:**Problem:**

Unreasonable results. The propeller of an aircraft is spinning at 10 rev/s when the pilot shuts off the engine. The propeller reduces its angular velocity at a constant 2.0 rad/s^2 for a time period of 40 s. What is the rotation rate of the propeller in 40 s? Is this a reasonable situation?

Solution:

The propeller takes only $\Delta t = \frac{\Delta\omega}{\alpha} = \frac{0 \text{ rad/s} - 10.0(2\pi) \text{ rad/s}}{-2.0 \text{ rad/s}^2} = 31.4 \text{ s}$ to come to rest, when the propeller is at 0 rad/s, it would start rotating in the opposite direction. This would be impossible due to the magnitude of forces involved in getting the propeller to stop and start rotating in the opposite direction.

Exercise:

Problem:

A gyroscope slows from an initial rate of 32.0 rad/s at a rate of 0.700 rad/s^2 . How long does it take to come to rest?

Exercise:**Problem:**

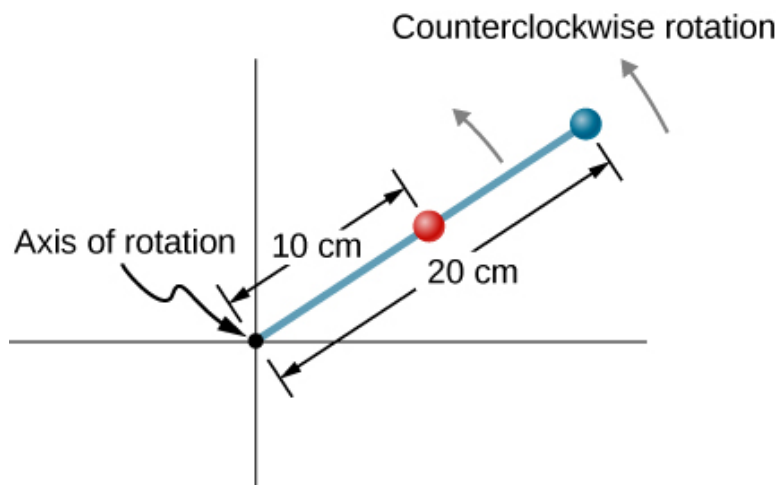
On takeoff, the propellers on a UAV (unmanned aerial vehicle) increase their angular velocity for 3.0 s from rest at a rate of $\omega = (25.0t) \text{ rad/s}$ where t is measured in seconds. (a) What is the instantaneous angular velocity of the propellers at $t = 2.0 \text{ s}$? (b) What is the angular acceleration?

Solution:

a. $\omega = 25.0(2.0 \text{ s}) = 50.0 \text{ rad/s}$; b. $\alpha = \frac{d\omega}{dt} = 25.0 \text{ rad/s}^2$

Exercise:**Problem:**

The angular position of a rod varies as $20.0t^2$ radians from time $t = 0$. The rod has two beads on it as shown in the following figure, one at 10 cm from the rotation axis and the other at 20 cm from the rotation axis. (a) What is the instantaneous angular velocity of the rod at $t = 5 \text{ s}$? (b) What is the angular acceleration of the rod? (c) What are the tangential speeds of the beads at $t = 5 \text{ s}$? (d) What are the tangential accelerations of the beads at $t = 5 \text{ s}$? (e) What are the centripetal accelerations of the beads at $t = 5 \text{ s}$?



Glossary

angular acceleration

time rate of change of angular velocity

angular position

angle a body has rotated through in a fixed coordinate system

angular velocity

time rate of change of angular position

instantaneous angular acceleration

derivative of angular velocity with respect to time

instantaneous angular velocity

derivative of angular position with respect to time

Rotation with Constant Angular Acceleration

By the end of this section, you will be able to:

- Derive the kinematic equations for rotational motion with constant angular acceleration
- Select from the kinematic equations for rotational motion with constant angular acceleration the appropriate equations to solve for unknowns in the analysis of systems undergoing fixed-axis rotation
- Use solutions found with the kinematic equations to verify the graphical analysis of fixed-axis rotation with constant angular acceleration

In the preceding section, we defined the rotational variables of angular displacement, angular velocity, and angular acceleration. In this section, we work with these definitions to derive relationships among these variables and use these relationships to analyze rotational motion for a rigid body about a fixed axis under a constant angular acceleration. This analysis forms the basis for rotational kinematics. If the angular acceleration is constant, the equations of rotational kinematics simplify, similar to the equations of linear kinematics discussed in [Motion along a Straight Line](#) and [Motion in Two and Three Dimensions](#). We can then use this simplified set of equations to describe many applications in physics and engineering where the angular acceleration of the system is constant. Rotational kinematics is also a prerequisite to the discussion of rotational dynamics later in this chapter.

Kinematics of Rotational Motion

Using our intuition, we can begin to see how the rotational quantities θ , ω , α , and t are related to one another. For example, we saw in the preceding section that if a flywheel has an angular acceleration in the same direction as its angular velocity vector, its angular velocity increases with time and its angular displacement also increases. On the contrary, if the angular acceleration is opposite to the angular velocity vector, its angular velocity decreases with time. We can describe these physical situations and many others with a consistent set of rotational kinematic equations under a constant angular acceleration. The method to investigate rotational motion in this way is called **kinematics of rotational motion**.

To begin, we note that if the system is rotating under a constant acceleration, then the average angular velocity follows a simple relation because the angular velocity is increasing linearly with time. The average angular velocity is just half the sum of the initial and final values:

Note:
Equation:

$$\bar{\omega} = \frac{\omega_0 + \omega_f}{2}.$$

From the definition of the average angular velocity, we can find an equation that relates the angular position, average angular velocity, and time:

Equation:

$$\bar{\omega} = \frac{\Delta\theta}{\Delta t}.$$

Solving for θ , we have

Note:

Equation:

$$\theta_f = \theta_0 + \bar{\omega}t,$$

where we have set $t_0 = 0$. This equation can be very useful if we know the average angular velocity of the system. Then we could find the angular displacement over a given time period. Next, we find an equation relating ω , α , and t . To determine this equation, we start with the definition of angular acceleration:

Equation:

$$\alpha = \frac{d\omega}{dt}.$$

We rearrange this to get $\alpha dt = d\omega$ and then we integrate both sides of this equation from initial values to final values, that is, from t_0 to t and ω_0 to ω_f . In uniform rotational motion, the angular acceleration is constant so it can be pulled out of the integral, yielding two definite integrals:

Equation:

$$\alpha \int_{t_0}^t dt' = \int_{\omega_0}^{\omega_f} d\omega.$$

Setting $t_0 = 0$, we have

Equation:

$$\alpha t = \omega_f - \omega_0.$$

We rearrange this to obtain

Note:
Equation:

$$\omega_f = \omega_0 + \alpha t,$$

where ω_0 is the initial angular velocity. [\[link\]](#) is the rotational counterpart to the linear kinematics equation $v_f = v_0 + at$. With [\[link\]](#), we can find the angular velocity of an object at any specified time t given the initial angular velocity and the angular acceleration.

Let's now do a similar treatment starting with the equation $\omega = \frac{d\theta}{dt}$. We rearrange it to obtain $\omega dt = d\theta$ and integrate both sides from initial to final values again, noting that the angular acceleration is constant and does not have a time dependence. However, this time, the angular velocity is not constant (in general), so we substitute in what we derived above:

Equation:

$$\int_{t_0}^{t_f} (\omega_0 + \alpha t') dt' = \int_{\theta_0}^{\theta_f} d\theta;$$
$$\int_{t_0}^t \omega_0 dt + \int_{t_0}^t \alpha t' dt' = \int_{\theta_0}^{\theta_f} d\theta = \left[\omega_0 t' + \alpha \left(\frac{(t')^2}{2} \right) \right]_{t_0}^t = \omega_0 t + \alpha \left(\frac{t^2}{2} \right) = \theta_f - \theta_0,$$

where we have set $t_0 = 0$. Now we rearrange to obtain

Note:
Equation:

$$\theta_f = \theta_0 + \omega_0 t + \frac{1}{2} \alpha t^2.$$

[\[link\]](#) is the rotational counterpart to the linear kinematics equation found in [Motion Along a Straight Line](#) for position as a function of time. This equation gives us the angular position of a rotating rigid body at any time t given the initial conditions (initial angular position and initial angular velocity) and the angular acceleration.

We can find an equation that is independent of time by solving for t in [\[link\]](#) and substituting into [\[link\]](#). [\[link\]](#) becomes

Equation:

$$\begin{aligned}
\theta_f &= \theta_0 + \omega_0 \left(\frac{\omega_f - \omega_0}{\alpha} \right) + \frac{1}{2} \alpha \left(\frac{\omega_f - \omega_0}{\alpha} \right)^2 \\
&= \theta_0 + \frac{\omega_0 \omega_f}{\alpha} - \frac{\omega_0^2}{\alpha} + \frac{1}{2} \frac{\omega_f^2}{\alpha} - \frac{\omega_0 \omega_f}{\alpha} + \frac{1}{2} \frac{\omega_0^2}{\alpha} \\
&= \theta_0 + \frac{1}{2} \frac{\omega_f^2}{\alpha} - \frac{1}{2} \frac{\omega_0^2}{\alpha}, \\
\theta_f - \theta_0 &= \frac{\omega_f^2 - \omega_0^2}{2\alpha}
\end{aligned}$$

or

Note:**Equation:**

$$\omega_f^2 = \omega_0^2 + 2\alpha(\Delta\theta).$$

[\[link\]](#) through [\[link\]](#) describe fixed-axis rotation for constant acceleration and are summarized in [\[link\]](#).

Angular displacement from average angular velocity	$\theta_f = \theta_0 + \bar{\omega}t$
Angular velocity from angular acceleration	$\omega_f = \omega_0 + \alpha t$
Angular displacement from angular velocity and angular acceleration	$\theta_f = \theta_0 + \omega_0 t + \frac{1}{2} \alpha t^2$
Angular velocity from angular displacement and angular acceleration	$\omega_f^2 = \omega_0^2 + 2\alpha(\Delta\theta)$

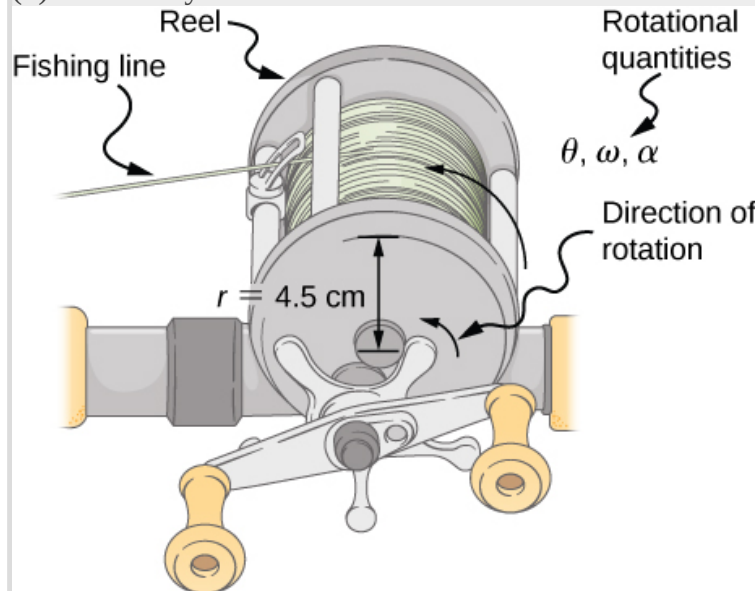
Kinematic Equations**Applying the Equations for Rotational Motion**

Now we can apply the key kinematic relations for rotational motion to some simple examples to get a feel for how the equations can be applied to everyday situations.

Example:**Calculating the Acceleration of a Fishing Reel**

A deep-sea fisherman hooks a big fish that swims away from the boat, pulling the fishing line from his fishing reel. The whole system is initially at rest, and the fishing line unwinds from the reel at a radius of 4.50 cm from its axis of rotation. The reel is given an angular acceleration of 110 rad/s^2 for 2.00 s ([link](#)).

- (a) What is the final angular velocity of the reel after 2 s?
(b) How many revolutions does the reel make?



Fishing line coming off a rotating reel moves linearly.

Strategy

Identify the knowns and compare with the kinematic equations for constant acceleration. Look for the appropriate equation that can be solved for the unknown, using the knowns given in the problem description.

Solution

- a. We are given α and t and want to determine ω . The most straightforward equation to use is $\omega_f = \omega_0 + \alpha t$, since all terms are known besides the unknown variable we are looking for. We are given that $\omega_0 = 0$ (it starts from rest), so

Equation:

$$\omega_f = 0 + (110 \text{ rad/s}^2)(2.00 \text{ s}) = 220 \text{ rad/s}.$$

- b. We are asked to find the number of revolutions. Because $1 \text{ rev} = 2\pi \text{ rad}$, we can find the number of revolutions by finding θ in radians. We are given α and t , and we know ω_0 is zero, so we can obtain θ by using

Equation:

$$\begin{aligned}\theta_f &= \theta_i + \omega_i t + \frac{1}{2} \alpha t^2 \\ &= 0 + 0 + (0.500) (110 \text{ rad/s}^2) (2.00 \text{ s})^2 = 220 \text{ rad}.\end{aligned}$$

Converting radians to revolutions gives

Equation:

$$\text{Number of rev} = (220 \text{ rad}) \frac{1 \text{ rev}}{2\pi \text{ rad}} = 35.0 \text{ rev}.$$

Significance

This example illustrates that relationships among rotational quantities are highly analogous to those among linear quantities. The answers to the questions are realistic. After unwinding for two seconds, the reel is found to spin at 220 rad/s, which is 2100 rpm. (No wonder reels sometimes make high-pitched sounds.)

In the preceding example, we considered a fishing reel with a positive angular acceleration. Now let us consider what happens with a negative angular acceleration.

Example:

Calculating the Duration When the Fishing Reel Slows Down and Stops

Now the fisherman applies a brake to the spinning reel, achieving an angular acceleration of -300 rad/s^2 . How long does it take the reel to come to a stop?

Strategy

We are asked to find the time t for the reel to come to a stop. The initial and final conditions are different from those in the previous problem, which involved the same fishing reel. Now we see that the initial angular velocity is $\omega_0 = 220 \text{ rad/s}$ and the final angular velocity ω is zero. The angular acceleration is given as $\alpha = -300 \text{ rad/s}^2$. Examining the available equations, we see all quantities but t are known in $\omega_f = \omega_0 + \alpha t$, making it easiest to use this equation.

Solution

The equation states

Equation:

$$\omega_f = \omega_0 + \alpha t.$$

We solve the equation algebraically for t and then substitute the known values as usual, yielding

Equation:

$$t = \frac{\omega_f - \omega_0}{\alpha} = \frac{0 - 220.0 \text{ rad/s}}{-300.0 \text{ rad/s}^2} = 0.733 \text{ s}.$$

Significance

Note that care must be taken with the signs that indicate the directions of various quantities. Also, note that the time to stop the reel is fairly small because the acceleration is rather large. Fishing lines sometimes snap because of the accelerations involved, and fishermen often let the fish swim for a while before applying brakes on the reel. A tired fish is slower, requiring a smaller acceleration.

Note:

Exercise:

Problem:

Check Your Understanding A centrifuge used in DNA extraction spins at a maximum rate of 7000 rpm, producing a “g-force” on the sample that is 6000 times the force of gravity. If the centrifuge takes 10 seconds to come to rest from the maximum spin rate: (a) What is the angular acceleration of the centrifuge? (b) What is the angular displacement of the centrifuge during this time?

Solution:

a. Using [\[link\]](#), we have $7000 \text{ rpm} = \frac{7000.0(2\pi \text{ rad})}{60.0 \text{ s}} = 733.0 \text{ rad/s}$,

$$\alpha = \frac{\omega - \omega_0}{t} = \frac{733.0 \text{ rad/s}}{10.0 \text{ s}} = 73.3 \text{ rad/s}^2;$$

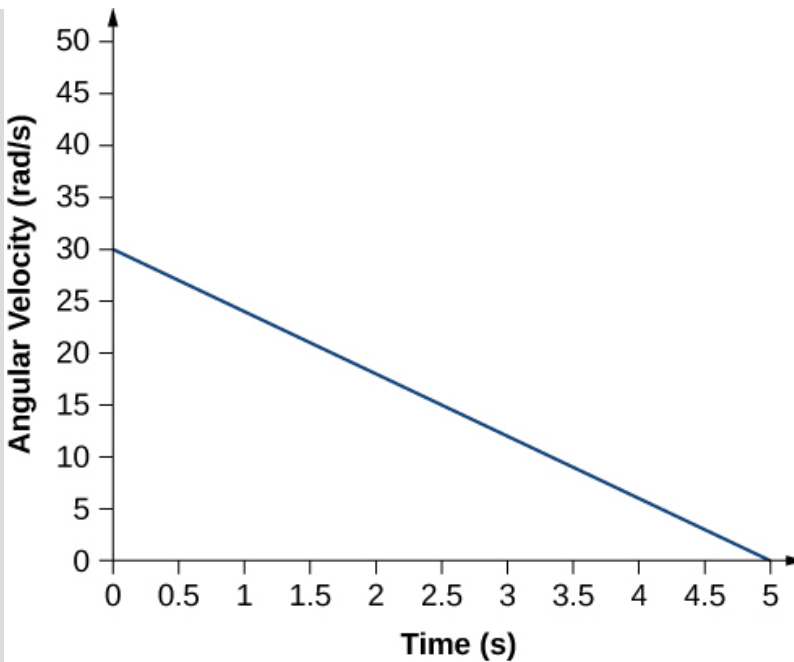
b. Using [\[link\]](#), we have

$$\omega^2 = \omega_0^2 + 2\alpha\Delta\theta \Rightarrow \Delta\theta = \frac{\omega^2 - \omega_0^2}{2\alpha} = \frac{0 - (733.0 \text{ rad/s})^2}{2(73.3 \text{ rad/s}^2)} = 3665.2 \text{ rad}$$

Example:

Angular Acceleration of a Propeller

[\[link\]](#) shows a graph of the angular velocity of a propeller on an aircraft as a function of time. Its angular velocity starts at 30 rad/s and drops linearly to 0 rad/s over the course of 5 seconds. (a) Find the angular acceleration of the object and verify the result using the kinematic equations. (b) Find the angle through which the propeller rotates during these 5 seconds and verify your result using the kinematic equations.



A graph of the angular velocity of a propeller versus time.

Strategy

- Since the angular velocity varies linearly with time, we know that the angular acceleration is constant and does not depend on the time variable. The angular acceleration is the slope of the angular velocity vs. time graph, $\alpha = \frac{d\omega}{dt}$. To calculate the slope, we read directly from [\[link\]](#), and see that $\omega_0 = 30 \text{ rad/s}$ at $t = 0 \text{ s}$ and $\omega_f = 0 \text{ rad/s}$ at $t = 5 \text{ s}$. Then, we can verify the result using $\omega = \omega_0 + \alpha t$.
- We use the equation $\omega = \frac{d\theta}{dt}$; since the time derivative of the angle is the angular velocity, we can find the angular displacement by integrating the angular velocity, which from the figure means taking the area under the angular velocity graph. In other words:

Equation:

$$\int_{\theta_0}^{\theta_f} d\theta = \theta_f - \theta_0 = \int_{t_0}^{t_f} \omega(t) dt.$$

Then we use the kinematic equations for constant acceleration to verify the result.

Solution

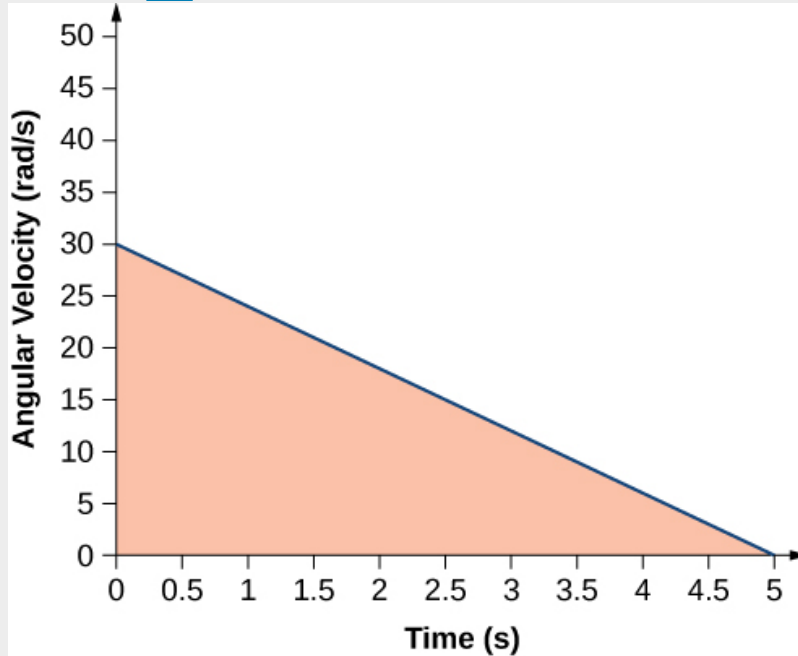
- Calculating the slope, we get

Equation:

$$\alpha = \frac{\omega - \omega_0}{t - t_0} = \frac{(0 - 30.0) \text{ rad/s}}{(5.0 - 0) \text{ s}} = -6.0 \text{ rad/s}^2.$$

We see that this is exactly [\[link\]](#) with a little rearranging of terms.

- b. We can find the area under the curve by calculating the area of the right triangle, as shown in [\[link\]](#).



The area under the curve is the area of the right triangle.

Equation:

$$\Delta\theta = \text{area (triangle);}$$

$$\Delta\theta = \frac{1}{2}(30 \text{ rad/s})(5 \text{ s}) = 75 \text{ rad.}$$

We verify the solution using [\[link\]](#):

Equation:

$$\theta_f = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2.$$

Setting $\theta_0 = 0$, we have

Equation:

$$\theta_0 = (30.0 \text{ rad/s})(5.0 \text{ s}) + \frac{1}{2}(-6.0 \text{ rad/s}^2)(5.0 \text{ rad/s})^2 = 150.0 - 75.0 = 75.0 \text{ rad.}$$

This verifies the solution found from finding the area under the curve.

Significance

We see from part (b) that there are alternative approaches to analyzing fixed-axis rotation with constant acceleration. We started with a graphical approach and verified the solution using the rotational kinematic equations. Since $\alpha = \frac{d\omega}{dt}$, we could do the same graphical analysis on an angular acceleration-vs.-time curve. The area under an α -vs.- t curve gives us the change in angular velocity. Since the angular acceleration is constant in this section, this is a straightforward exercise.

Summary

- The kinematics of rotational motion describes the relationships among rotation angle (angular position), angular velocity, angular acceleration, and time.
- For a constant angular acceleration, the angular velocity varies linearly. Therefore, the average angular velocity is 1/2 the initial plus final angular velocity over a given time period:

Equation:

$$\bar{\omega} = \frac{\omega_0 + \omega_f}{2}.$$

- We used a graphical analysis to find solutions to fixed-axis rotation with constant angular acceleration. From the relation $\omega = \frac{d\theta}{dt}$, we found that the area under an angular

velocity-vs.-time curve gives the angular displacement, $\theta_f - \theta_0 = \Delta\theta = \int_{t_0}^t \omega(t)dt$. The

results of the graphical analysis were verified using the kinematic equations for constant angular acceleration. Similarly, since $\alpha = \frac{d\omega}{dt}$, the area under an angular acceleration-

vs.-time graph gives the change in angular velocity: $\omega_f - \omega_0 = \Delta\omega = \int_{t_0}^t \alpha(t)dt$.

Conceptual Questions

Exercise:

Problem:

If a rigid body has a constant angular acceleration, what is the functional form of the angular velocity in terms of the time variable?

Solution:

straight line, linear in time variable

Exercise:

Problem:

If a rigid body has a constant angular acceleration, what is the functional form of the angular position?

Exercise:

Problem:

If the angular acceleration of a rigid body is zero, what is the functional form of the angular velocity?

Solution:

constant

Exercise:

Problem:

A massless tether with a masses tied to both ends rotates about a fixed axis through the center. Can the total acceleration of the tether/mass combination be zero if the angular velocity is constant?

Problems

Exercise:

Problem:

A wheel has a constant angular acceleration of 5.0 rad/s^2 . Starting from rest, it turns through 300 rad. (a) What is its final angular velocity? (b) How much time elapses while it turns through the 300 radians?

Solution:

- a. $\omega = 54.8 \text{ rad/s}$;
- b. $t = 11.0 \text{ s}$

Exercise:

Problem:

During a 6.0-s time interval, a flywheel with a constant angular acceleration turns through 500 radians that acquire an angular velocity of 100 rad/s. (a) What is the angular velocity at the beginning of the 6.0 s? (b) What is the angular acceleration of the flywheel?

Exercise:

Problem:

The angular velocity of a rotating rigid body increases from 500 to 1500 rev/min in 120 s. (a) What is the angular acceleration of the body? (b) Through what angle does it turn in this 120 s?

Solution:

- a. 0.87 rad/s^2 ;
- b. $\theta = 12,600 \text{ rad}$

Exercise:**Problem:**

A flywheel slows from 600 to 400 rev/min while rotating through 40 revolutions. (a) What is the angular acceleration of the flywheel? (b) How much time elapses during the 40 revolutions?

Exercise:**Problem:**

A wheel 1.0 m in radius rotates with an angular acceleration of 4.0 rad/s^2 . (a) If the wheel's initial angular velocity is 2.0 rad/s , what is its angular velocity after 10 s? (b) Through what angle does it rotate in the 10-s interval? (c) What are the tangential speed and acceleration of a point on the rim of the wheel at the end of the 10-s interval?

Solution:

- a. $\omega = 42.0 \text{ rad/s}$;
- b. $\theta = 220 \text{ rad}$; c. $v_t = 42 \text{ m/s}$
 $a_t = 4.0 \text{ m/s}^2$

Exercise:**Problem:**

A vertical wheel with a diameter of 50 cm starts from rest and rotates with a constant angular acceleration of 5.0 rad/s^2 around a fixed axis through its center counterclockwise. (a) Where is the point that is initially at the bottom of the wheel at $t = 10 \text{ s}$? (b) What is the point's linear acceleration at this instant?

Exercise:

Problem:

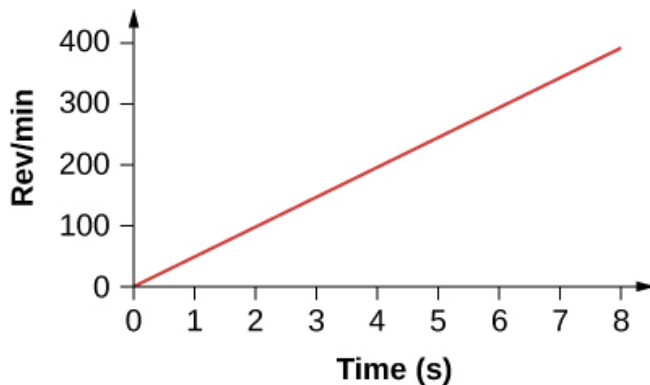
A circular disk of radius 10 cm has a constant angular acceleration of 1.0 rad/s^2 ; at $t = 0$ its angular velocity is 2.0 rad/s . (a) Determine the disk's angular velocity at $t = 5.0 \text{ s}$. (b) What is the angle it has rotated through during this time? (c) What is the tangential acceleration of a point on the disk at $t = 5.0 \text{ s}$?

Solution:

- a. $\omega = 7.0 \text{ rad/s}$;
b. $\theta = 22.5 \text{ rad}$; c. $a_t = 0.1 \text{ m/s}$

Exercise:**Problem:**

The angular velocity vs. time for a fan on a hovercraft is shown below. (a) What is the angle through which the fan blades rotate in the first 8 seconds? (b) Verify your result using the kinematic equations.

**Exercise:****Problem:**

A rod of length 20 cm has two beads attached to its ends. The rod with beads starts rotating from rest. If the beads are to have a tangential speed of 20 m/s in 7 s , what is the angular acceleration of the rod to achieve this?

Solution:

$$\alpha = 28.6 \text{ rad/s}^2.$$

Glossary

kinematics of rotational motion

describes the relationships among rotation angle, angular velocity, angular acceleration, and time

Relating Angular and Translational Quantities

By the end of this section, you will be able to:

- Given the linear kinematic equation, write the corresponding rotational kinematic equation
- Calculate the linear distances, velocities, and accelerations of points on a rotating system given the angular velocities and accelerations

In this section, we relate each of the rotational variables to the translational variables defined in [Motion Along a Straight Line](#) and [Motion in Two and Three Dimensions](#). This will complete our ability to describe rigid-body rotations.

Angular vs. Linear Variables

In [Rotational Variables](#), we introduced angular variables. If we compare the rotational definitions with the definitions of linear kinematic variables from [Motion Along a Straight Line](#) and [Motion in Two and Three Dimensions](#), we find that there is a mapping of the linear variables to the rotational ones. Linear position, velocity, and acceleration have their rotational counterparts, as we can see when we write them side by side:

	Linear	Rotational
Position	x	θ
Velocity	$v = \frac{dx}{dt}$	$\omega = \frac{d\theta}{dt}$
Acceleration	$a = \frac{dv}{dt}$	$\alpha = \frac{d\omega}{dt}$

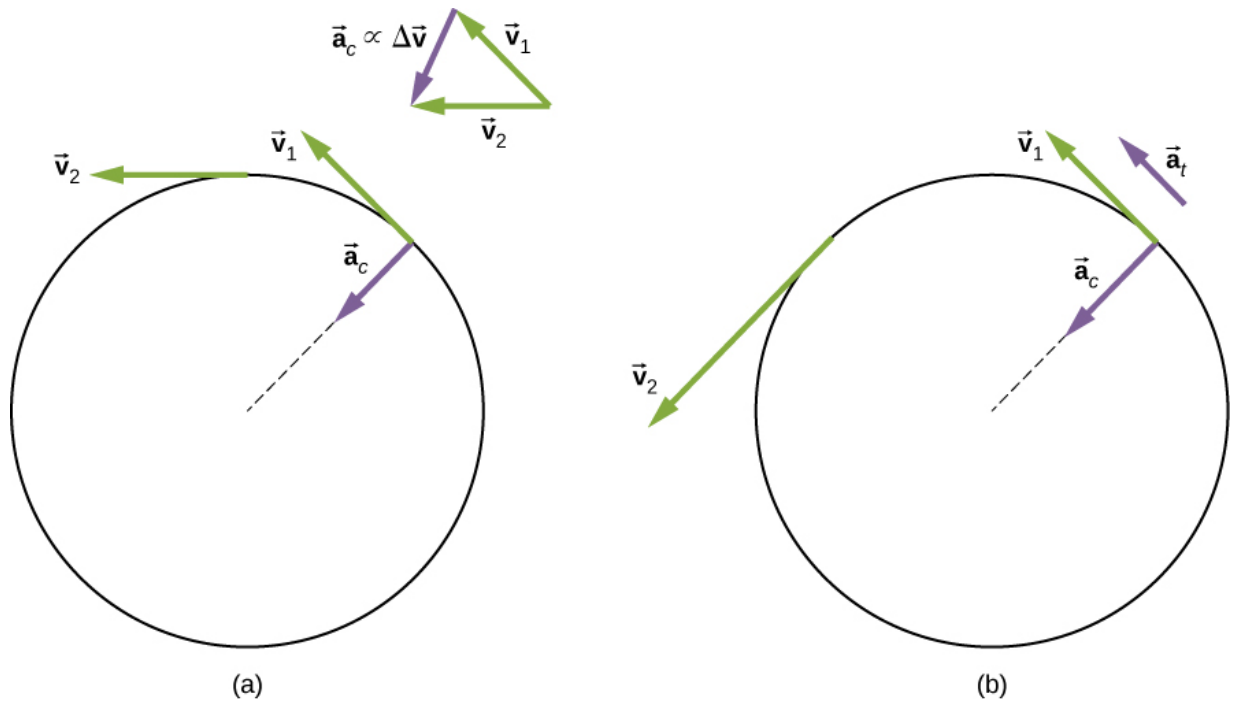
Let's compare the linear and rotational variables individually. The linear variable of position has physical units of meters, whereas the angular position variable has dimensionless units of radians, as can be seen from the definition of $\theta = \frac{s}{r}$, which is the ratio of two lengths. The linear velocity has units of m/s, and its counterpart, the angular velocity, has units of rad/s. In [Rotational Variables](#), we saw in the case of circular motion that the linear tangential speed of a particle at a radius r from the axis of rotation is related to the angular velocity by the relation $v_t = r\omega$. This could also apply to points on a rigid body rotating about a fixed axis. Here, we consider only circular motion. In circular motion, both uniform and nonuniform, there exists a centripetal acceleration ([Motion in Two and Three Dimensions](#)). The centripetal acceleration vector points inward from the particle executing circular motion toward the axis of rotation. The derivation of the magnitude of the centripetal acceleration is given in [Motion in Two and Three Dimensions](#). From that derivation, the magnitude of the centripetal acceleration was found to be

Equation:

$$a_c = \frac{v_t^2}{r},$$

where r is the radius of the circle.

Thus, in uniform circular motion when the angular velocity is constant and the angular acceleration is zero, we have a linear acceleration—that is, centripetal acceleration—since the tangential speed in [\[link\]](#) is a constant. If nonuniform circular motion is present, the rotating system has an angular acceleration, and we have both a linear centripetal acceleration that is changing (because v_t is changing) as well as a linear tangential acceleration. These relationships are shown in [\[link\]](#), where we show the centripetal and tangential accelerations for uniform and nonuniform circular motion.



- (a) Uniform circular motion: The centripetal acceleration a_c has its vector inward toward the axis of rotation. There is no tangential acceleration. (b) Nonuniform circular motion: An angular acceleration produces an inward centripetal acceleration that is changing in magnitude, plus a tangential acceleration a_t .

The centripetal acceleration is due to the change in the direction of tangential velocity, whereas the tangential acceleration is due to any change in the magnitude of the tangential velocity. The tangential and centripetal acceleration vectors \vec{a}_t and \vec{a}_c are always perpendicular to each other, as seen in [\[link\]](#). To complete this description, we can assign a **total linear acceleration** vector to a point on a rotating rigid body or a particle executing circular motion at a radius r from a fixed axis. The total linear acceleration vector \vec{a} is the vector sum of the centripetal and tangential accelerations,

Note:

Equation:

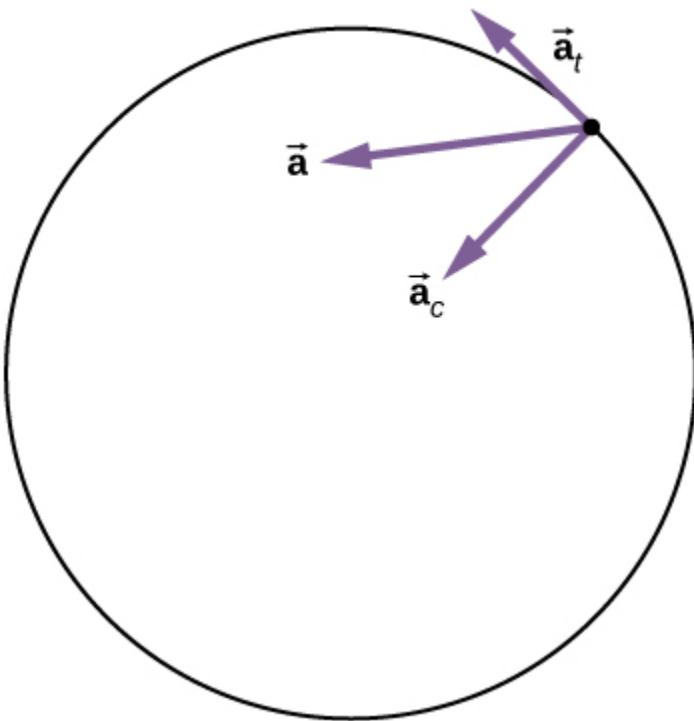
$$\vec{a} = \vec{a}_c + \vec{a}_t.$$

The total linear acceleration vector in the case of nonuniform circular motion points at an angle between the centripetal and tangential acceleration vectors, as shown in [\[link\]](#). Since $\vec{a}_c \perp \vec{a}_t$, the magnitude of the total linear acceleration is

Equation:

$$|\vec{a}| = \sqrt{a_c^2 + a_t^2}.$$

Note that if the angular acceleration is zero, the total linear acceleration is equal to the centripetal acceleration.



A particle is executing circular motion and has an angular acceleration. The total linear acceleration of the particle is the vector sum of the centripetal acceleration and tangential acceleration vectors. The total linear acceleration vector is at an angle in between the centripetal and tangential accelerations.

Relationships between Rotational and Translational Motion

We can look at two relationships between rotational and translational motion.

1. Generally speaking, the linear kinematic equations have their rotational counterparts. [\[link\]](#) lists the four linear kinematic equations and the corresponding rotational counterpart. The two sets of equations look similar to each other, but describe two different physical situations, that is, rotation and translation.

Rotational	Translational
$\theta_f = \theta_0 + \bar{\omega}t$	$x = x_0 + \bar{v}t$
$\omega_f = \omega_0 + \alpha t$	$v_f = v_0 + at$
$\theta_f = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$	$x_f = x_0 + v_0 t + \frac{1}{2}at^2$

Rotational	Translational
$\omega_f^2 = \omega_0^2 + 2\alpha(\Delta\theta)$	$v_f^2 = v_0^2 + 2a(\Delta x)$

Rotational and Translational Kinematic Equations

2. The second correspondence has to do with relating linear and rotational variables in the special case of circular motion. This is shown in [\[link\]](#), where in the third column, we have listed the connecting equation that relates the linear variable to the rotational variable. The rotational variables of angular velocity and acceleration have subscripts that indicate their definition in circular motion.

Rotational	Translational	Relationship (r = radius)
θ	s	$\theta = \frac{s}{r}$
ω	v_t	$\omega = \frac{v_t}{r}$
α	a_t	$\alpha = \frac{a_t}{r}$
	a_c	$a_c = \frac{v_t^2}{r}$

Rotational and Translational Quantities: Circular Motion

Example:

Linear Acceleration of a Centrifuge

A centrifuge has a radius of 20 cm and accelerates from a maximum rotation rate of 10,000 rpm to rest in 30 seconds under a constant angular

acceleration. It is rotating counterclockwise. What is the magnitude of the total acceleration of a point at the tip of the centrifuge at $t = 29.0\text{s}$? What is the direction of the total acceleration vector?

Strategy

With the information given, we can calculate the angular acceleration, which then will allow us to find the tangential acceleration. We can find the centripetal acceleration at $t = 0$ by calculating the tangential speed at this time. With the magnitudes of the accelerations, we can calculate the total linear acceleration. From the description of the rotation in the problem, we can sketch the direction of the total acceleration vector.

Solution

The angular acceleration is

Equation:

$$\alpha = \frac{\omega - \omega_0}{t} = \frac{0 - (1.0 \times 10^4)2\pi/60.0 \text{ s}(\text{rad/s})}{30.0 \text{ s}} = -34.9 \text{ rad/s}^2.$$

Therefore, the tangential acceleration is

Equation:

$$a_t = r\alpha = 0.2 \text{ m}(-34.9 \text{ rad/s}^2) = -7.0 \text{ m/s}^2.$$

The angular velocity at $t = 29.0 \text{ s}$ is

Equation:

$$\begin{aligned}\omega &= \omega_0 + \alpha t = 1.0 \times 10^4 \left(\frac{2\pi}{60.0 \text{ s}} \right) + (-34.9 \text{ rad/s}^2)(29.0 \text{ s}) \\ &= 1047.2 \text{ rad/s} - 1012.71 = 35.1 \text{ rad/s}.\end{aligned}$$

Thus, the tangential speed at $t = 29.0 \text{ s}$ is

Equation:

$$v_t = r\omega = 0.2 \text{ m}(35.1 \text{ rad/s}) = 7.0 \text{ m/s}.$$

We can now calculate the centripetal acceleration at $t = 29.0 \text{ s}$:

Equation:

$$a_c = \frac{v^2}{r} = \frac{(7.0 \text{ m/s})^2}{0.2 \text{ m}} = 245.0 \text{ m/s}^2.$$

Since the two acceleration vectors are perpendicular to each other, the magnitude of the total linear acceleration is

Equation:

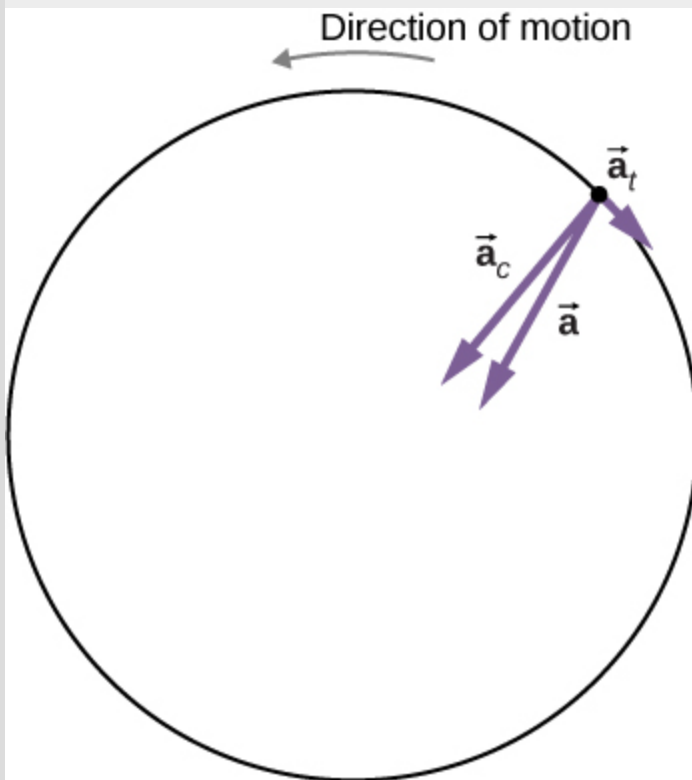
$$|\vec{a}| = \sqrt{a_c^2 + a_t^2} = \sqrt{(245.0)^2 + (-7.0)^2} = 245.1 \text{ m/s}^2.$$

Since the centrifuge has a negative angular acceleration, it is slowing down. The total acceleration vector is as shown in [\[link\]](#). The angle with respect to the centripetal acceleration vector is

Equation:

$$\theta = \tan^{-1} \frac{-7.0}{245.0} = -1.6^\circ.$$

The negative sign means that the total acceleration vector is angled toward the clockwise direction.



The centripetal, tangential, and total acceleration vectors. The centrifuge is slowing down, so the tangential acceleration is clockwise, opposite the direction of rotation (counterclockwise).

Significance

From [\[link\]](#), we see that the tangential acceleration vector is opposite the direction of rotation. The magnitude of the tangential acceleration is much smaller than the centripetal acceleration, so the total linear acceleration vector will make a very small angle with respect to the centripetal acceleration vector.

Note:

Exercise:

Problem:

Check Your Understanding A boy jumps on a merry-go-round with a radius of 5 m that is at rest. It starts accelerating at a constant rate up to an angular velocity of 5 rad/s in 20 seconds. What is the distance travelled by the boy?

Solution:

The angular acceleration is $\alpha = \frac{(5.0-0)\text{rad/s}}{20.0\text{ s}} = 0.25\text{ rad/s}^2$.

Therefore, the total angle that the boy passes through is

$$\Delta\theta = \frac{\omega^2 - \omega_0^2}{2\alpha} = \frac{(5.0)^2 - 0}{2(0.25)} = 50\text{ rad}.$$

Thus, we calculate

$$s = r\theta = 5.0\text{ m}(50.0\text{ rad}) = 250.0\text{ m}.$$

Note:

Check out this [PhET simulation](#) to change the parameters of a rotating disk (the initial angle, angular velocity, and angular acceleration), and place bugs at different radial distances from the axis. The simulation then lets you explore how circular motion relates to the bugs' xy -position, velocity, and acceleration using vectors or graphs.

Summary

- The linear kinematic equations have their rotational counterparts such that there is a mapping $x \rightarrow \theta$, $v \rightarrow \omega$, $a \rightarrow \alpha$.
- A system undergoing uniform circular motion has a constant angular velocity, but points at a distance r from the rotation axis have a linear centripetal acceleration.
- A system undergoing nonuniform circular motion has an angular acceleration and therefore has both a linear centripetal and linear tangential acceleration at a point a distance r from the axis of rotation.
- The total linear acceleration is the vector sum of the centripetal acceleration vector and the tangential acceleration vector. Since the centripetal and tangential acceleration vectors are perpendicular to each other for circular motion, the magnitude of the total linear acceleration is $|\vec{a}| = \sqrt{a_c^2 + a_t^2}$.

Conceptual Questions

Exercise:**Problem:**

Explain why centripetal acceleration changes the direction of velocity in circular motion but not its magnitude.

Solution:

The centripetal acceleration vector is perpendicular to the velocity vector.

Exercise:

Problem:

In circular motion, a tangential acceleration can change the magnitude of the velocity but not its direction. Explain your answer.

Exercise:

Problem:

Suppose a piece of food is on the edge of a rotating microwave oven plate. Does it experience nonzero tangential acceleration, centripetal acceleration, or both when: (a) the plate starts to spin faster? (b) The plate rotates at constant angular velocity? (c) The plate slows to a halt?

Solution:

a. both; b. nonzero centripetal acceleration; c. both

Problems

Exercise:

Problem:

At its peak, a tornado is 60.0 m in diameter and carries 500 km/h winds. What is its angular velocity in revolutions per second?

Exercise:

Problem:

A man stands on a merry-go-round that is rotating at 2.5 rad/s. If the coefficient of static friction between the man's shoes and the merry-go-round is $\mu_s = 0.5$, how far from the axis of rotation can he stand without sliding?

Solution:

$$r = 0.78 \text{ m}$$

Exercise:**Problem:**

An ultracentrifuge accelerates from rest to 100,000 rpm in 2.00 min.

(a) What is the average angular acceleration in rad/s^2 ? (b) What is the tangential acceleration of a point 9.50 cm from the axis of rotation? (c) What is the centripetal acceleration in m/s^2 and multiples of g of this point at full rpm? (d) What is the total distance traveled during the acceleration by a point 9.5 cm from the axis of rotation of the ultracentrifuge?

Exercise:**Problem:**

A wind turbine is rotating counterclockwise at 0.5 rev/s and slows to a stop in 10 s. Its blades are 20 m in length. (a) What is the angular acceleration of the turbine? (b) What is the centripetal acceleration of the tip of the blades at $t = 0$ s? (c) What is the magnitude and direction of the total linear acceleration of the tip of the blades at $t = 0$ s?

Solution:

a. $\alpha = -0.314 \text{ rad/s}^2$,

b. $a_c = 197.4 \text{ m/s}^2$; c.

$$a = \sqrt{a_c^2 + a_t^2} = \sqrt{197.4^2 + (-6.28)^2} = 197.5 \text{ m/s}^2$$

$$\theta = \tan^{-1} \frac{-6.28}{197.4} = -1.8^\circ \text{ in the clockwise direction from the centripetal acceleration vector}$$

Exercise:

Problem:

What is (a) the angular speed and (b) the linear speed of a point on Earth's surface at latitude 30° N. Take the radius of the Earth to be 6309 km. (c) At what latitude would your linear speed be 10 m/s?

Exercise:**Problem:**

A child with mass 40 kg sits on the edge of a merry-go-round at a distance of 3.0 m from its axis of rotation. The merry-go-round accelerates from rest up to 0.4 rev/s in 10 s. If the coefficient of static friction between the child and the surface of the merry-go-round is 0.6, does the child fall off before 5 s?

Solution:

$$ma = 40.0 \text{ kg}(5.1 \text{ m/s}^2) = 204.0 \text{ N}$$

The maximum friction force is

$$\mu_s N = 0.6(40.0 \text{ kg})(9.8 \text{ m/s}^2) = 235.2 \text{ N} \text{ so the child does not fall off yet.}$$

Exercise:**Problem:**

A bicycle wheel with radius 0.3 m rotates from rest to 3 rev/s in 5 s. What is the magnitude and direction of the total acceleration vector at the edge of the wheel at 1.0 s?

Exercise:**Problem:**

The angular velocity of a flywheel with radius 1.0 m varies according to $\omega(t) = 2.0t$. Plot $a_c(t)$ and $a_t(t)$ from $t = 0$ to 3.0 s for $r = 1.0$ m. Analyze these results to explain when $a_c \gg a_t$ and when $a_c \ll a_t$ for a point on the flywheel at a radius of 1.0 m.

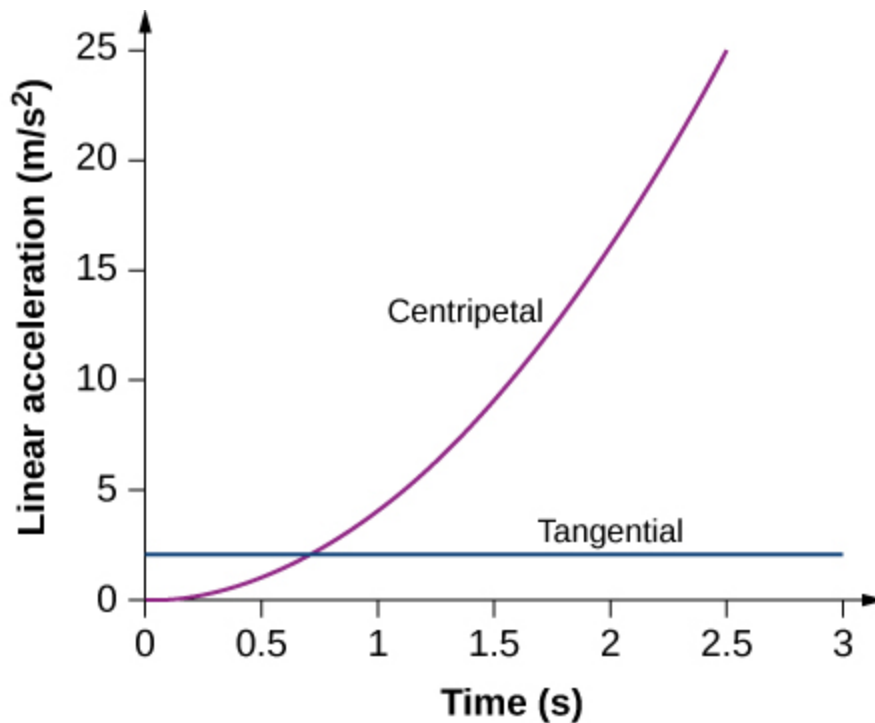
Solution:

$$v_t = r\omega = 1.0(2.0t) \text{ m/s}$$

$$a_c = \frac{v_t^2}{r} = \frac{(2.0t)^2}{1.0 \text{ m}} = 4.0t^2 \text{ m/s}^2$$

$$a_t(t) = r\alpha(t) = r \frac{d\omega}{dt} = 1.0 \text{ m}(2.0) = 2.0 \text{ m/s}^2.$$

Plotting both accelerations gives



The tangential acceleration is constant, while the centripetal acceleration is time dependent, and increases with time to values much greater than the tangential acceleration after $t = 1\text{ s}$. For times less than 0.7 s and approaching zero the centripetal acceleration is much less than the tangential acceleration.

Glossary

total linear acceleration

vector sum of the centripetal acceleration vector and the tangential acceleration vector

Moment of Inertia and Rotational Kinetic Energy

By the end of this section, you will be able to:

- Describe the differences between rotational and translational kinetic energy
- Define the physical concept of moment of inertia in terms of the mass distribution from the rotational axis
- Explain how the moment of inertia of rigid bodies affects their rotational kinetic energy
- Use conservation of mechanical energy to analyze systems undergoing both rotation and translation
- Calculate the angular velocity of a rotating system when there are energy losses due to nonconservative forces

So far in this chapter, we have been working with rotational kinematics: the description of motion for a rotating rigid body with a fixed axis of rotation. In this section, we define two new quantities that are helpful for analyzing properties of rotating objects: moment of inertia and rotational kinetic energy. With these properties defined, we will have two important tools we need for analyzing rotational dynamics.

Rotational Kinetic Energy

Any moving object has kinetic energy. We know how to calculate this for a body undergoing translational motion, but how about for a rigid body undergoing rotation? This might seem complicated because each point on the rigid body has a different velocity. However, we can make use of angular velocity—which is the same for the entire rigid body—to express the kinetic energy for a rotating object. [\[link\]](#) shows an example of a very energetic rotating body: an electric grindstone propelled by a motor. Sparks are flying, and noise and vibration are generated as the grindstone does its work. This system has considerable energy, some of it in the form of heat, light, sound, and vibration. However, most of this energy is in the form of **rotational kinetic energy**.



The rotational kinetic energy of the grindstone is converted to heat, light, sound, and vibration. (credit: Zachary David Bell, US Navy)

Energy in rotational motion is not a new form of energy; rather, it is the energy associated with rotational motion, the same as kinetic energy in translational motion. However, because kinetic energy is given by $K = \frac{1}{2}mv^2$, and velocity is a quantity that is different for every point on a rotating body about an axis, it makes sense to find a way to write kinetic energy in terms of the variable ω , which is the same for all points on a rigid rotating body. For a single particle rotating around a fixed axis, this is straightforward to calculate. We can relate the angular velocity to the magnitude of the translational velocity using the relation $v_t = \omega r$, where r is the distance of the particle from the axis of rotation and v_t is its tangential speed. Substituting into the equation for kinetic energy, we find

Equation:

$$K = \frac{1}{2}mv_t^2 = \frac{1}{2}m(\omega r)^2 = \frac{1}{2}(mr^2)\omega^2.$$

In the case of a rigid rotating body, we can divide up any body into a large number of smaller masses, each with a mass m_j and distance to the axis of rotation r_j , such that the total mass of the body is equal to the sum of the individual masses: $M = \sum_j m_j$. Each smaller mass has tangential speed v_j , where we have

dropped the subscript t for the moment. The total kinetic energy of the rigid rotating body is

Equation:

$$K = \sum_j \frac{1}{2}m_j v_j^2 = \sum_j \frac{1}{2}m_j (r_j \omega_j)^2$$

and since $\omega_j = \omega$ for all masses,

Note:

Equation:

$$K = \frac{1}{2} \left(\sum_j m_j r_j^2 \right) \omega^2.$$

The units of [\[link\]](#) are joules (J). The equation in this form is complete, but awkward; we need to find a way to generalize it.

Moment of Inertia

If we compare [\[link\]](#) to the way we wrote kinetic energy in [Work and Kinetic Energy](#), $(\frac{1}{2}mv^2)$, this suggests we have a new rotational variable to add to our list of our relations between rotational and

translational variables. The quantity $\sum_j m_j r_j^2$ is the counterpart for mass in the equation for rotational kinetic energy. This is an important new term for rotational motion. This quantity is called the **moment of inertia** I , with units of $\text{kg} \cdot \text{m}^2$:

Note:
Equation:

$$I = \sum_j m_j r_j^2.$$

For now, we leave the expression in summation form, representing the moment of inertia of a system of point particles rotating about a fixed axis. We note that the moment of inertia of a single point particle about a fixed axis is simply mr^2 , with r being the distance from the point particle to the axis of rotation. In the next section, we explore the integral form of this equation, which can be used to calculate the moment of inertia of some regular-shaped rigid bodies.

The moment of inertia is the quantitative measure of rotational inertia, just as in translational motion, and mass is the quantitative measure of linear inertia—that is, the more massive an object is, the more inertia it has, and the greater is its resistance to change in linear velocity. Similarly, the greater the moment of inertia of a rigid body or system of particles, the greater is its resistance to change in angular velocity about a fixed axis of rotation. It is interesting to see how the moment of inertia varies with r , the distance to the axis of rotation of the mass particles in [\[link\]](#). Rigid bodies and systems of particles with more mass concentrated at a greater distance from the axis of rotation have greater moments of inertia than bodies and systems of the same mass, but concentrated near the axis of rotation. In this way, we can see that a hollow cylinder has more rotational inertia than a solid cylinder of the same mass when rotating about an axis through the center. Substituting [\[link\]](#) into [\[link\]](#), the expression for the kinetic energy of a rotating rigid body becomes

Note:
Equation:

$$K = \frac{1}{2} I \omega^2.$$

We see from this equation that the kinetic energy of a rotating rigid body is directly proportional to the moment of inertia and the square of the angular velocity. This is exploited in flywheel energy-storage devices, which are designed to store large amounts of rotational kinetic energy. Many carmakers are now testing flywheel energy storage devices in their automobiles, such as the flywheel, or kinetic energy recovery system, shown in [\[link\]](#).



A KERS (kinetic energy recovery system) flywheel used in cars. (credit: "cmonville"/Flickr)

The rotational and translational quantities for kinetic energy and inertia are summarized in [\[link\]](#). The relationship column is not included because a constant doesn't exist by which we could multiply the rotational quantity to get the translational quantity, as can be done for the variables in [\[link\]](#).

Rotational	Translational
$I = \sum_j m_j r_j^2$	m
$K = \frac{1}{2} I \omega^2$	$K = \frac{1}{2} m v^2$

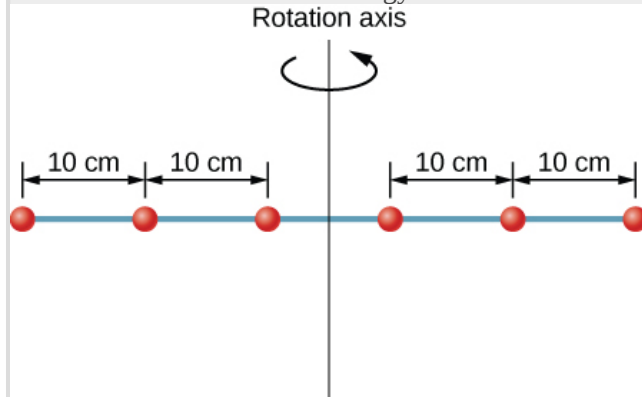
Rotational and Translational Kinetic Energies and Inertia

Example:

Moment of Inertia of a System of Particles

Six small washers are spaced 10 cm apart on a rod of negligible mass and 0.5 m in length. The mass of each washer is 20 g. The rod rotates about an axis located at 25 cm, as shown in [\[link\]](#). (a) What is the moment of inertia of the system? (b) If the two washers closest to the axis are removed, what is the

moment of inertia of the remaining four washers? (c) If the system with six washers rotates at 5 rev/s, what is its rotational kinetic energy?



Six washers are spaced 10 cm apart on a rod of negligible mass and rotating about a vertical axis.

Strategy

- We use the definition for moment of inertia for a system of particles and perform the summation to evaluate this quantity. The masses are all the same so we can pull that quantity in front of the summation symbol.
- We do a similar calculation.
- We insert the result from (a) into the expression for rotational kinetic energy.

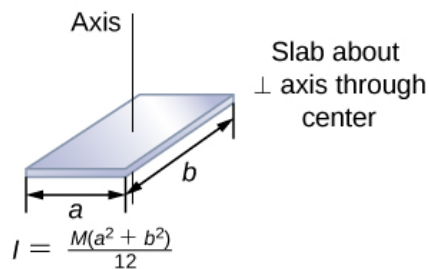
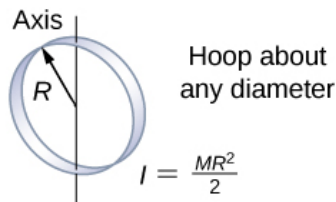
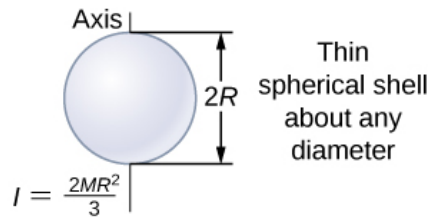
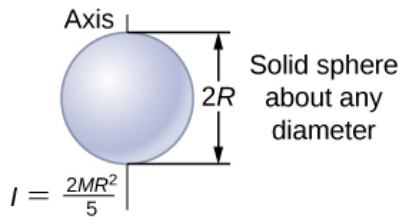
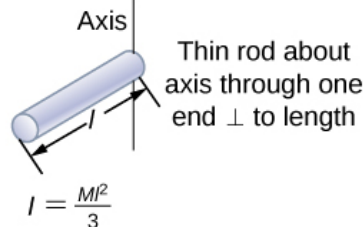
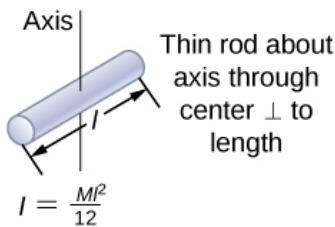
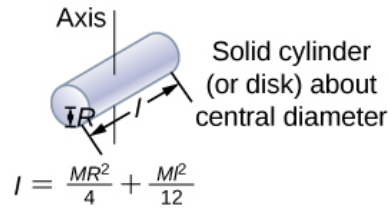
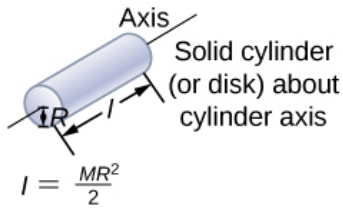
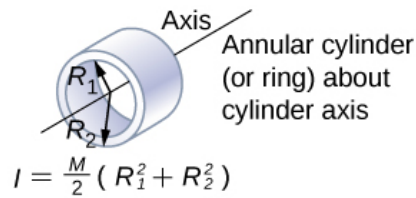
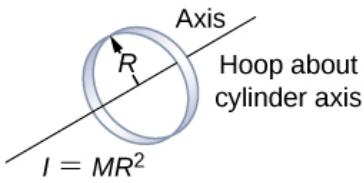
Solution

- $$I = \sum_j m_j r_j^2 = (0.02 \text{ kg})(2 \times (0.25 \text{ m})^2 + 2 \times (0.15 \text{ m})^2 + 2 \times (0.05 \text{ m})^2) = 0.0035 \text{ kg} \cdot \text{m}^2$$
- $$I = \sum_j m_j r_j^2 = (0.02 \text{ kg})(2 \times (0.25 \text{ m})^2 + 2 \times (0.15 \text{ m})^2) = 0.0034 \text{ kg} \cdot \text{m}^2.$$
- $$K = \frac{1}{2} I \omega^2 = \frac{1}{2} (0.0035 \text{ kg} \cdot \text{m}^2) (5.0 \times 2\pi \text{ rad/s})^2 = 1.73 \text{ J}.$$

Significance

We can see the individual contributions to the moment of inertia. The masses close to the axis of rotation have a very small contribution. When we removed them, it had a very small effect on the moment of inertia.

In the next section, we generalize the summation equation for point particles and develop a method to calculate moments of inertia for rigid bodies. For now, though, [link](#) gives values of rotational inertia for common object shapes around specified axes.



Values of rotational inertia for common shapes of objects.

Applying Rotational Kinetic Energy

Now let's apply the ideas of rotational kinetic energy and the moment of inertia table to get a feeling for the energy associated with a few rotating objects. The following examples will also help get you comfortable using these equations. First, let's look at a general problem-solving strategy for rotational energy.

Note:

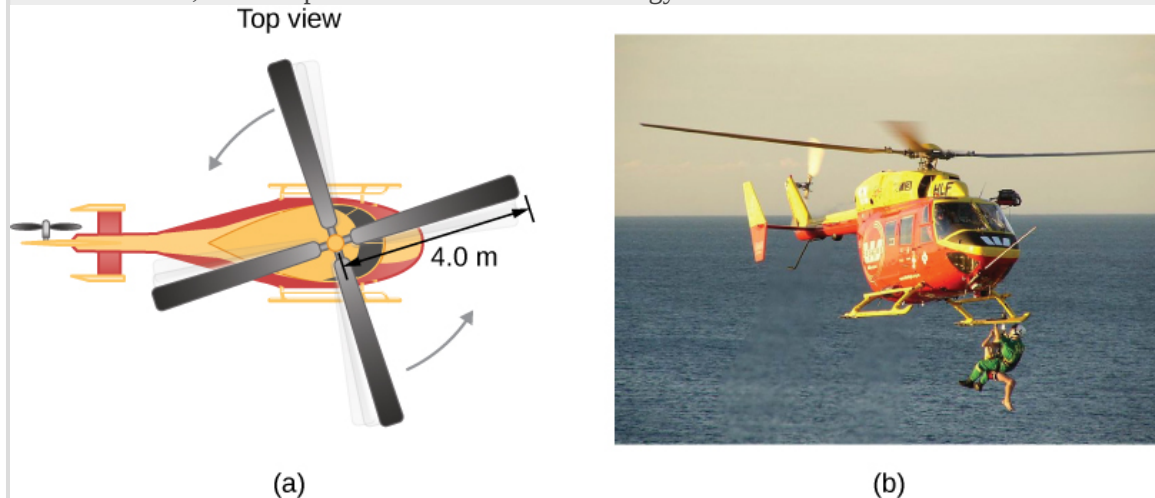
Rotational Energy

1. Determine that energy or work is involved in the rotation.
2. Determine the system of interest. A sketch usually helps.
3. Analyze the situation to determine the types of work and energy involved.
4. If there are no losses of energy due to friction and other nonconservative forces, mechanical energy is conserved, that is, $K_i + U_i = K_f + U_f$.
5. If nonconservative forces are present, mechanical energy is not conserved, and other forms of energy, such as heat and light, may enter or leave the system. Determine what they are and calculate them as necessary.
6. Eliminate terms wherever possible to simplify the algebra.
7. Evaluate the numerical solution to see if it makes sense in the physical situation presented in the wording of the problem.

Example:

Calculating Helicopter Energies

A typical small rescue helicopter has four blades: Each is 4.00 m long and has a mass of 50.0 kg ([link](#)). The blades can be approximated as thin rods that rotate about one end of an axis perpendicular to their length. The helicopter has a total loaded mass of 1000 kg. (a) Calculate the rotational kinetic energy in the blades when they rotate at 300 rpm. (b) Calculate the translational kinetic energy of the helicopter when it flies at 20.0 m/s, and compare it with the rotational energy in the blades.



(a) Sketch of a four-blade helicopter. (b) A water rescue operation featuring a helicopter from the Auckland Westpac Rescue Helicopter Service. (credit b: modification of work by “111 Emergency”/Flickr)

Strategy

Rotational and translational kinetic energies can be calculated from their definitions. The wording of the problem gives all the necessary constants to evaluate the expressions for the rotational and translational kinetic energies.

Solution

- a. The rotational kinetic energy is

Equation:

$$K = \frac{1}{2} I \omega^2.$$

We must convert the angular velocity to radians per second and calculate the moment of inertia before we can find K . The angular velocity ω is

Equation:

$$\omega = \frac{300 \text{ rev}}{1.00 \text{ min}} \frac{2\pi \text{ rad}}{1 \text{ rev}} \frac{1.00 \text{ min}}{60.0 \text{ s}} = 31.4 \frac{\text{rad}}{\text{s}}.$$

The moment of inertia of one blade is that of a thin rod rotated about its end, listed in [\[link\]](#). The total I is four times this moment of inertia because there are four blades. Thus,

Equation:

$$I = 4 \frac{Ml^2}{3} = 4 \times \frac{(50.0 \text{ kg})(4.00 \text{ m})^2}{3} = 1067.0 \text{ kg} \cdot \text{m}^2.$$

Entering ω and I into the expression for rotational kinetic energy gives

Equation:

$$K = 0.5(1067 \text{ kg} \cdot \text{m}^2)(31.4 \text{ rad/s})^2 = 5.26 \times 10^5 \text{ J}.$$

b. Entering the given values into the equation for translational kinetic energy, we obtain

Equation:

$$K = \frac{1}{2} mv^2 = (0.5)(1000.0 \text{ kg})(20.0 \text{ m/s})^2 = 2.00 \times 10^5 \text{ J}.$$

To compare kinetic energies, we take the ratio of translational kinetic energy to rotational kinetic energy. This ratio is

Equation:

$$\frac{2.00 \times 10^5 \text{ J}}{5.26 \times 10^5 \text{ J}} = 0.380.$$

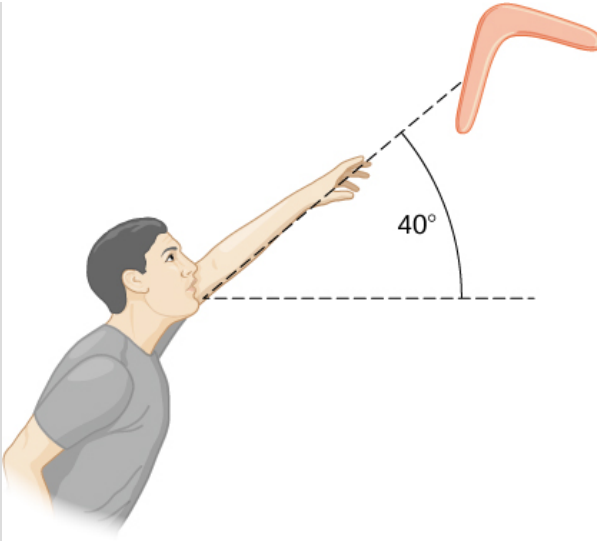
Significance

The ratio of translational energy to rotational kinetic energy is only 0.380. This ratio tells us that most of the kinetic energy of the helicopter is in its spinning blades.

Example:

Energy in a Boomerang

A person hurls a boomerang into the air with a velocity of 30.0 m/s at an angle of 40.0° with respect to the horizontal ([\[link\]](#)). It has a mass of 1.0 kg and is rotating at 10.0 rev/s. The moment of inertia of the boomerang is given as $I = \frac{1}{12} mL^2$ where $L = 0.7 \text{ m}$. (a) What is the total energy of the boomerang when it leaves the hand? (b) How high does the boomerang go from the elevation of the hand, neglecting air resistance?



A boomerang is hurled into the air at an initial angle of 40° .

Strategy

We use the definitions of rotational and linear kinetic energy to find the total energy of the system. The problem states to neglect air resistance, so we don't have to worry about energy loss. In part (b), we use conservation of mechanical energy to find the maximum height of the boomerang.

Solution

- a. Moment of inertia: $I = \frac{1}{12}mL^2 = \frac{1}{12}(1.0 \text{ kg})(0.7\text{m})^2 = 0.041 \text{ kg} \cdot \text{m}^2$.
Angular velocity: $\omega = (10.0 \text{ rev/s})(2\pi) = 62.83 \text{ rad/s}$.

The rotational kinetic energy is therefore

Equation:

$$K_R = \frac{1}{2}(0.041 \text{ kg} \cdot \text{m}^2)(62.83 \text{ rad/s})^2 = 80.93 \text{ J}.$$

The translational kinetic energy is

Equation:

$$K_T = \frac{1}{2}mv^2 = \frac{1}{2}(1.0 \text{ kg})(30.0 \text{ m/s})^2 = 450.0 \text{ J}.$$

Thus, the total energy in the boomerang is

Equation:

$$K_{\text{Total}} = K_R + K_T = 80.93 + 450.0 = 530.93 \text{ J}.$$

- b. We use conservation of mechanical energy. Since the boomerang is launched at an angle, we need to write the total energies of the system in terms of its linear kinetic energies using the velocity in the x- and y-directions. The total energy when the boomerang leaves the hand is

Equation:

$$E_{\text{Before}} = \frac{1}{2}mv_x^2 + \frac{1}{2}mv_y^2 + \frac{1}{2}I\omega^2.$$

The total energy at maximum height is

Equation:

$$E_{\text{Final}} = \frac{1}{2}mv_x^2 + \frac{1}{2}I\omega^2 + mgh.$$

By conservation of mechanical energy, $E_{\text{Before}} = E_{\text{Final}}$ so we have, after canceling like terms,

Equation:

$$\frac{1}{2}mv_y^2 = mgh.$$

Since $v_y = 30.0 \text{ m/s}(\sin 40^\circ) = 19.28 \text{ m/s}$, we find

Equation:

$$h = \frac{(19.28 \text{ m/s})^2}{2(9.8 \text{ m/s}^2)} = 18.97 \text{ m}.$$

Significance

In part (b), the solution demonstrates how energy conservation is an alternative method to solve a problem that normally would be solved using kinematics. In the absence of air resistance, the rotational kinetic energy was not a factor in the solution for the maximum height.

Note:

Exercise:

Problem:

Check Your Understanding A nuclear submarine propeller has a moment of inertia of $800.0 \text{ kg} \cdot \text{m}^2$. If the submerged propeller has a rotation rate of 4.0 rev/s when the engine is cut, what is the rotation rate of the propeller after 5.0 s when water resistance has taken $50,000 \text{ J}$ out of the system?

Solution:

The initial rotational kinetic energy of the propeller is

$$K_0 = \frac{1}{2}I\omega^2 = \frac{1}{2}(800.0 \text{ kg} \cdot \text{m}^2)(4.0 \times 2\pi \text{ rad/s})^2 = 2.53 \times 10^5 \text{ J}.$$

At 5.0 s the new rotational kinetic energy of the propeller is

$$K_f = 2.03 \times 10^5 \text{ J}.$$

and the new angular velocity is

$$\omega = \sqrt{\frac{2(2.03 \times 10^5 \text{ J})}{800.0 \text{ kg} \cdot \text{m}^2}} = 22.53 \text{ rad/s}$$

which is 3.58 rev/s .

Summary

- The rotational kinetic energy is the kinetic energy of rotation of a rotating rigid body or system of particles, and is given by $K = \frac{1}{2}I\omega^2$, where I is the moment of inertia, or “rotational mass” of the rigid body or system of particles.
- The moment of inertia for a system of point particles rotating about a fixed axis is $I = \sum_j m_j r_j^2$,

where m_j is the mass of the point particle and r_j is the distance of the point particle to the rotation axis. Because of the r^2 term, the moment of inertia increases as the square of the distance to the fixed rotational axis. The moment of inertia is the rotational counterpart to the mass in linear motion.

- In systems that are both rotating and translating, conservation of mechanical energy can be used if there are no nonconservative forces at work. The total mechanical energy is then conserved and is the sum of the rotational and translational kinetic energies, and the gravitational potential energy.

Conceptual Questions

Exercise:

Problem:

What if another planet the same size as Earth were put into orbit around the Sun along with Earth. Would the moment of inertia of the system increase, decrease, or stay the same?

Exercise:

Problem:

A solid sphere is rotating about an axis through its center at a constant rotation rate. Another hollow sphere of the same mass and radius is rotating about its axis through the center at the same rotation rate. Which sphere has a greater rotational kinetic energy?

Solution:

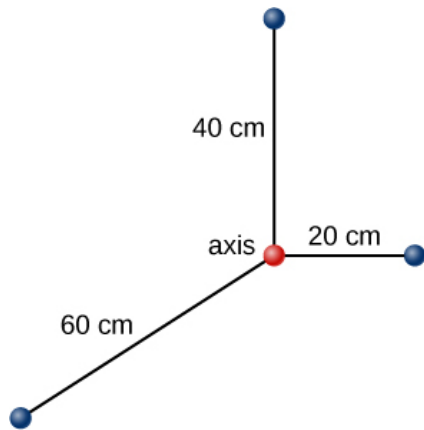
The hollow sphere, since the mass is distributed further away from the rotation axis.

Problems

Exercise:

Problem:

A system of point particles is shown in the following figure. Each particle has mass 0.3 kg and they all lie in the same plane. (a) What is the moment of inertia of the system about the given axis? (b) If the system rotates at 5 rev/s, what is its rotational kinetic energy?



Exercise:

Problem:

(a) Calculate the rotational kinetic energy of Earth on its axis. (b) What is the rotational kinetic energy of Earth in its orbit around the Sun?

Solution:

a. $K = 2.56 \times 10^{29} \text{ J};$

b. $K = 2.68 \times 10^{33} \text{ J}$

Exercise:

Problem:

Calculate the rotational kinetic energy of a 12-kg motorcycle wheel if its angular velocity is 120 rad/s and its inner radius is 0.280 m and outer radius 0.330 m.

Exercise:

Problem:

A baseball pitcher throws the ball in a motion where there is rotation of the forearm about the elbow joint as well as other movements. If the linear velocity of the ball relative to the elbow joint is 20.0 m/s at a distance of 0.480 m from the joint and the moment of inertia of the forearm is $0.500 \text{ kg}\cdot\text{m}^2$, what is the rotational kinetic energy of the forearm?

Solution:

$$K = 434.0 \text{ J}$$

Exercise:

Problem:

A diver goes into a somersault during a dive by tucking her limbs. If her rotational kinetic energy is 100 J and her moment of inertia in the tuck is $9.0 \text{ kg}\cdot\text{m}^2$, what is her rotational rate during the somersault?

Exercise:

Problem:

An aircraft is coming in for a landing at 300 meters height when the propeller falls off. The aircraft is flying at 40.0 m/s horizontally. The propeller has a rotation rate of 20 rev/s, a moment of inertia of $70.0 \text{ kg}\cdot\text{m}^2$, and a mass of 200 kg. Neglect air resistance. (a) With what translational velocity does the propeller hit the ground? (b) What is the rotation rate of the propeller at impact?

Solution:

- a. $v_f = 86.5 \text{ m/s}$;
b. The rotational rate of the propeller stays the same at 20 rev/s.

Exercise:**Problem:**

If air resistance is present in the preceding problem and reduces the propeller's rotational kinetic energy at impact by 30%, what is the propeller's rotation rate at impact?

Exercise:**Problem:**

A neutron star of mass $2 \times 10^{30} \text{ kg}$ and radius 10 km rotates with a period of 0.02 seconds. What is its rotational kinetic energy?

Solution:

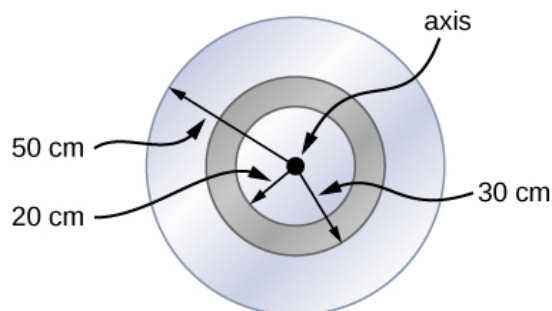
$$K = 3.95 \times 10^{42} \text{ J}$$

Exercise:**Problem:**

An electric sander consisting of a rotating disk of mass 0.7 kg and radius 10 cm rotates at 15 rev/s. When applied to a rough wooden wall the rotation rate decreases by 20%. (a) What is the final rotational kinetic energy of the rotating disk? (b) How much has its rotational kinetic energy decreased?

Exercise:**Problem:**

A system consists of a disk of mass 2.0 kg and radius 50 cm upon which is mounted an annular cylinder of mass 1.0 kg with inner radius 20 cm and outer radius 30 cm (see below). The system rotates about an axis through the center of the disk and annular cylinder at 10 rev/s. (a) What is the moment of inertia of the system? (b) What is its rotational kinetic energy?



Solution:

a. $I = 0.315 \text{ kg} \cdot \text{m}^2$;

b. $K = 621.8 \text{ J}$

Glossary

moment of inertia

rotational mass of rigid bodies that relates to how easy or hard it will be to change the angular velocity of the rotating rigid body

rotational kinetic energy

kinetic energy due to the rotation of an object; this is part of its total kinetic energy

Calculating Moments of Inertia

By the end of this section, you will be able to:

- Calculate the moment of inertia for uniformly shaped, rigid bodies
- Apply the parallel axis theorem to find the moment of inertia about any axis parallel to one already known
- Calculate the moment of inertia for compound objects

In the preceding section, we defined the moment of inertia but did not show how to calculate it. In this section, we show how to calculate the moment of inertia for several standard types of objects, as well as how to use known moments of inertia to find the moment of inertia for a shifted axis or for a compound object. This section is very useful for seeing how to apply a general equation to complex objects (a skill that is critical for more advanced physics and engineering courses).

Moment of Inertia

We defined the moment of inertia I of an object to be $I = \sum_i m_i r_i^2$ for all the point

masses that make up the object. Because r is the distance to the axis of rotation from each piece of mass that makes up the object, the moment of inertia for any object depends on the chosen axis. To see this, let's take a simple example of two masses at the end of a massless (negligibly small mass) rod ([link](#)) and calculate the moment of inertia about two different axes. In this case, the summation over the masses is simple because the two masses at the end of the barbell can be approximated as point masses, and the sum therefore has only two terms.

In the case with the axis in the center of the barbell, each of the two masses m is a distance R away from the axis, giving a moment of inertia of

Equation:

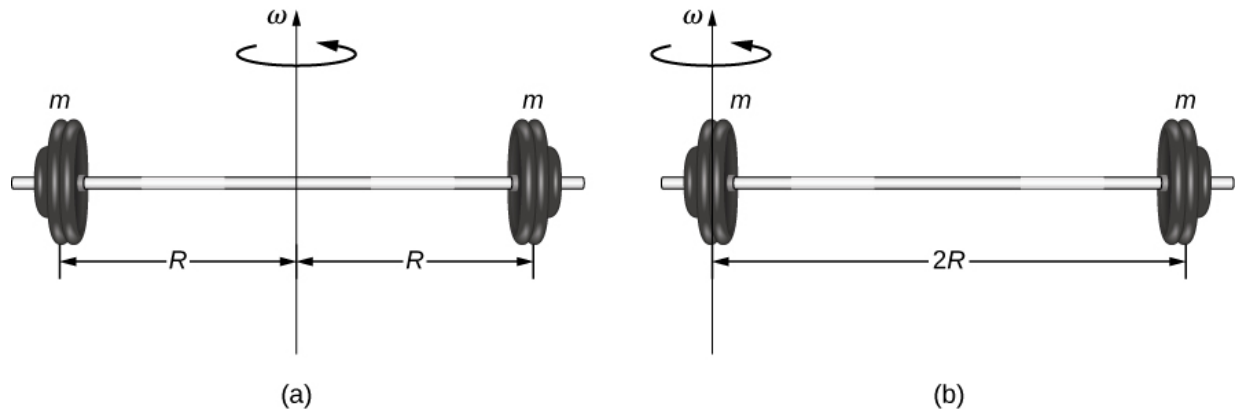
$$I_1 = mR^2 + mR^2 = 2mR^2.$$

In the case with the axis at the end of the barbell—passing through one of the masses—the moment of inertia is

Equation:

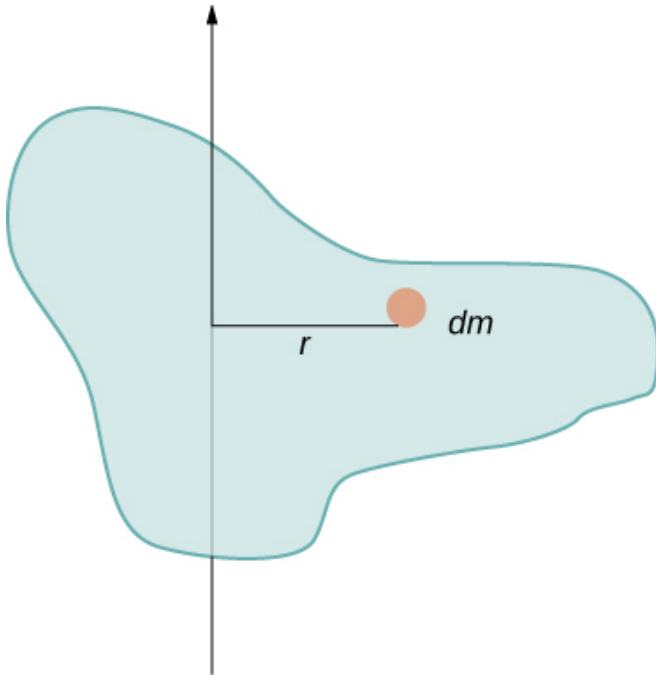
$$I_2 = m(0)^2 + m(2R)^2 = 4mR^2.$$

From this result, we can conclude that it is twice as hard to rotate the barbell about the end than about its center.



(a) A barbell with an axis of rotation through its center; (b) a barbell with an axis of rotation through one end.

In this example, we had two point masses and the sum was simple to calculate. However, to deal with objects that are not point-like, we need to think carefully about each of the terms in the equation. The equation asks us to sum over each ‘piece of mass’ a certain distance from the axis of rotation. But what exactly does each ‘piece of mass’ mean? Recall that in our derivation of this equation, each piece of mass had the same magnitude of velocity, which means the whole piece had to have a single distance r to the axis of rotation. However, this is not possible unless we take an infinitesimally small piece of mass dm , as shown in [\[link\]](#).



Using an infinitesimally small piece of mass to calculate the contribution to the total moment of inertia.

The need to use an infinitesimally small piece of mass dm suggests that we can write the moment of inertia by evaluating an integral over infinitesimal masses rather than doing a discrete sum over finite masses:

Note:

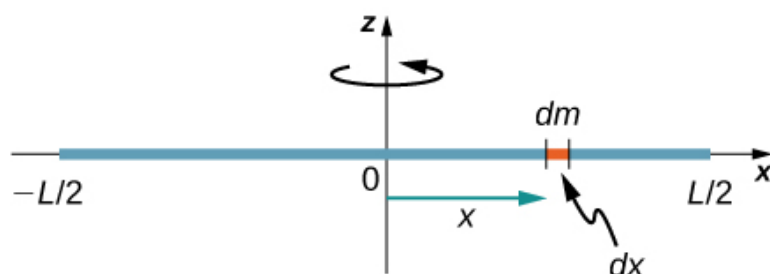
Equation:

$$I = \sum_i m_i r_i^2 \text{ becomes } I = \int r^2 dm.$$

This, in fact, is the form we need to generalize the equation for complex shapes. It is best to work out specific examples in detail to get a feel for how to calculate the moment of inertia for specific shapes. This is the focus of most of the rest of this section.

A uniform thin rod with an axis through the center

Consider a uniform (density and shape) thin rod of mass M and length L as shown in [\[link\]](#). We want a thin rod so that we can assume the cross-sectional area of the rod is small and the rod can be thought of as a string of masses along a one-dimensional straight line. In this example, the axis of rotation is perpendicular to the rod and passes through the midpoint for simplicity. Our task is to calculate the moment of inertia about this axis. We orient the axes so that the z -axis is the axis of rotation and the x -axis passes through the length of the rod, as shown in the figure. This is a convenient choice because we can then integrate along the x -axis.



Calculation of the moment of inertia I for a uniform thin rod about an axis through the center of the rod.

We define dm to be a small element of mass making up the rod. The moment of inertia integral is an integral over the mass distribution. However, we know how to integrate over space, not over mass. We therefore need to find a way to relate mass to spatial variables. We do this using the **linear mass density** λ of the object, which is the mass per unit length. Since the mass density of this object is uniform, we can write

Equation:

$$\lambda = \frac{m}{l} \text{ or } m = \lambda l.$$

If we take the differential of each side of this equation, we find

Equation:

$$dm = d(\lambda l) = \lambda(dl)$$

since λ is constant. We chose to orient the rod along the x -axis for convenience—this is where that choice becomes very helpful. Note that a piece of the rod dl lies completely along the x -axis and has a length dx ; in fact, $dl = dx$ in this situation. We can therefore write $dm = \lambda(dx)$, giving us an integration variable that we know how to deal with. The distance of each piece of mass dm from the axis is given by the variable x , as shown in the figure. Putting this all together, we obtain

Equation:

$$I = \int r^2 dm = \int x^2 dm = \int x^2 \lambda dx.$$

The last step is to be careful about our limits of integration. The rod extends from $x = -L/2$ to $x = L/2$, since the axis is in the middle of the rod at $x = 0$. This gives us

Equation:

$$\begin{aligned} I &= \int_{-L/2}^{L/2} x^2 \lambda dx = \lambda \frac{x^3}{3} \bigg|_{-L/2}^{L/2} = \lambda \left(\frac{1}{3} \right) \left[\left(\frac{L}{2} \right)^3 - \left(\frac{-L}{2} \right)^3 \right] \\ &= \lambda \left(\frac{1}{3} \right) \frac{L^3}{8} (2) = \frac{M}{L} \left(\frac{1}{3} \right) \frac{L^3}{8} (2) = \frac{1}{12} ML^2. \end{aligned}$$

Next, we calculate the moment of inertia for the same uniform thin rod but with a different axis choice so we can compare the results. We would expect the moment of inertia to be smaller about an axis through the center of mass than the endpoint axis, just as it was for the barbell example at the start of this section. This happens because more mass is distributed farther from the axis of rotation.

A uniform thin rod with axis at the end

Now consider the same uniform thin rod of mass M and length L , but this time we move the axis of rotation to the end of the rod. We wish to find the moment of inertia about this new axis ([link](#)). The quantity dm is again defined to be a small element of mass making up the rod. Just as before, we obtain

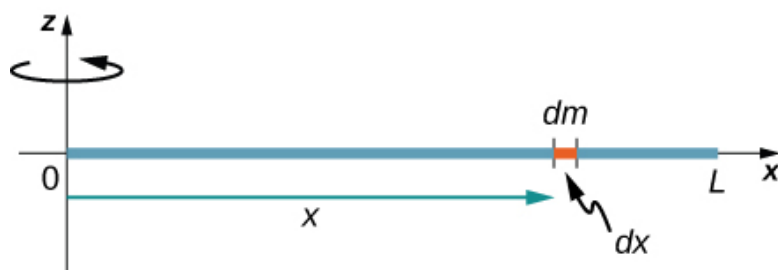
Equation:

$$I = \int r^2 dm = \int x^2 dm = \int x^2 \lambda dx.$$

However, this time we have different limits of integration. The rod extends from $x = 0$ to $x = L$, since the axis is at the end of the rod at $x = 0$. Therefore we find

Equation:

$$\begin{aligned} I &= \int_0^L x^2 \lambda dx = \lambda \frac{x^3}{3} \bigg|_0^L = \lambda \left(\frac{1}{3} \right) [(L)^3 - (0)^3] \\ &= \lambda \left(\frac{1}{3} \right) L^3 = \frac{M}{L} \left(\frac{1}{3} \right) L^3 = \frac{1}{3} ML^2. \end{aligned}$$



Calculation of the moment of inertia I for a uniform thin rod about an axis through the end of the rod.

Note the rotational inertia of the rod about its endpoint is larger than the rotational inertia about its center (consistent with the barbell example) by a factor of four.

The Parallel-Axis Theorem

The similarity between the process of finding the moment of inertia of a rod about an axis through its middle and about an axis through its end is striking, and suggests that there might be a simpler method for determining the moment of inertia for a rod about any axis parallel to the axis through the center of mass. Such an axis is called a **parallel axis**. There is a theorem for this, called the **parallel-axis theorem**, which we state here but do not derive in this text.

Note:

Parallel-Axis Theorem

Let m be the mass of an object and let d be the distance from an axis through the object's center of mass to a new axis. Then we have

Equation:

$$I_{\text{parallel-axis}} = I_{\text{center of mass}} + md^2.$$

Let's apply this to the rod examples solved above:

Equation:

$$I_{\text{end}} = I_{\text{center of mass}} + md^2 = \frac{1}{12}mL^2 + m\left(\frac{L}{2}\right)^2 = \left(\frac{1}{12} + \frac{1}{4}\right)mL^2 = \frac{1}{3}mL^2.$$

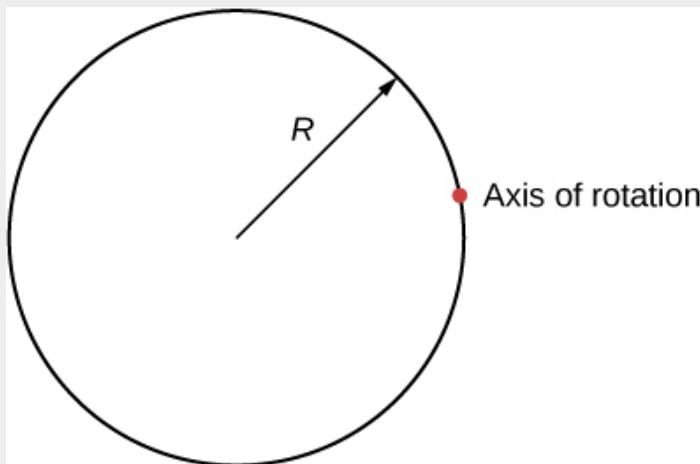
This result agrees with our more lengthy calculation from above. This is a useful equation that we apply in some of the examples and problems.

Note:

Exercise:

Problem:

Check Your Understanding What is the moment of inertia of a cylinder of radius R and mass m about an axis through a point on the surface, as shown below?

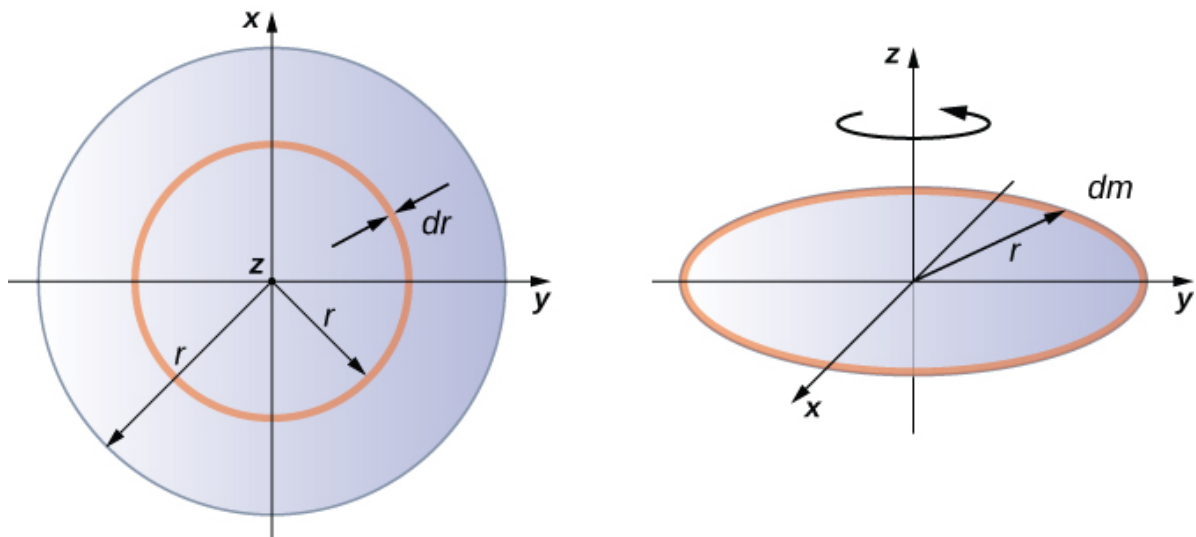


Solution:

$$I_{\text{parallel-axis}} = I_{\text{center of mass}} + md^2 = mR^2 + mR^2 = 2mR^2$$

A uniform thin disk about an axis through the center

Integrating to find the moment of inertia of a two-dimensional object is a little bit trickier, but one shape is commonly done at this level of study—a uniform thin disk about an axis through its center ([link](#)).



Calculating the moment of inertia for a thin disk about an axis through its center.

Since the disk is thin, we can take the mass as distributed entirely in the xy -plane. We again start with the relationship for the **surface mass density**, which is the mass per unit surface area. Since it is uniform, the surface mass density σ is constant:

Equation:

$$\sigma = \frac{m}{A} \quad \text{or} \quad \sigma A = m, \text{ so } dm = \sigma(dA).$$

Now we use a simplification for the area. The area can be thought of as made up of a series of thin rings, where each ring is a mass increment dm of radius r equidistant from the axis, as shown in part (b) of the figure. The infinitesimal area of each ring dA is therefore given by the length of each ring ($2\pi r$) times the infinitesimal width of each ring dr :

Equation:

$$A = \pi r^2, dA = d(\pi r^2) = \pi dr^2 = 2\pi r dr.$$

The full area of the disk is then made up from adding all the thin rings with a radius range from 0 to R . This radius range then becomes our limits of integration for dr , that is, we integrate from $r = 0$ to $r = R$. Putting this all together, we have

Equation:

$$\begin{aligned} I &= \int_0^R r^2 \sigma (2\pi r) dr = 2\pi \sigma \int_0^R r^3 dr = 2\pi \sigma \frac{r^4}{4} \Big|_0^R = 2\pi \sigma \left(\frac{R^4}{4} - 0 \right) \\ &= 2\pi \frac{m}{A} \left(\frac{R^4}{4} \right) = 2\pi \frac{m}{\pi R^2} \left(\frac{R^4}{4} \right) = \frac{1}{2} m R^2. \end{aligned}$$

Note that this agrees with the value given in [\[link\]](#).

Calculating the moment of inertia for compound objects

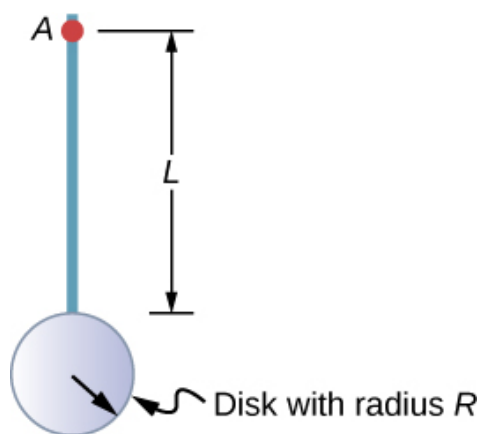
Now consider a compound object such as that in [\[link\]](#), which depicts a thin disk at the end of a thin rod. This cannot be easily integrated to find the moment of inertia because it is not a uniformly shaped object. However, if we go back to the initial definition of moment of inertia as a summation, we can reason that a compound object's moment of inertia can be found from the sum of each part of the object:

Note:

Equation:

$$I_{\text{total}} = \sum_i I_i.$$

It is important to note that the moments of inertia of the objects in [\[link\]](#) are *about a common axis*. In the case of this object, that would be a rod of length L rotating about its end, and a thin disk of radius R rotating about an axis shifted off of the center by a distance $L + R$, where R is the radius of the disk. Let's define the mass of the rod to be m_r and the mass of the disk to be m_d .



Compound object consisting of a disk at the end of a rod. The axis of rotation is located at A.

The moment of inertia of the rod is simply $\frac{1}{3}m_rL^2$, but we have to use the parallel-axis theorem to find the moment of inertia of the disk about the axis shown. The moment of inertia of the disk about its center is $\frac{1}{2}m_dR^2$ and we apply the parallel-axis theorem $I_{\text{parallel-axis}} = I_{\text{center of mass}} + md^2$ to find

Equation:

$$I_{\text{parallel-axis}} = \frac{1}{2}m_dR^2 + m_d(L + R)^2.$$

Adding the moment of inertia of the rod plus the moment of inertia of the disk with a shifted axis of rotation, we find the moment of inertia for the compound object to be

Equation:

$$I_{\text{total}} = \frac{1}{3}m_rL^2 + \frac{1}{2}m_dR^2 + m_d(L + R)^2.$$

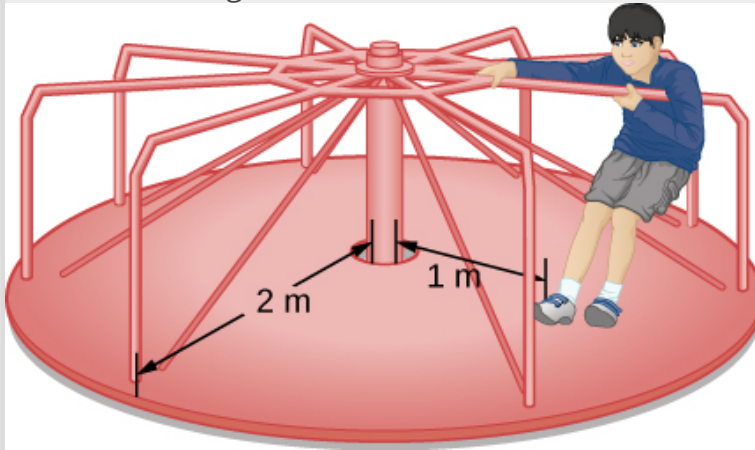
Applying moment of inertia calculations to solve problems

Now let's examine some practical applications of moment of inertia calculations.

Example:

Person on a Merry-Go-Round

A 25-kg child stands at a distance $r = 1.0$ m from the axis of a rotating merry-go-round ([link](#)). The merry-go-round can be approximated as a uniform solid disk with a mass of 500 kg and a radius of 2.0 m. Find the moment of inertia of this system.



Calculating the moment of inertia for a child on a merry-go-round.

Strategy

This problem involves the calculation of a moment of inertia. We are given the mass and distance to the axis of rotation of the child as well as the mass and radius of the merry-go-round. Since the mass and size of the child are much smaller than the merry-go-round, we can approximate the child as a point mass. The notation we use is $m_c = 25$ kg, $r_c = 1.0$ m, $m_m = 500$ kg, $r_m = 2.0$ m.

Our goal is to find $I_{\text{total}} = \sum_i I_i$.

Solution

For the child, $I_c = m_c r^2$, and for the merry-go-round, $I_m = \frac{1}{2} m_m r^2$. Therefore

Equation:

$$I_{\text{total}} = 25(1)^2 + \frac{1}{2}(500)(2)^2 = 25 + 1000 = 1025 \text{ kg} \cdot \text{m}^2.$$

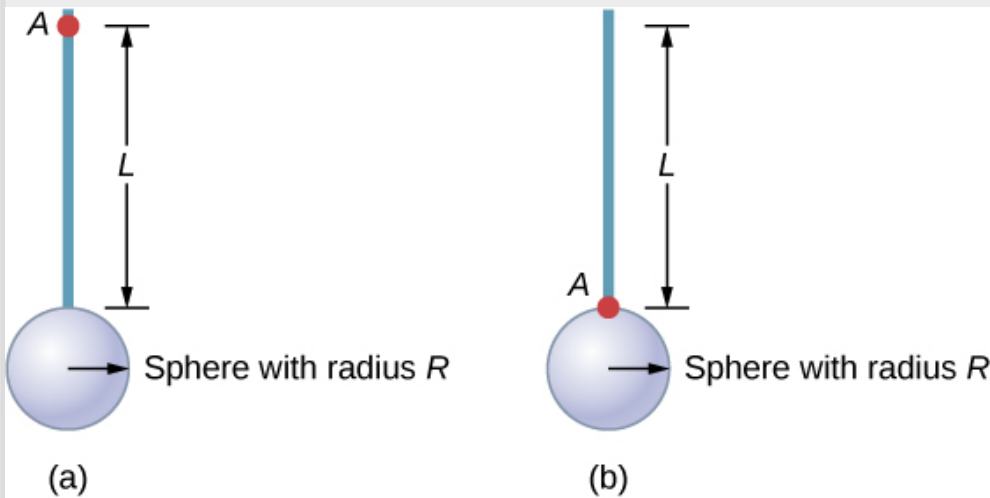
Significance

The value should be close to the moment of inertia of the merry-go-round by itself because it has much more mass distributed away from the axis than the child does.

Example:

Rod and Solid Sphere

Find the moment of inertia of the rod and solid sphere combination about the two axes as shown below. The rod has length 0.5 m and mass 2.0 kg. The radius of the sphere is 20.0 cm and has mass 1.0 kg.



Strategy

Since we have a compound object in both cases, we can use the parallel-axis theorem to find the moment of inertia about each axis. In (a), the center of mass of the sphere is located at a distance $L + R$ from the axis of rotation. In (b), the center of mass of the sphere is located a distance R from the axis of rotation. In both cases, the moment of inertia of the rod is about an axis at one end. Refer to [\[link\]](#) for the moments of inertia for the individual objects.

$$\text{a. } I_{\text{total}} = \sum_i I_i = I_{\text{Rod}} + I_{\text{Sphere}};$$

$$I_{\text{Sphere}} = I_{\text{center of mass}} + m_{\text{Sphere}}(L + R)^2 = \frac{2}{5} m_{\text{Sphere}} R^2 + m_{\text{Sphere}}(L + R)^2;$$

$$I_{\text{total}} = I_{\text{Rod}} + I_{\text{Sphere}} = \frac{1}{3} m_{\text{Rod}} L^2 + \frac{2}{5} m_{\text{Sphere}} R^2 + m_{\text{Sphere}}(L + R)^2;$$

$$I_{\text{total}} = \frac{1}{3}(2.0 \text{ kg})(0.5 \text{ m})^2 + \frac{2}{5}(1.0 \text{ kg})(0.2 \text{ m})^2 + (1.0 \text{ kg})(0.5 \text{ m} + 0.2 \text{ m})^2;$$

$$I_{\text{total}} = (0.167 + 0.016 + 0.490) \text{ kg} \cdot \text{m}^2 = 0.673 \text{ kg} \cdot \text{m}^2.$$

b. $I_{\text{Sphere}} = \frac{2}{5}m_{\text{Sphere}}R^2 + m_{\text{Sphere}}R^2;$

$$I_{\text{total}} = I_{\text{Rod}} + I_{\text{Sphere}} = \frac{1}{3}m_{\text{Rod}}L^2 + \frac{2}{5}m_{\text{Sphere}}R^2 + m_{\text{Sphere}}R^2;$$

$$I_{\text{total}} = \frac{1}{3}(2.0 \text{ kg})(0.5 \text{ m})^2 + \frac{2}{5}(1.0 \text{ kg})(0.2 \text{ m})^2 + (1.0 \text{ kg})(0.2 \text{ m})^2;$$

$$I_{\text{total}} = (0.167 + 0.016 + 0.04) \text{ kg} \cdot \text{m}^2 = 0.223 \text{ kg} \cdot \text{m}^2.$$

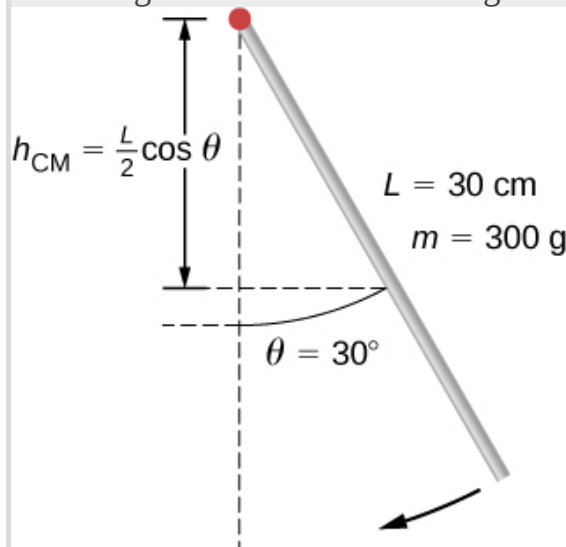
Significance

Using the parallel-axis theorem eases the computation of the moment of inertia of compound objects. We see that the moment of inertia is greater in (a) than (b). This is because the axis of rotation is closer to the center of mass of the system in (b). The simple analogy is that of a rod. The moment of inertia about one end is $\frac{1}{3}mL^2$, but the moment of inertia through the center of mass along its length is $\frac{1}{12}mL^2$.

Example:

Angular Velocity of a Pendulum

A pendulum in the shape of a rod ([link](#)) is released from rest at an angle of 30° . It has a length 30 cm and mass 300 g. What is its angular velocity at its lowest point?



A pendulum in the form of a rod is released from rest at an angle of 30° .

Strategy

Use conservation of energy to solve the problem. At the point of release, the pendulum has gravitational potential energy, which is determined from the height of the center of mass above its lowest point in the swing. At the bottom of the swing, all of the gravitational potential energy is converted into rotational kinetic energy.

Solution

The change in potential energy is equal to the change in rotational kinetic energy, $\Delta U + \Delta K = 0$.

At the top of the swing: $U = mgh_{\text{cm}} = mg\frac{L}{2}(\cos \theta)$. At the bottom of the swing, $U = mg\frac{L}{2}$.

At the top of the swing, the rotational kinetic energy is $K = 0$. At the bottom of the swing, $K = \frac{1}{2}I\omega^2$. Therefore:

Equation:

$$\Delta U + \Delta K = 0 \Rightarrow \left(mg\frac{L}{2}(1 - \cos \theta) - 0\right) + \left(0 - \frac{1}{2}I\omega^2\right) = 0$$

or

Equation:

$$\frac{1}{2}I\omega^2 = mg\frac{L}{2}(1 - \cos \theta).$$

Solving for ω , we have

Equation:

$$\omega = \sqrt{mg\frac{L}{I}(1 - \cos \theta)} = \sqrt{mg\frac{L}{\frac{1}{3}mL^2}(1 - \cos \theta)} = \sqrt{g\frac{3}{L}(1 - \cos \theta)}.$$

Inserting numerical values, we have

Equation:

$$\omega = \sqrt{9.8 \text{ m/s}^2 \frac{3}{0.3 \text{ m}}(1 - \cos 30)} = 3.6 \text{ rad/s}.$$

Significance

Note that the angular velocity of the pendulum does not depend on its mass.

Summary

- Moments of inertia can be found by summing or integrating over every ‘piece of mass’ that makes up an object, multiplied by the square of the distance of each ‘piece of mass’ to the axis. In integral form the moment of inertia is $I = \int r^2 dm$.
- Moment of inertia is larger when an object’s mass is farther from the axis of rotation.
- It is possible to find the moment of inertia of an object about a new axis of rotation once it is known for a parallel axis. This is called the parallel axis theorem given by $I_{\text{parallel-axis}} = I_{\text{center of mass}} + md^2$, where d is the distance from the initial axis to the parallel axis.
- Moment of inertia for a compound object is simply the sum of the moments of inertia for each individual object that makes up the compound object.

Conceptual Questions

Exercise:

Problem:

If a child walks toward the center of a merry-go-round, does the moment of inertia increase or decrease?

Exercise:

Problem:

A discus thrower rotates with a discus in his hand before letting it go. (a) How does his moment of inertia change after releasing the discus? (b) What would be a good approximation to use in calculating the moment of inertia of the discus thrower and discus?

Solution:

a. It decreases. b. The arms could be approximated with rods and the discus with a disk. The torso is near the axis of rotation so it doesn’t contribute much to the moment of inertia.

Exercise:

Problem:

Does increasing the number of blades on a propeller increase or decrease its moment of inertia, and why?

Exercise:

Problem:

The moment of inertia of a long rod spun around an axis through one end perpendicular to its length is $mL^2/3$. Why is this moment of inertia greater than it would be if you spun a point mass m at the location of the center of mass of the rod (at $L/2$) (that would be $mL^2/4$)?

Solution:

Because the moment of inertia varies as the square of the distance to the axis of rotation. The mass of the rod located at distances greater than $L/2$ would provide the larger contribution to make its moment of inertia greater than the point mass at $L/2$.

Exercise:**Problem:**

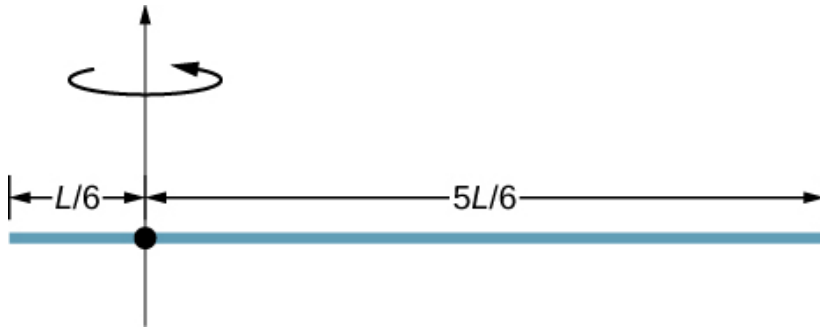
Why is the moment of inertia of a hoop that has a mass M and a radius R greater than the moment of inertia of a disk that has the same mass and radius?

Problems**Exercise:****Problem:**

While punting a football, a kicker rotates his leg about the hip joint. The moment of inertia of the leg is $3.75 \text{ kg}\cdot\text{m}^2$ and its rotational kinetic energy is 175 J . (a) What is the angular velocity of the leg? (b) What is the velocity of tip of the punter's shoe if it is 1.05 m from the hip joint?

Exercise:**Problem:**

Using the parallel axis theorem, what is the moment of inertia of the rod of mass m about the axis shown below?



Solution:

$$I = \frac{7}{36}mL^2$$

Exercise:

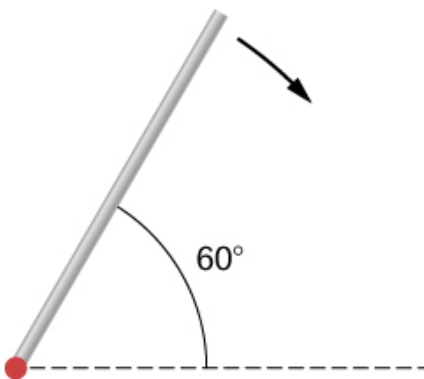
Problem:

Find the moment of inertia of the rod in the previous problem by direct integration.

Exercise:

Problem:

A uniform rod of mass 1.0 kg and length 2.0 m is free to rotate about one end (see the following figure). If the rod is released from rest at an angle of 60° with respect to the horizontal, what is the speed of the tip of the rod as it passes the horizontal position?



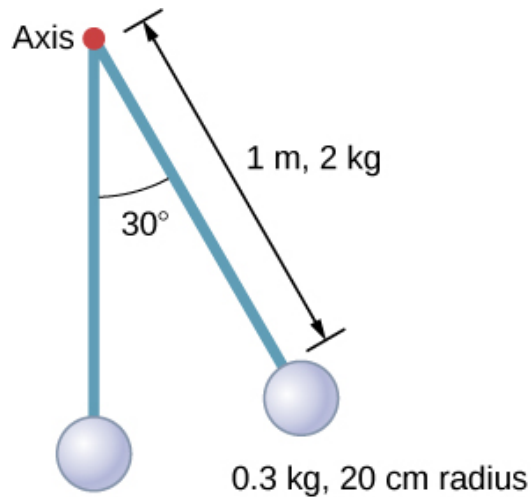
Solution:

$$v = 7.14 \text{ m/s.}$$

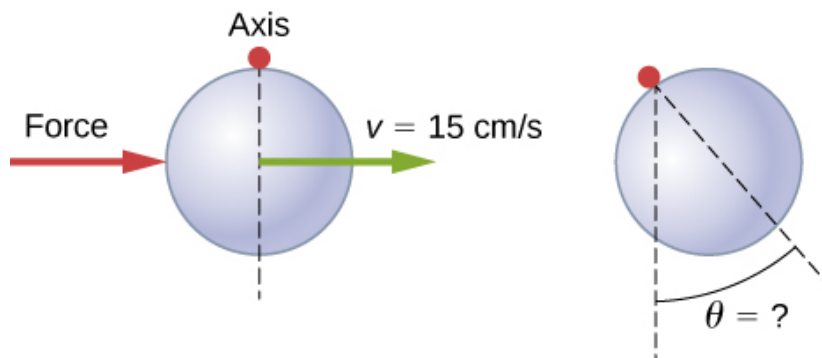
Exercise:

Problem:

A pendulum consists of a rod of mass 2 kg and length 1 m with a solid sphere at one end with mass 0.3 kg and radius 20 cm (see the following figure). If the pendulum is released from rest at an angle of 30° , what is the angular velocity at the lowest point?

**Exercise:****Problem:**

A solid sphere of radius 10 cm is allowed to rotate freely about an axis. The sphere is given a sharp blow so that its center of mass starts from the position shown in the following figure with speed 15 cm/s. What is the maximum angle that the diameter makes with the vertical?

**Solution:**

$$\theta = 10.2^\circ$$

Exercise:

Problem:

Calculate the moment of inertia by direct integration of a thin rod of mass M and length L about an axis through the rod at $L/3$, as shown below. Check your answer with the parallel-axis theorem.



Glossary

linear mass density

the mass per unit length λ of a one dimensional object

parallel axis

axis of rotation that is parallel to an axis about which the moment of inertia of an object is known

parallel-axis theorem

if the moment of inertia is known for a given axis, it can be found for any axis parallel to it

surface mass density

mass per unit area σ of a two dimensional object

Torque

By the end of this section, you will be able to:

- Describe how the magnitude of a torque depends on the magnitude of the lever arm and the angle the force vector makes with the lever arm
- Determine the sign (positive or negative) of a torque using the right-hand rule
- Calculate individual torques about a common axis and sum them to find the net torque

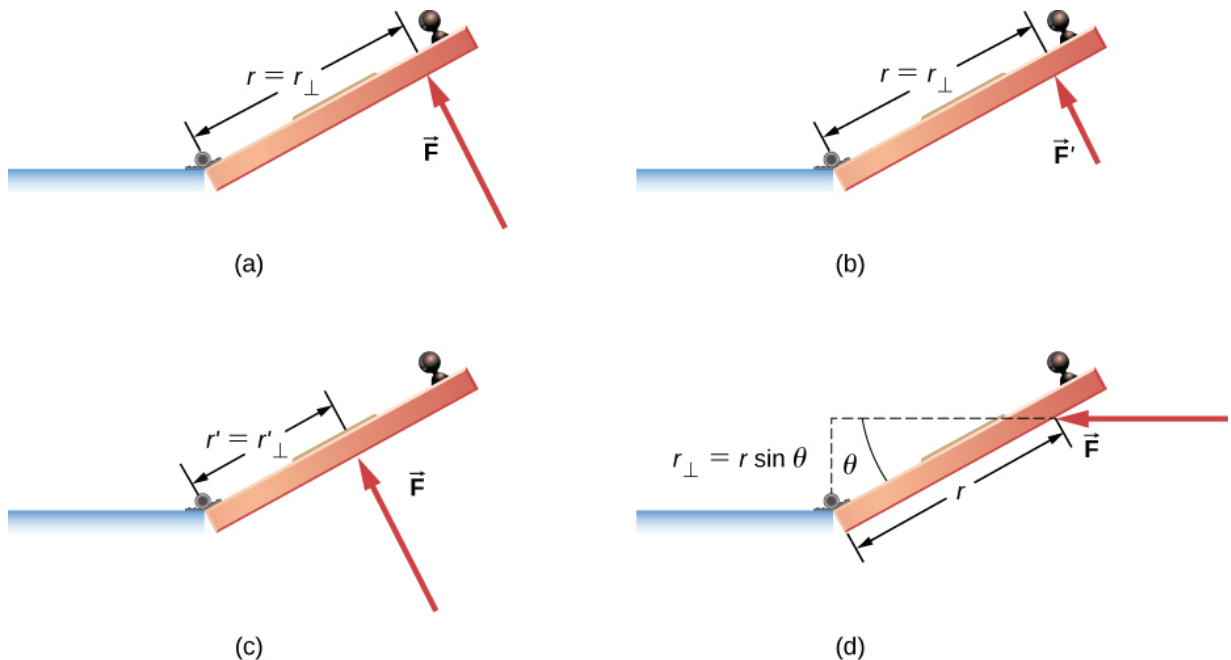
An important quantity for describing the dynamics of a rotating rigid body is torque. We see the application of torque in many ways in our world. We all have an intuition about torque, as when we use a large wrench to unscrew a stubborn bolt. Torque is at work in unseen ways, as when we press on the accelerator in a car, causing the engine to put additional torque on the drive train. Or every time we move our bodies from a standing position, we apply a torque to our limbs. In this section, we define torque and make an argument for the equation for calculating torque for a rigid body with fixed-axis rotation.

Defining Torque

So far we have defined many variables that are rotational equivalents to their translational counterparts. Let's consider what the counterpart to force must be. Since forces change the translational motion of objects, the rotational counterpart must be related to changing the rotational motion of an object about an axis. We call this rotational counterpart **torque**.

In everyday life, we rotate objects about an axis all the time, so intuitively we already know much about torque. Consider, for example, how we rotate a door to open it. First, we know that a door opens slowly if we push too close to its hinges; it is more efficient to rotate a door open if we push far from the hinges. Second, we know that we should push perpendicular to the plane of the door; if we push parallel to the plane of the door, we are not able to rotate it. Third, the larger the force, the more effective it is in opening the door; the harder you push, the more rapidly the door opens. The

first point implies that the farther the force is applied from the axis of rotation, the greater the angular acceleration; the second implies that the effectiveness depends on the angle at which the force is applied; the third implies that the magnitude of the force must also be part of the equation. Note that for rotation in a plane, torque has two possible directions. Torque is either clockwise or counterclockwise relative to the chosen pivot point. [\[link\]](#) shows counterclockwise rotations.



Torque is the turning or twisting effectiveness of a force, illustrated here for door rotation on its hinges (as viewed from overhead). Torque has both magnitude and direction. (a) A counterclockwise torque is produced by a force \vec{F} acting at a distance r from the hinges (the pivot point). (b) A smaller counterclockwise torque is produced when a smaller force \vec{F}' acts at the same distance r from the hinges. (c) The same force as in (a) produces a smaller counterclockwise torque when applied at a smaller distance from the hinges. (d) A smaller counterclockwise torque is produced by the same magnitude force as (a) acting at the same distance as (a) but at an angle θ that is less than 90° .

Now let's consider how to define torques in the general three-dimensional case.

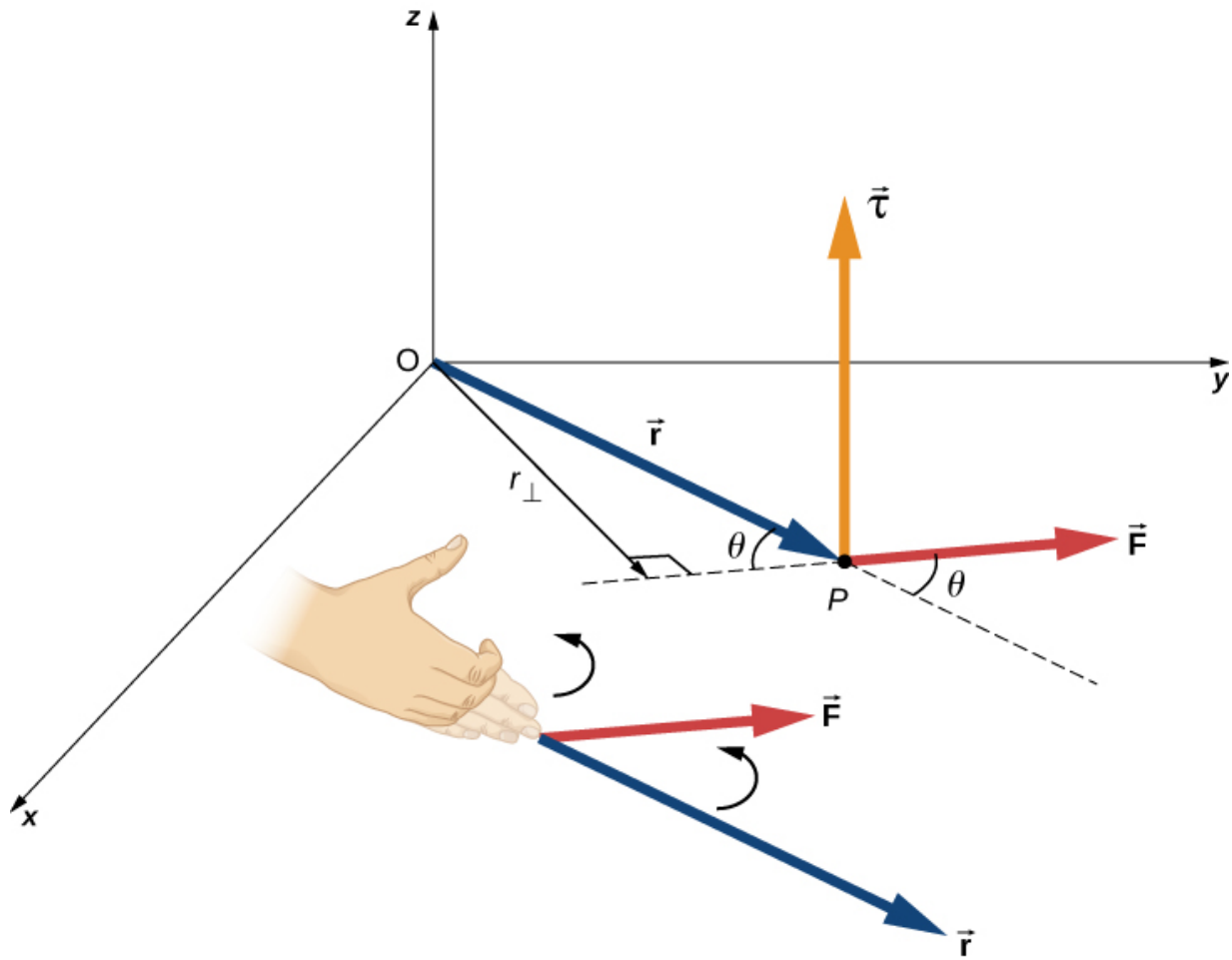
Note:

Torque

When a force $\vec{\mathbf{F}}$ is applied to a point P whose position is $\vec{\mathbf{r}}$ relative to O ([\[link\]](#)), the torque $\vec{\boldsymbol{\tau}}$ around O is

Equation:

$$\vec{\boldsymbol{\tau}} = \vec{\mathbf{r}} \times \vec{\mathbf{F}}.$$



The torque is perpendicular to the plane defined by \vec{r} and \vec{F} and its direction is determined by the right-hand rule.

From the definition of the cross product, the torque $\vec{\tau}$ is perpendicular to the plane containing \vec{r} and \vec{F} and has magnitude

Equation:

$$|\vec{\tau}| = |\vec{r} \times \vec{F}| = rF \sin \theta,$$

where θ is the angle between the vectors \vec{r} and \vec{F} . The SI unit of torque is newtons times meters, usually written as $\text{N} \cdot \text{m}$. The quantity $r_{\perp} = r \sin \theta$

is the perpendicular distance from O to the line determined by the vector $\vec{\mathbf{F}}$ and is called the **lever arm**. Note that the greater the lever arm, the greater the magnitude of the torque. In terms of the lever arm, the magnitude of the torque is

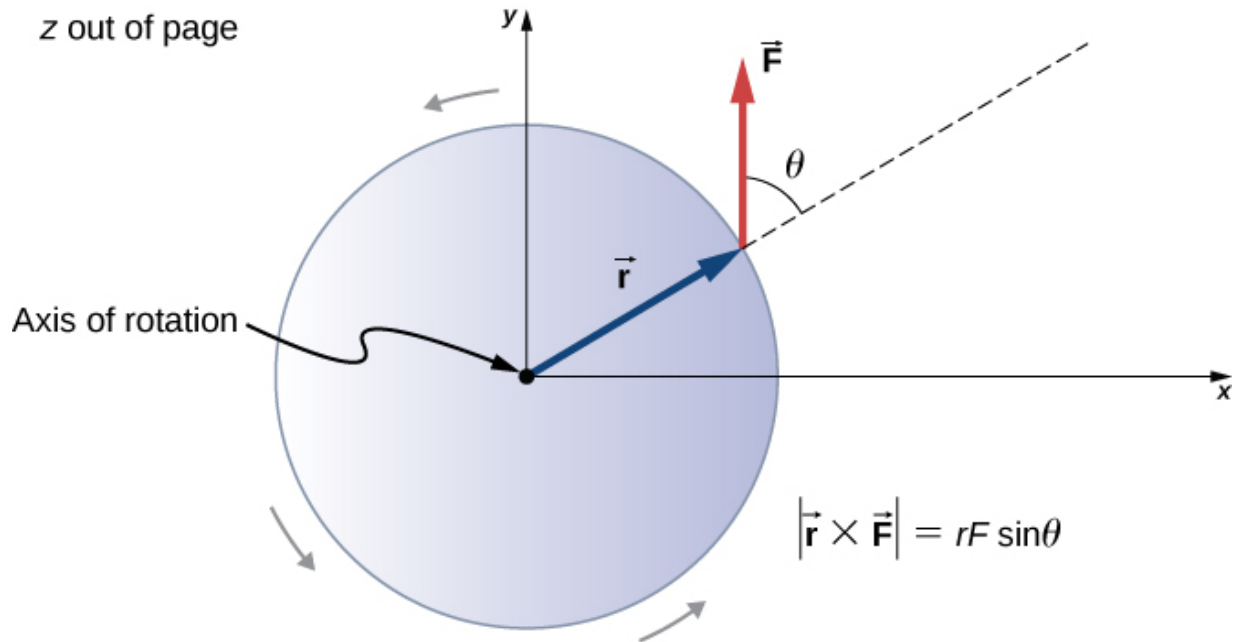
Note:

Equation:

$$|\vec{\tau}| = r_{\perp} F.$$

The cross product $\vec{\mathbf{r}} \times \vec{\mathbf{F}}$ also tells us the sign of the torque. In [\[link\]](#), the cross product $\vec{\mathbf{r}} \times \vec{\mathbf{F}}$ is along the positive z -axis, which by convention is a positive torque. If $\vec{\mathbf{r}} \times \vec{\mathbf{F}}$ is along the negative z -axis, this produces a negative torque.

If we consider a disk that is free to rotate about an axis through the center, as shown in [\[link\]](#), we can see how the angle between the radius $\vec{\mathbf{r}}$ and the force $\vec{\mathbf{F}}$ affects the magnitude of the torque. If the angle is zero, the torque is zero; if the angle is 90° , the torque is maximum. The torque in [\[link\]](#) is positive because the direction of the torque by the right-hand rule is out of the page along the positive z -axis. The disk rotates counterclockwise due to the torque, in the same direction as a positive angular acceleration.



A disk is free to rotate about its axis through the center. The magnitude of the torque on the disk is $rF \sin \theta$. When $\theta = 0^\circ$, the torque is zero and the disk does not rotate. When $\theta = 90^\circ$, the torque is maximum and the disk rotates with maximum angular acceleration.

Any number of torques can be calculated about a given axis. The individual torques add to produce a net torque about the axis. When the appropriate sign (positive or negative) is assigned to the magnitudes of individual torques about a specified axis, the net torque about the axis is the sum of the individual torques:

Note:

Equation:

$$\vec{\tau}_{\text{net}} = \sum_i |\vec{\tau}_i|.$$

Calculating Net Torque for Rigid Bodies on a Fixed Axis

In the following examples, we calculate the torque both abstractly and as applied to a rigid body.

We first introduce a problem-solving strategy.

Note:

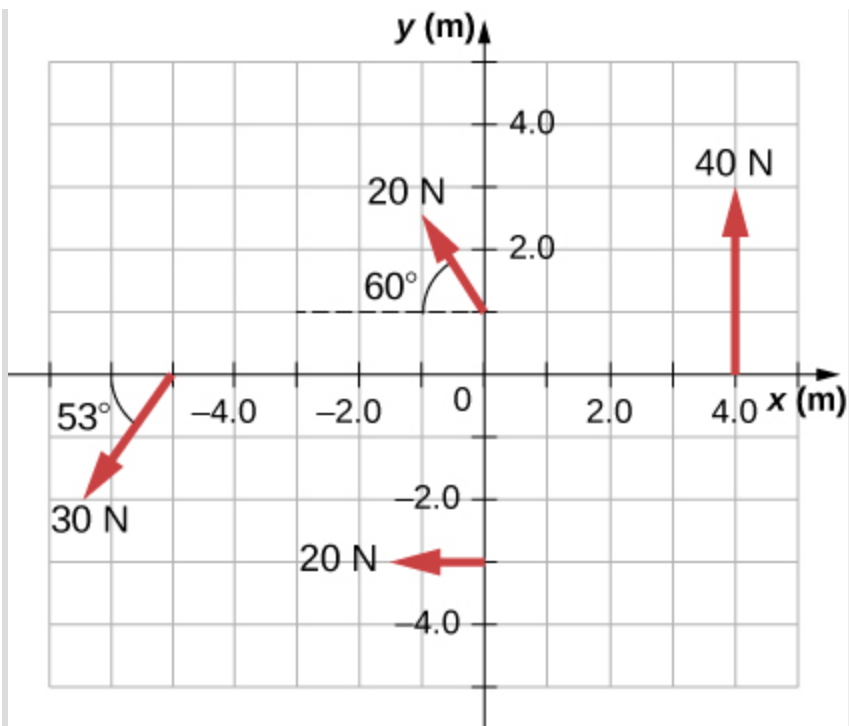
Finding Net Torque

1. Choose a coordinate system with the pivot point or axis of rotation as the origin of the selected coordinate system.
2. Determine the angle between the lever arm \vec{r} and the force vector.
3. Take the cross product of \vec{r} and \vec{F} to determine if the torque is positive or negative about the pivot point or axis.
4. Evaluate the magnitude of the torque using $r_{\perp} F$.
5. Assign the appropriate sign, positive or negative, to the magnitude.
6. Sum the torques to find the net torque.

Example:

Calculating Torque

Four forces are shown in [\[link\]](#) at particular locations and orientations with respect to a given xy-coordinate system. Find the torque due to each force about the origin, then use your results to find the net torque about the origin.



Four forces producing torques.

Strategy

This problem requires calculating torque. All known quantities—forces with directions and lever arms—are given in the figure. The goal is to find each individual torque and the net torque by summing the individual torques. Be careful to assign the correct sign to each torque by using the cross product of \vec{r} and the force vector \vec{F} .

Solution

Use $|\vec{\tau}| = r_{\perp} F = r F \sin \theta$ to find the magnitude and $\vec{\tau} = \vec{r} \times \vec{F}$ to determine the sign of the torque.

The torque from force 40 N in the first quadrant is given by $(4)(40)\sin 90^{\circ} = 160 \text{ N} \cdot \text{m}$.

The cross product of \vec{r} and \vec{F} is out of the page, positive.

The torque from force 20 N in the third quadrant is given by $-(3)(20)\sin 90^{\circ} = -60 \text{ N} \cdot \text{m}$.

The cross product of \vec{r} and \vec{F} is into the page, so it is negative.

The torque from force 30 N in the third quadrant is given by $(5)(30)\sin 53^\circ = 120 \text{ N} \cdot \text{m}$.

The cross product of \vec{r} and \vec{F} is out of the page, positive.

The torque from force 20 N in the second quadrant is given by $(1)(20)\sin 30^\circ = 10 \text{ N} \cdot \text{m}$.

The cross product of \vec{r} and \vec{F} is out of the page.

The net torque is therefore

$$\tau_{\text{net}} = \sum_i |\tau_i| = 160 - 60 + 120 + 10 = 230 \text{ N} \cdot \text{m}.$$

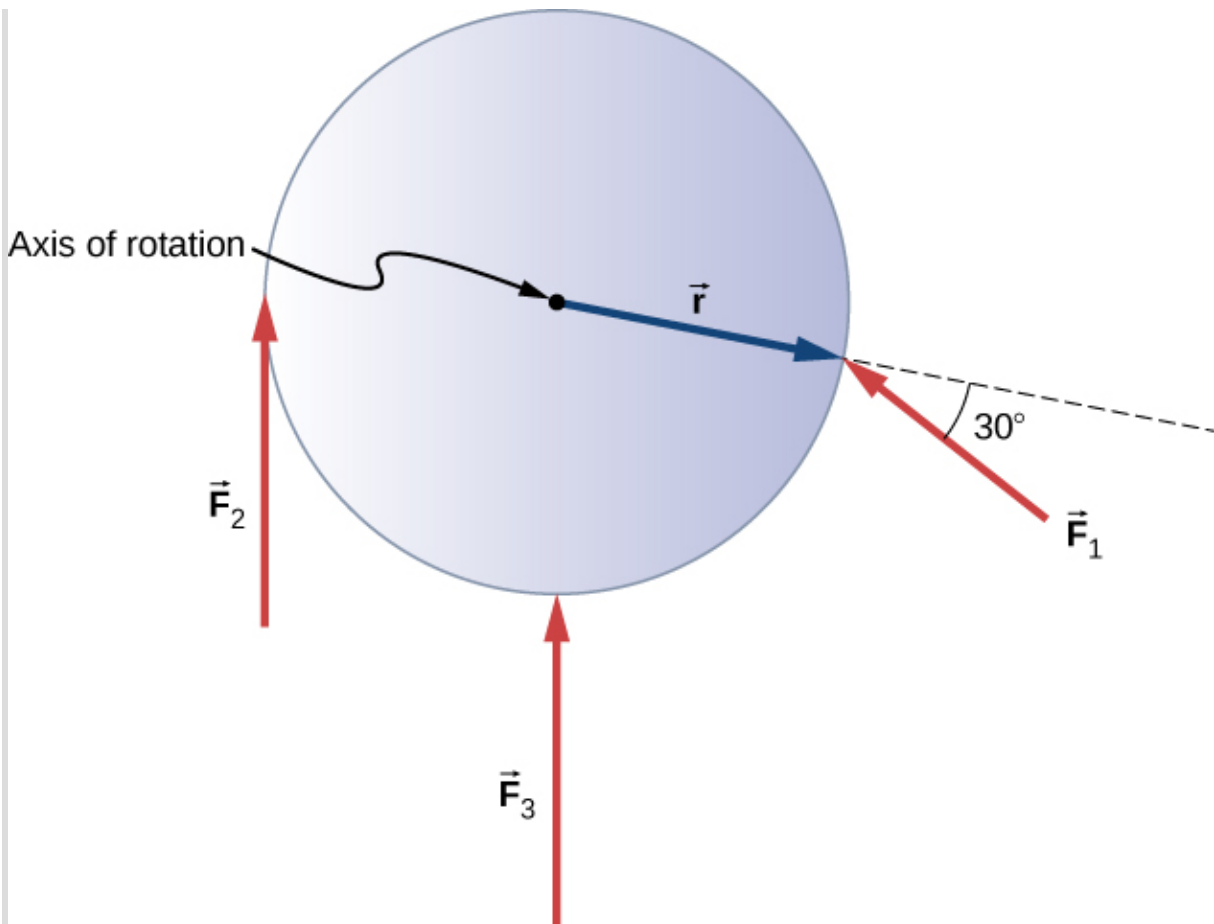
Significance

Note that each force that acts in the counterclockwise direction has a positive torque, whereas each force that acts in the clockwise direction has a negative torque. The torque is greater when the distance, force, or perpendicular components are greater.

Example:

Calculating Torque on a rigid body

[\[link\]](#) shows several forces acting at different locations and angles on a flywheel. We have $|\vec{F}_1| = 20 \text{ N}$, $|\vec{F}_2| = 30 \text{ N}$, $|\vec{F}_3| = 30 \text{ N}$, and $r = 0.5 \text{ m}$. Find the net torque on the flywheel about an axis through the center.



Three forces acting on a flywheel.

Strategy

We calculate each torque individually, using the cross product, and determine the sign of the torque. Then we sum the torques to find the net torque.

Solution

We start with \vec{F}_1 . If we look at [\[link\]](#), we see that \vec{F}_1 makes an angle of $90^\circ + 60^\circ$ with the radius vector \vec{r} . Taking the cross product, we see that it is out of the page and so is positive. We also see this from calculating its magnitude:

Equation:

$$|\vec{\tau}_1| = rF_1 \sin 150^\circ = 0.5 \text{ m}(20 \text{ N})(0.5) = 5.0 \text{ N} \cdot \text{m}.$$

Next we look at \vec{F}_2 . The angle between \vec{F}_2 and \vec{r} is 90° and the cross product is into the page so the torque is negative. Its value is

Equation:

$$|\vec{\tau}_2| = -rF_2\sin 90^\circ = -0.5\text{ m}(30\text{ N}) = -15.0\text{ N}\cdot\text{m}.$$

When we evaluate the torque due to \vec{F}_3 , we see that the angle it makes with \vec{r} is zero so $\vec{r} \times \vec{F}_3 = 0$. Therefore, \vec{F}_3 does not produce any torque on the flywheel.

We evaluate the sum of the torques:

Equation:

$$\tau_{\text{net}} = \sum_i |\tau_i| = 5 - 15 = -10\text{ N}\cdot\text{m}.$$

Significance

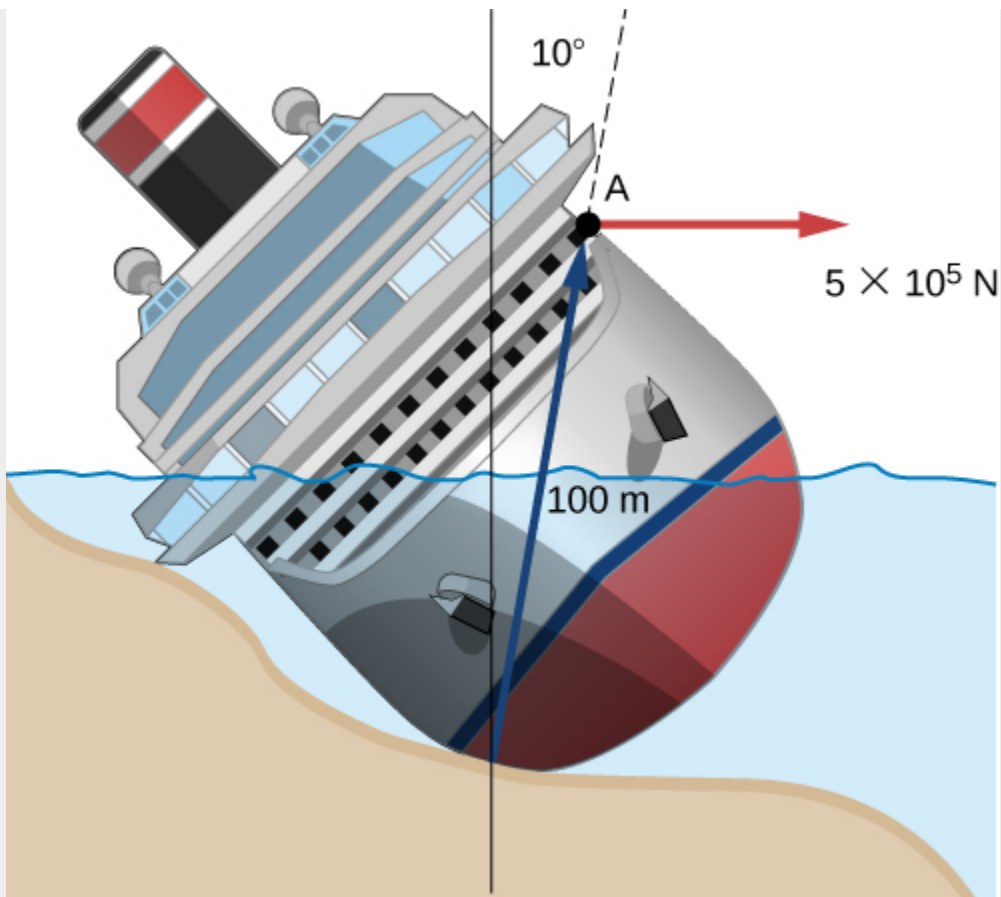
The axis of rotation is at the center of mass of the flywheel. Since the flywheel is on a fixed axis, it is not free to translate. If it were on a frictionless surface and not fixed in place, \vec{F}_3 would cause the flywheel to translate, as well as \vec{F}_1 . Its motion would be a combination of translation and rotation.

Note:

Exercise:

Problem:

Check Your Understanding A large ocean-going ship runs aground near the coastline, similar to the fate of the *Costa Concordia*, and lies at an angle as shown below. Salvage crews must apply a torque to right the ship in order to float the vessel for transport. A force of $5.0 \times 10^5\text{ N}$ acting at point A must be applied to right the ship. What is the torque about the point of contact of the ship with the ground ([link](#))?



A ship runs aground and tilts, requiring torque to be applied to return the vessel to an upright position.

Solution:

The angle between the lever arm and the force vector is 80° ; therefore, $r_\perp = 100\text{m}(\sin 80^\circ) = 98.5 \text{ m}$.

The cross product $\vec{\tau} = \vec{r} \times \vec{F}$ gives a negative or clockwise torque.

The torque is then

$$\tau = -r_\perp F = -98.5 \text{ m}(5.0 \times 10^5 \text{ N}) = -4.9 \times 10^7 \text{ N} \cdot \text{m}.$$

Summary

- The magnitude of a torque about a fixed axis is calculated by finding the lever arm to the point where the force is applied and using the relation $|\vec{\tau}| = r_{\perp} F$, where r_{\perp} is the perpendicular distance from the axis to the line upon which the force vector lies.
- The sign of the torque is found using the right hand rule. If the page is the plane containing \vec{r} and \vec{F} , then $\vec{r} \times \vec{F}$ is out of the page for positive torques and into the page for negative torques.
- The net torque can be found from summing the individual torques about a given axis.

Conceptual Questions

Exercise:

Problem:

What three factors affect the torque created by a force relative to a specific pivot point?

Solution:

magnitude of the force, length of the lever arm, and angle of the lever arm and force vector

Exercise:

Problem:

Give an example in which a small force exerts a large torque. Give another example in which a large force exerts a small torque.

Exercise:

Problem:

When reducing the mass of a racing bike, the greatest benefit is realized from reducing the mass of the tires and wheel rims. Why does this allow a racer to achieve greater accelerations than would an identical reduction in the mass of the bicycle's frame?

Solution:

The moment of inertia of the wheels is reduced, so a smaller torque is needed to accelerate them.

Exercise:

Problem: Can a single force produce a zero torque?

Exercise:**Problem:**

Can a set of forces have a net torque that is zero and a net force that is not zero?

Solution:

yes

Exercise:**Problem:**

Can a set of forces have a net force that is zero and a net torque that is not zero?

Exercise:**Problem:**

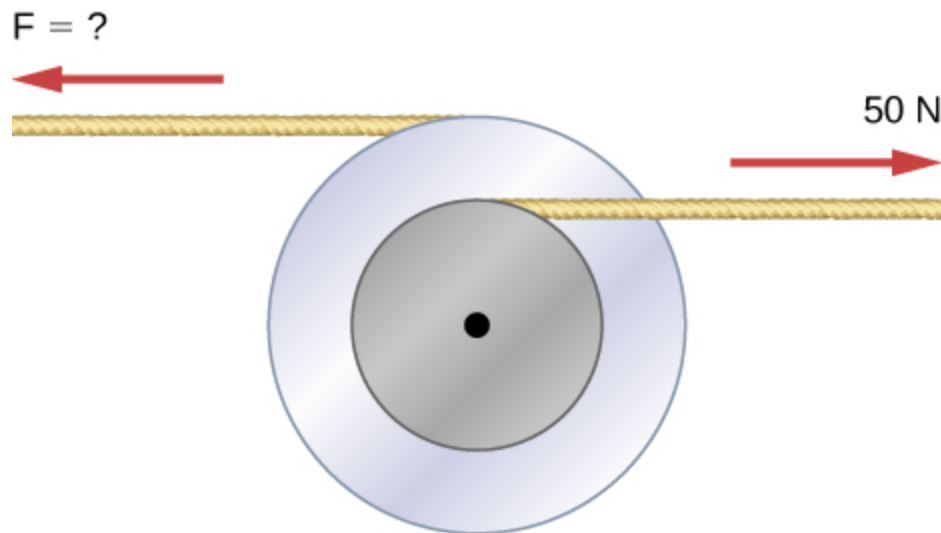
In the expression $\vec{r} \times \vec{F}$ can $|\vec{r}|$ ever be less than the lever arm? Can it be equal to the lever arm?

Solution:

$|\vec{r}|$ can be equal to the lever arm but never less than the lever arm

Problems**Exercise:****Problem:**

Two flywheels of negligible mass and different radii are bonded together and rotate about a common axis (see below). The smaller flywheel of radius 30 cm has a cord that has a pulling force of 50 N on it. What pulling force needs to be applied to the cord connecting the larger flywheel of radius 50 cm such that the combination does not rotate?



Solution:

$$F = 30 \text{ N}$$

Exercise:

Problem:

The cylinder head bolts on a car are to be tightened with a torque of $62.0 \text{ N}\cdot\text{m}$. If a mechanic uses a wrench of length 20 cm , what perpendicular force must he exert on the end of the wrench to tighten a bolt correctly?

Exercise:**Problem:**

(a) When opening a door, you push on it perpendicularly with a force of 55.0 N at a distance of 0.850 m from the hinges. What torque are you exerting relative to the hinges? (b) Does it matter if you push at the same height as the hinges? There is only one pair of hinges.

Solution:

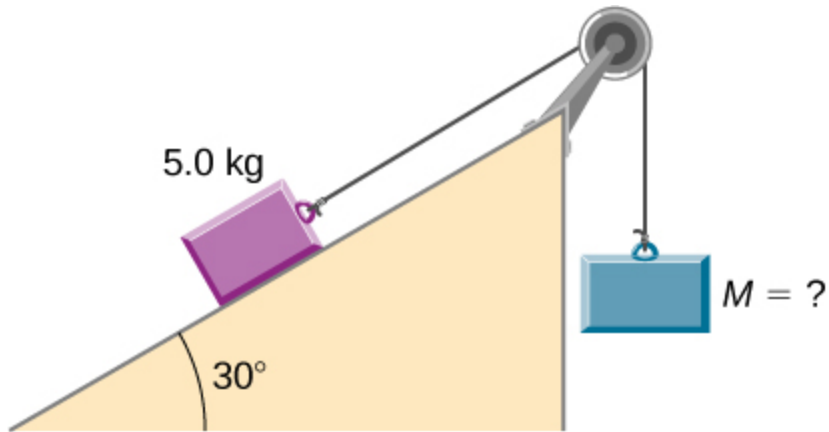
a. $0.85 \text{ m} (55.0 \text{ N}) = 46.75 \text{ N}\cdot\text{m}$; b. It does not matter at what height you push.

Exercise:**Problem:**

When tightening a bolt, you push perpendicularly on a wrench with a force of 165 N at a distance of 0.140 m from the center of the bolt. How much torque are you exerting in newton-meters (relative to the center of the bolt)?

Exercise:**Problem:**

What hanging mass must be placed on the cord to keep the pulley from rotating (see the following figure)? The mass on the frictionless plane is 5.0 kg . The inner radius of the pulley is 20 cm and the outer radius is 30 cm .



Solution:

$$m_2 = \frac{4.9 \text{ N}\cdot\text{m}}{9.8(0.3 \text{ m})} = 1.67 \text{ kg}$$

Exercise:

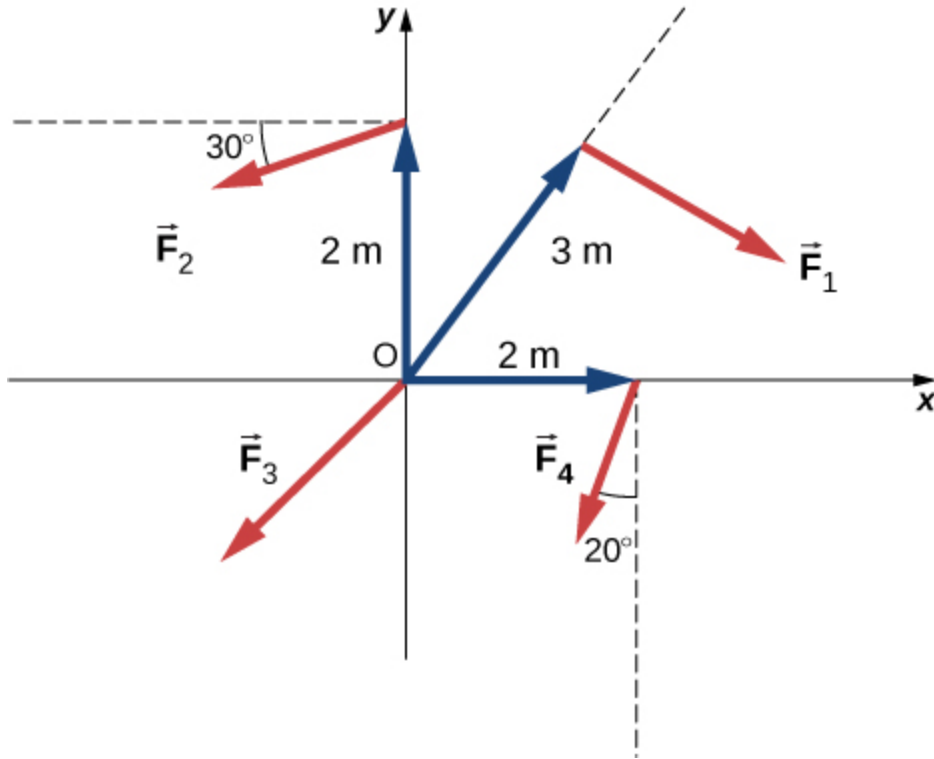
Problem:

A simple pendulum consists of a massless tether 50 cm in length connected to a pivot and a small mass of 1.0 kg attached at the other end. What is the torque about the pivot when the pendulum makes an angle of 40° with respect to the vertical?

Exercise:

Problem:

Calculate the torque about the z-axis that is out of the page at the origin in the following figure, given that $F_1 = 3 \text{ N}$, $F_2 = 2 \text{ N}$, $F_3 = 3 \text{ N}$, $F_4 = 1.8 \text{ N}$.



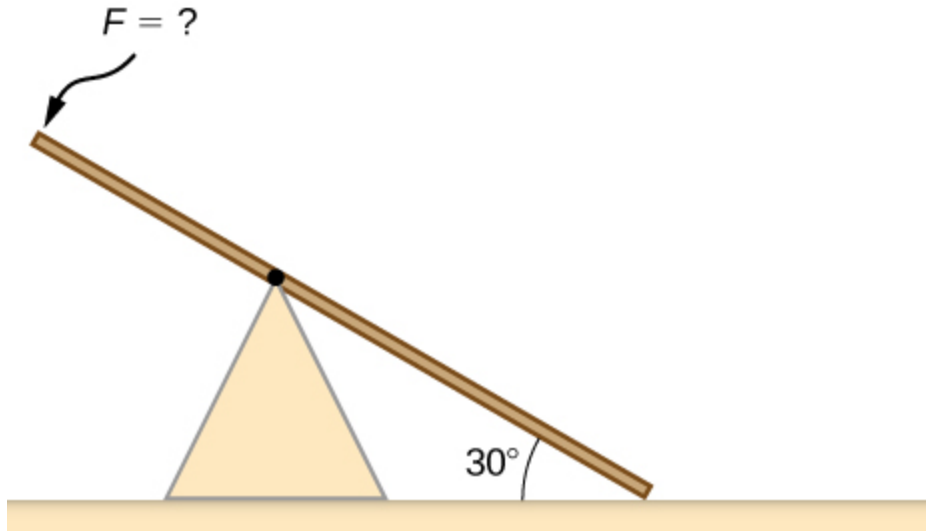
Solution:

$$\tau_{net} = -9.0 \text{ N} \cdot \text{m} + 3.46 \text{ N} \cdot \text{m} + 0 - 3.38 \text{ N} \cdot \text{m} = -8.92 \text{ N} \cdot \text{m}$$

Exercise:

Problem:

A seesaw has length 10.0 m and uniform mass 10.0 kg and is resting at an angle of 30° with respect to the ground (see the following figure). The pivot is located at 6.0 m . What magnitude of force needs to be applied perpendicular to the seesaw at the raised end so as to allow the seesaw to barely start to rotate?



Exercise:

Problem:

A pendulum consists of a rod of mass 1 kg and length 1 m connected to a pivot with a solid sphere attached at the other end with mass 0.5 kg and radius 30 cm. What is the torque about the pivot when the pendulum makes an angle of 30° with respect to the vertical?

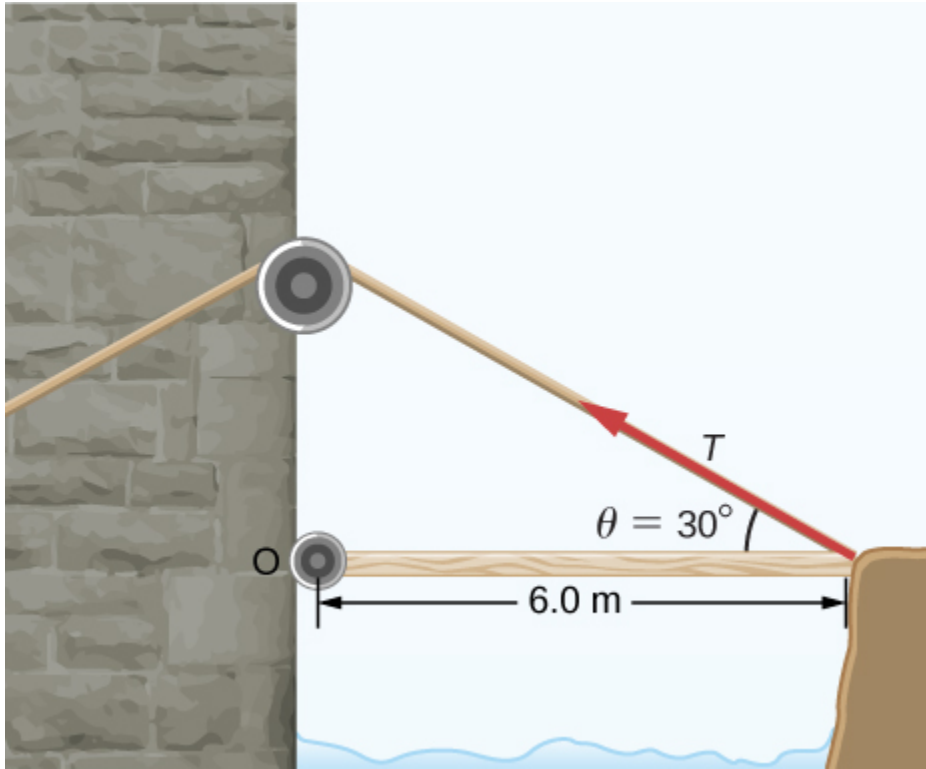
Solution:

$$\tau = 5.66 \text{ N} \cdot \text{m}$$

Exercise:

Problem:

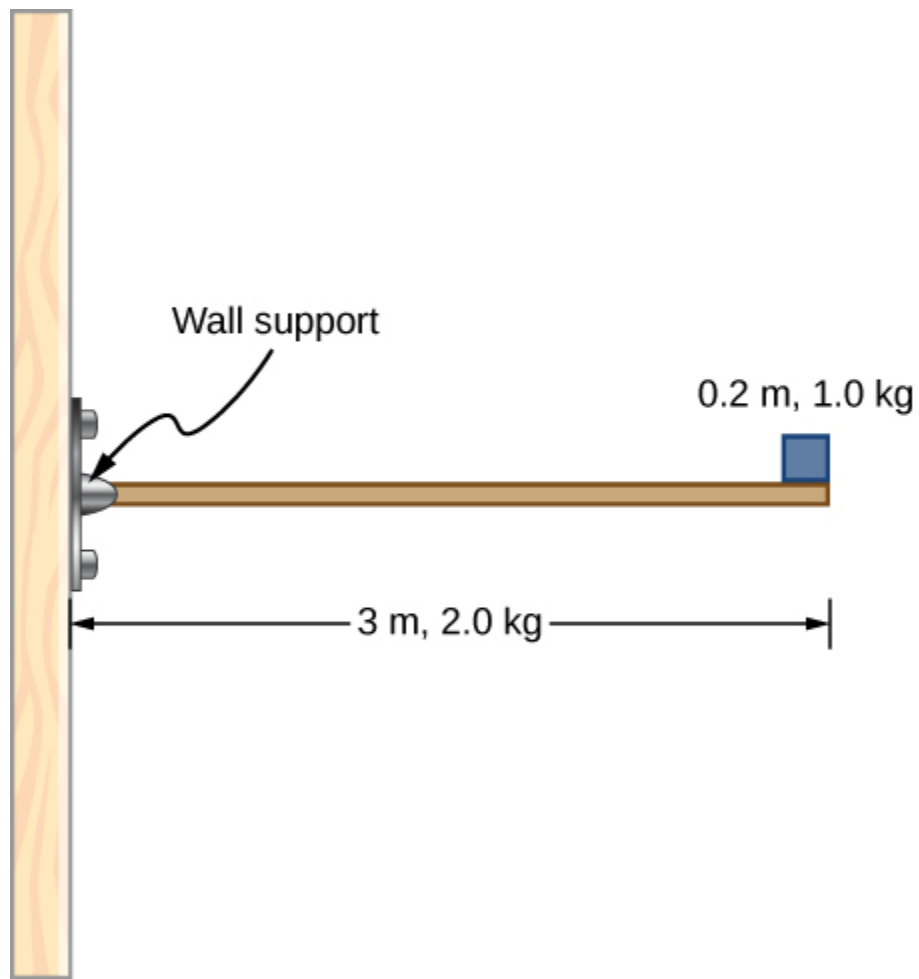
A torque of $5.00 \times 10^3 \text{ N} \cdot \text{m}$ is required to raise a drawbridge (see the following figure). What is the tension necessary to produce this torque? Would it be easier to raise the drawbridge if the angle θ were larger or smaller?



Exercise:

Problem:

A horizontal beam of length 3 m and mass 2.0 kg has a mass of 1.0 kg and width 0.2 m sitting at the end of the beam (see the following figure). What is the torque of the system about the support at the wall?



Solution:

$$\sum \tau = 57.82 \text{ N} \cdot \text{m}$$

Exercise:

Problem:

What force must be applied to end of a rod along the x -axis of length 2.0 m in order to produce a torque on the rod about the origin of $8.0\hat{\mathbf{k}} \text{ N} \cdot \text{m}$?

Exercise:

Problem:

What is the torque about the origin of the force

$(5.0\hat{\mathbf{i}} - 2.0\hat{\mathbf{j}} + 1.0\hat{\mathbf{k}})$ N if it is applied at the point whose position is:

$\vec{\mathbf{r}} = (-2.0\hat{\mathbf{i}} + 4.0\hat{\mathbf{j}})$ m?

Solution:

$$\vec{\mathbf{r}} \times \vec{\mathbf{F}} = 4.0\hat{\mathbf{i}} + 2.0\hat{\mathbf{j}} - 16.0\hat{\mathbf{k}} \text{ N} \cdot \text{m}$$

Glossary

lever arm

perpendicular distance from the line that the force vector lies on to a given axis

torque

cross product of a force and a lever arm to a given axis

Newton's Second Law for Rotation

By the end of this section, you will be able to:

- Calculate the torques on rotating systems about a fixed axis to find the angular acceleration
- Explain how changes in the moment of inertia of a rotating system affect angular acceleration with a fixed applied torque

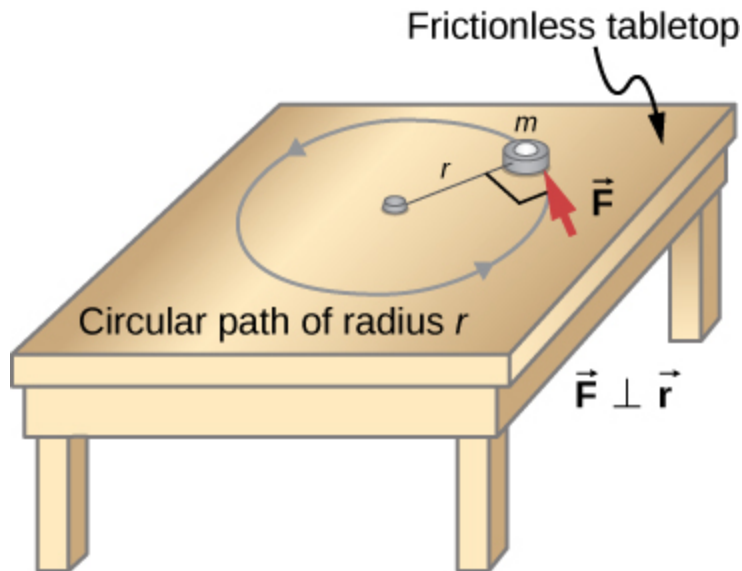
In this section, we put together all the pieces learned so far in this chapter to analyze the dynamics of rotating rigid bodies. We have analyzed motion with kinematics and rotational kinetic energy but have not yet connected these ideas with force and/or torque. In this section, we introduce the rotational equivalent to Newton's second law of motion and apply it to rigid bodies with fixed-axis rotation.

Newton's Second Law for Rotation

We have thus far found many counterparts to the translational terms used throughout this text, most recently, torque, the rotational analog to force. This raises the question: Is there an analogous equation to Newton's second law, $\Sigma \vec{F} = m\vec{a}$, which involves torque and rotational motion? To investigate this, we start with Newton's second law for a single particle rotating around an axis and executing circular motion. Let's exert a force \vec{F} on a point mass m that is at a distance r from a pivot point ([link](#)). The particle is constrained to move in a circular path with fixed radius and the force is tangent to the circle. We apply Newton's second law to determine the magnitude of the acceleration $a = F/m$ in the direction of \vec{F} . Recall that the magnitude of the tangential acceleration is proportional to the magnitude of the angular acceleration by $a = r\alpha$. Substituting this expression into Newton's second law, we obtain

Equation:

$$F = mr\alpha.$$



An object is supported by a horizontal frictionless table and is attached to a pivot point by a cord that supplies centripetal force. A force \vec{F} is applied to the object perpendicular to the radius r , causing it to accelerate about the pivot point. The force is perpendicular to r .

Multiply both sides of this equation by r ,

Equation:

$$rF = mr^2\alpha.$$

Note that the left side of this equation is the torque about the axis of rotation, where r is the lever arm and F is the force, perpendicular to r . Recall that the moment of inertia for a point particle is $I = mr^2$. The torque applied perpendicularly to the point mass in [\[link\]](#) is therefore

Equation:

$$\tau = I\alpha.$$

The torque on the particle is equal to the moment of inertia about the rotation axis times the angular acceleration. We can generalize this equation to a rigid body rotating about a fixed axis.

Note:

Newton's Second Law for Rotation

If more than one torque acts on a rigid body about a fixed axis, then the sum of the torques equals the moment of inertia times the angular acceleration:

Equation:

$$\sum_i \tau_i = I\alpha.$$

The term $I\alpha$ is a scalar quantity and can be positive or negative (counterclockwise or clockwise) depending upon the sign of the net torque. Remember the convention that counterclockwise angular acceleration is positive. Thus, if a rigid body is rotating clockwise and experiences a positive torque (counterclockwise), the angular acceleration is positive.

[\[link\]](#) is **Newton's second law for rotation** and tells us how to relate torque, moment of inertia, and rotational kinematics. This is called the equation for **rotational dynamics**. With this equation, we can solve a whole class of problems involving force and rotation. It makes sense that the relationship for how much force it takes to rotate a body would include the moment of inertia, since that is the quantity that tells us how easy or hard it is to change the rotational motion of an object.

Deriving Newton's Second Law for Rotation in Vector Form

As before, when we found the angular acceleration, we may also find the torque vector. The second law $\Sigma \vec{\mathbf{F}} = m\vec{\mathbf{a}}$ tells us the relationship between

net force and how to change the translational motion of an object. We have a vector rotational equivalent of this equation, which can be found by using [\[link\]](#) and [\[link\]](#). [\[link\]](#) relates the angular acceleration to the position and tangential acceleration vectors:

Equation:

$$\vec{a} = \vec{\alpha} \times \vec{r}.$$

We form the cross product of this equation with \vec{r} and use a cross product identity (note that $\vec{r} \cdot \vec{\alpha} = 0$):

Equation:

$$\vec{r} \times \vec{a} = \vec{r} \times (\vec{\alpha} \times \vec{r}) = \vec{\alpha}(\vec{r} \cdot \vec{r}) - \vec{r}(\vec{r} \cdot \vec{\alpha}) = \vec{\alpha}(\vec{r} \cdot \vec{r}) = \vec{\alpha}r^2.$$

We now form the cross product of Newton's second law with the position vector \vec{r} ,

Equation:

$$\Sigma(\vec{r} \times \vec{F}) = \vec{r} \times (m\vec{a}) = m\vec{r} \times \vec{a} = mr^2\vec{\alpha}.$$

Identifying the first term on the left as the sum of the torques, and mr^2 as the moment of inertia, we arrive at Newton's second law of rotation in vector form:

Equation:

$$\Sigma\vec{\tau} = I\vec{\alpha}.$$

This equation is exactly [\[link\]](#) but with the torque and angular acceleration as vectors. An important point is that the torque vector is in the same direction as the angular acceleration.

Applying the Rotational Dynamics Equation

Before we apply the rotational dynamics equation to some everyday situations, let's review a general problem-solving strategy for use with this category of problems.

Note:

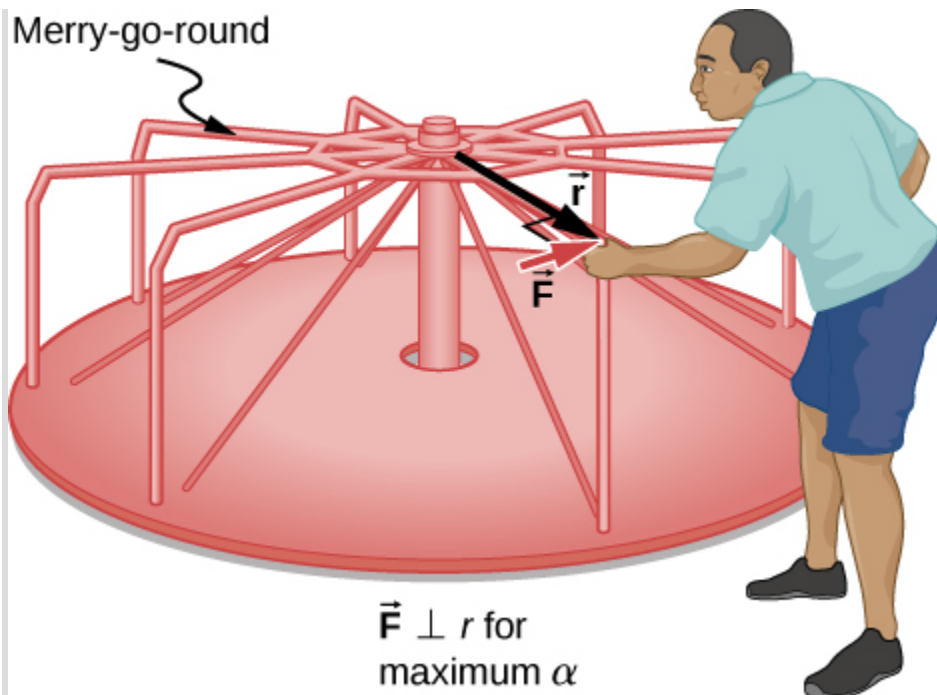
Rotational Dynamics

1. Examine the situation to determine that torque and mass are involved in the rotation. Draw a careful sketch of the situation.
2. Determine the system of interest.
3. Draw a free-body diagram. That is, draw and label all external forces acting on the system of interest.
4. Identify the pivot point. If the object is in equilibrium, it must be in equilibrium for all possible pivot points—choose the one that simplifies your work the most.
5. Apply $\sum_i \tau_i = I\alpha$, the rotational equivalent of Newton's second law, to solve the problem. Care must be taken to use the correct moment of inertia and to consider the torque about the point of rotation.
6. As always, check the solution to see if it is reasonable.

Example:

Calculating the Effect of Mass Distribution on a Merry-Go-Round

Consider the father pushing a playground merry-go-round in [\[link\]](#). He exerts a force of 250 N at the edge of the 50.0-kg merry-go-round, which has a 1.50-m radius. Calculate the angular acceleration produced (a) when no one is on the merry-go-round and (b) when an 18.0-kg child sits 1.25 m away from the center. Consider the merry-go-round itself to be a uniform disk with negligible friction.



A father pushes a playground merry-go-round at its edge and perpendicular to its radius to achieve maximum torque.

Strategy

The net torque is given directly by the expression $\sum_i \tau_i = I\alpha$. To solve for α , we must first calculate the net torque τ (which is the same in both cases) and moment of inertia I (which is greater in the second case).

Solution

- The moment of inertia of a solid disk about this axis is given in [\[link\]](#) to be

Equation:

$$\frac{1}{2}MR^2.$$

We have $M = 50.0$ kg and $R = 1.50$ m, so

Equation:

$$I = (0.500)(50.0 \text{ kg})(1.50 \text{ m})^2 = 56.25 \text{ kg}\cdot\text{m}^2.$$

To find the net torque, we note that the applied force is perpendicular to the radius and friction is negligible, so that

Equation:

$$\tau = rF\sin\theta = (1.50 \text{ m})(250.0 \text{ N}) = 375.0 \text{ N}\cdot\text{m}.$$

Now, after we substitute the known values, we find the angular acceleration to be

Equation:

$$\alpha = \frac{\tau}{I} = \frac{375.0 \text{ N}\cdot\text{m}}{56.25 \text{ kg}\cdot\text{m}^2} = 6.67 \frac{\text{rad}}{\text{s}^2}.$$

- b. We expect the angular acceleration for the system to be less in this part because the moment of inertia is greater when the child is on the merry-go-round. To find the total moment of inertia I , we first find the child's moment of inertia I_c by approximating the child as a point mass at a distance of 1.25 m from the axis. Then

Equation:

$$I_c = mR^2 = (18.0 \text{ kg})(1.25 \text{ m})^2 = 28.13 \text{ kg}\cdot\text{m}^2.$$

The total moment of inertia is the sum of the moments of inertia of the merry-go-round and the child (about the same axis):

Equation:

$$I = 28.13 \text{ kg}\cdot\text{m}^2 + 56.25 \text{ kg}\cdot\text{m}^2 = 84.38 \text{ kg}\cdot\text{m}^2.$$

Substituting known values into the equation for α gives

Equation:

$$\alpha = \frac{\tau}{I} = \frac{375.0 \text{ N}\cdot\text{m}}{84.38 \text{ kg}\cdot\text{m}^2} = 4.44 \frac{\text{rad}}{\text{s}^2}.$$

Significance

The angular acceleration is less when the child is on the merry-go-round than when the merry-go-round is empty, as expected. The angular accelerations found are quite large, partly due to the fact that friction was considered to be negligible. If, for example, the father kept pushing perpendicularly for 2.00 s, he would give the merry-go-round an angular velocity of 13.3 rad/s when it is empty but only 8.89 rad/s when the child is on it. In terms of revolutions per second, these angular velocities are 2.12 rev/s and 1.41 rev/s, respectively. The father would end up running at about 50 km/h in the first case.

Note:

Exercise:

Problem:

Check Your Understanding The fan blades on a jet engine have a moment of inertia $30.0 \text{ kg}\cdot\text{m}^2$. In 10 s, they rotate counterclockwise from rest up to a rotation rate of 20 rev/s. (a) What torque must be applied to the blades to achieve this angular acceleration? (b) What is the torque required to bring the fan blades rotating at 20 rev/s to a rest in 20 s?

Solution:

a. The angular acceleration is $\alpha = \frac{20.0(2\pi)\text{rad/s}-0}{10.0 \text{ s}} = 12.56 \text{ rad/s}^2$.

Solving for the torque, we have

$$\sum_i \tau_i = I\alpha = (30.0 \text{ kg}\cdot\text{m}^2)(12.56 \text{ rad/s}^2) = 376.80 \text{ N}\cdot\text{m}; \text{ b.}$$

The angular acceleration is $\alpha = \frac{0-20.0(2\pi)\text{rad/s}}{20.0 \text{ s}} = -6.28 \text{ rad/s}^2$.

Solving for the torque, we have

$$\sum_i \tau_i = I\alpha = (30.0 \text{ kg}\cdot\text{m}^2)(-6.28 \text{ rad/s}^2) = -188.50 \text{ N}\cdot\text{m}$$

Summary

- Newton's second law for rotation, $\sum_i \tau_i = I\alpha$, says that the sum of the torques on a rotating system about a fixed axis equals the product of the moment of inertia and the angular acceleration. This is the rotational analog to Newton's second law of linear motion.
- In the vector form of Newton's second law for rotation, the torque vector $\vec{\tau}$ is in the same direction as the angular acceleration $\vec{\alpha}$. If the angular acceleration of a rotating system is positive, the torque on the system is also positive, and if the angular acceleration is negative, the torque is negative.

Conceptual Questions

Exercise:

Problem:

If you were to stop a spinning wheel with a constant force, where on the wheel would you apply the force to produce the maximum negative acceleration?

Exercise:

Problem:

A rod is pivoted about one end. Two forces \vec{F} and $-\vec{F}$ are applied to it. Under what circumstances will the rod not rotate?

Solution:

If the forces are along the axis of rotation, or if they have the same lever arm and are applied at a point on the rod.

Problems

Exercise:**Problem:**

You have a grindstone (a disk) that is 90.0 kg, has a 0.340-m radius, and is turning at 90.0 rpm, and you press a steel axe against it with a radial force of 20.0 N. (a) Assuming the kinetic coefficient of friction between steel and stone is 0.20, calculate the angular acceleration of the grindstone. (b) How many turns will the stone make before coming to rest?

Exercise:**Problem:**

Suppose you exert a force of 180 N tangential to a 0.280-m-radius, 75.0-kg grindstone (a solid disk). (a) What torque is exerted? (b) What is the angular acceleration assuming negligible opposing friction? (c) What is the angular acceleration if there is an opposing frictional force of 20.0 N exerted 1.50 cm from the axis?

Solution:

- a. $\tau = (0.280 \text{ m})(180.0 \text{ N}) = 50.4 \text{ N} \cdot \text{m}$; b. $\alpha = 17.14 \text{ rad/s}^2$;
c. $\alpha = 17.04 \text{ rad/s}^2$

Exercise:**Problem:**

A flywheel ($I = 50 \text{ kg} \cdot \text{m}^2$) starting from rest acquires an angular velocity of 200.0 rad/s while subject to a constant torque from a motor for 5 s. (a) What is the angular acceleration of the flywheel? (b) What is the magnitude of the torque?

Exercise:

Problem:

A constant torque is applied to a rigid body whose moment of inertia is $4.0 \text{ kg}\cdot\text{m}^2$ around the axis of rotation. If the wheel starts from rest and attains an angular velocity of 20.0 rad/s in 10.0 s , what is the applied torque?

Solution:

$$\tau = 8.0 \text{ N}\cdot\text{m}$$

Exercise:**Problem:**

A torque of $50.0 \text{ N}\cdot\text{m}$ is applied to a grinding wheel ($I = 20.0 \text{ kg}\cdot\text{m}^2$) for 20 s . (a) If it starts from rest, what is the angular velocity of the grinding wheel after the torque is removed? (b) Through what angle does the wheel move while the torque is applied?

Exercise:**Problem:**

A flywheel ($I = 100.0 \text{ kg}\cdot\text{m}^2$) rotating at 500.0 rev/min is brought to rest by friction in 2.0 min . What is the frictional torque on the flywheel?

Solution:

$$\tau = -43.6 \text{ N}\cdot\text{m}$$

Exercise:

Problem:

A uniform cylindrical grinding wheel of mass 50.0 kg and diameter 1.0 m is turned on by an electric motor. The friction in the bearings is negligible. (a) What torque must be applied to the wheel to bring it from rest to 120 rev/min in 20 revolutions? (b) A tool whose coefficient of kinetic friction with the wheel is 0.60 is pressed perpendicularly against the wheel with a force of 40.0 N. What torque must be supplied by the motor to keep the wheel rotating at a constant angular velocity?

Exercise:**Problem:**

Suppose when Earth was created, it was not rotating. However, after the application of a uniform torque after 6 days, it was rotating at 1 rev/day. (a) What was the angular acceleration during the 6 days? (b) What torque was applied to Earth during this period? (c) What force tangent to Earth at its equator would produce this torque?

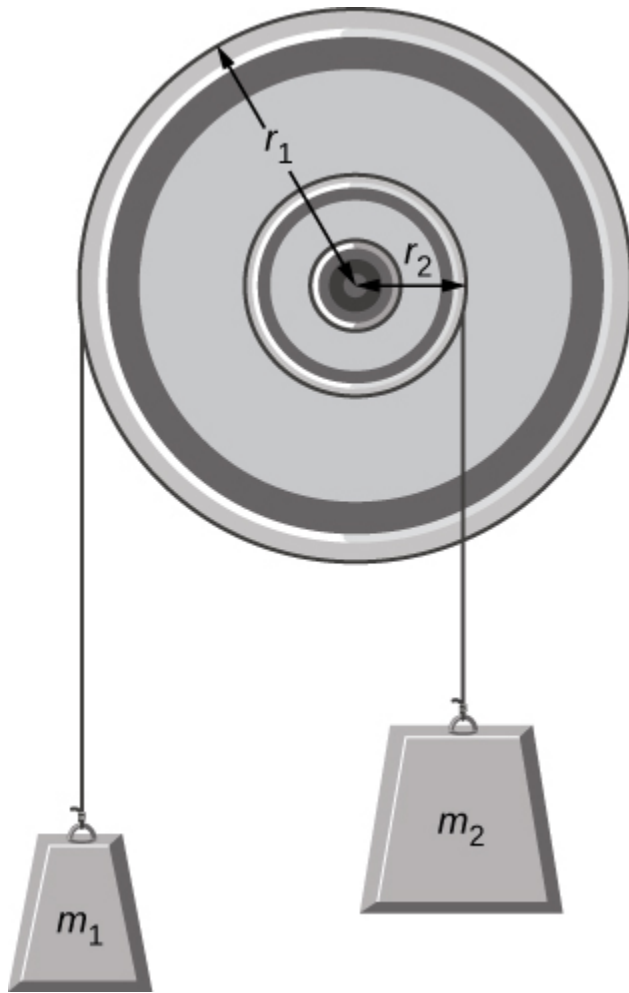
Solution:

- a. $\alpha = 1.4 \times 10^{-10} \text{ rad/s}^2$;
b. $\tau = 1.36 \times 10^{28} \text{ N-m}$; c. $F = 2.1 \times 10^{21} \text{ N}$

Exercise:**Problem:**

A pulley of moment of inertia 2.0 kg-m^2 is mounted on a wall as shown in the following figure. Light strings are wrapped around two circumferences of the pulley and weights are attached. What are (a) the angular acceleration of the pulley and (b) the linear acceleration of the weights? Assume the following data:

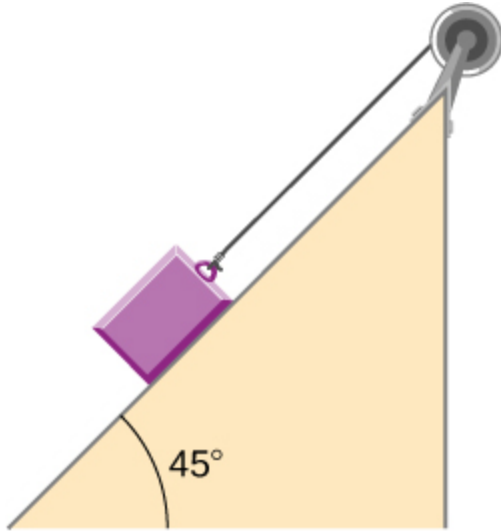
$$r_1 = 50 \text{ cm}, \quad r_2 = 20 \text{ cm}, \quad m_1 = 1.0 \text{ kg}, \quad m_2 = 2.0 \text{ kg}.$$



Exercise:

Problem:

A block of mass 3 kg slides down an inclined plane at an angle of 45° with a massless tether attached to a pulley with mass 1 kg and radius 0.5 m at the top of the incline (see the following figure). The pulley can be approximated as a disk. The coefficient of kinetic friction on the plane is 0.4. What is the acceleration of the block?



Solution:

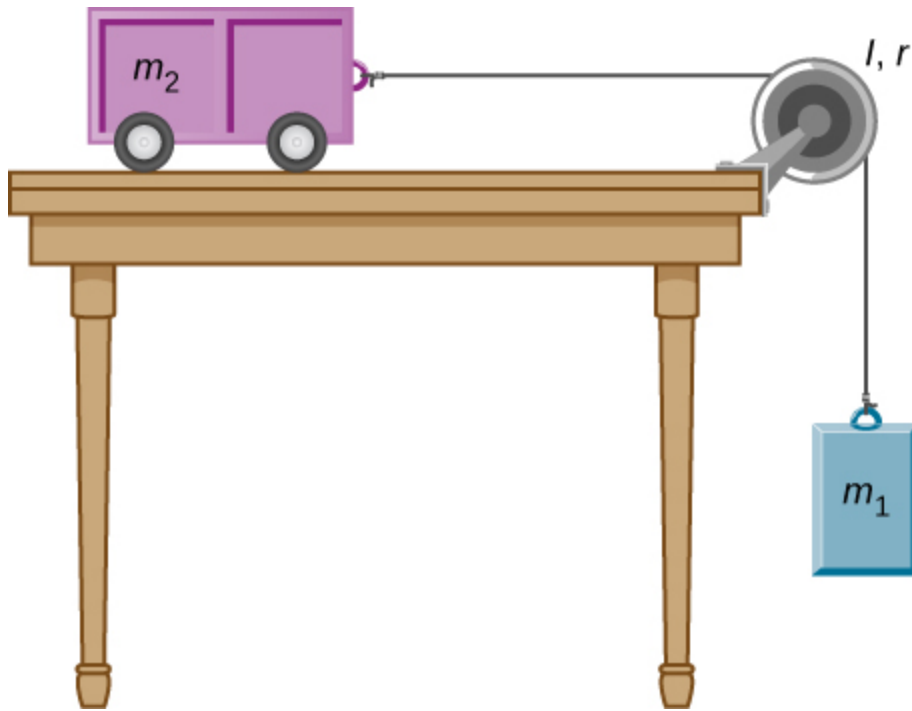
$$a = 3.6 \text{ m/s}^2$$

Exercise:

Problem:

The cart shown below moves across the table top as the block falls. What is the acceleration of the cart? Neglect friction and assume the following data:

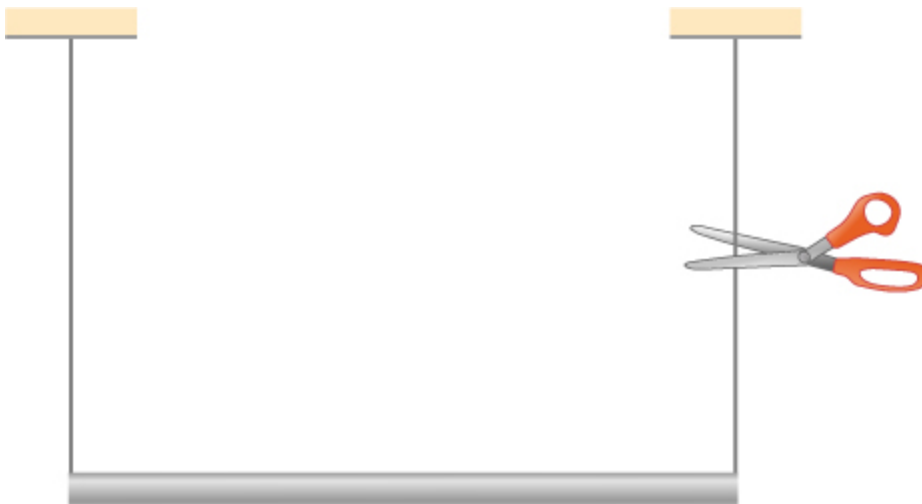
$$m_1 = 2.0 \text{ kg}, m_2 = 4.0 \text{ kg}, I = 0.4 \text{ kg}\cdot\text{m}^2, r = 20 \text{ cm}$$



Exercise:

Problem:

A uniform rod of mass and length is held vertically by two strings of negligible mass, as shown below. (a) Immediately after the string is cut, what is the linear acceleration of the free end of the stick? (b) Of the middle of the stick?

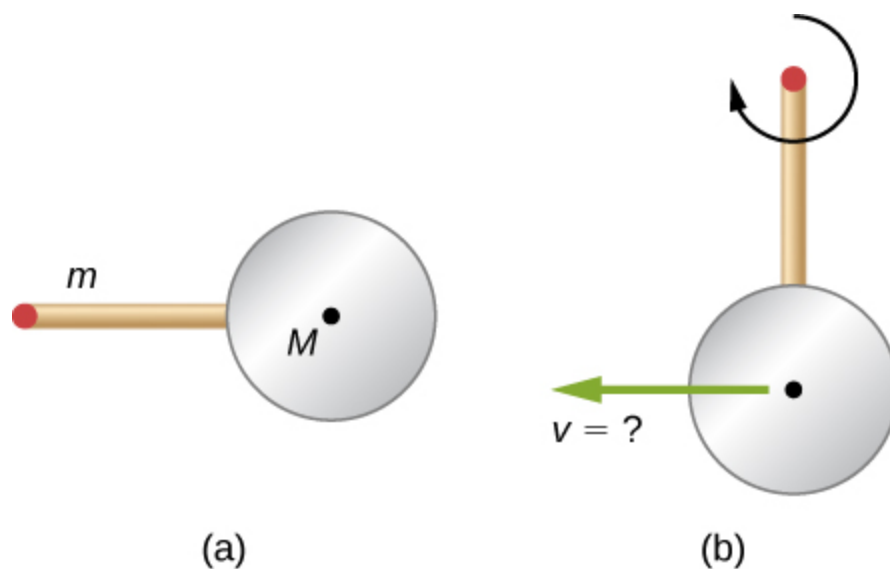


Solution:

a. $a = r\alpha = 14.7 \text{ m/s}^2$; b. $a = \frac{L}{2}\alpha = \frac{3}{4}g$

Exercise:**Problem:**

A thin stick of mass 0.2 kg and length $L = 0.5 \text{ m}$ is attached to the rim of a metal disk of mass $M = 2.0 \text{ kg}$ and radius $R = 0.3 \text{ m}$. The stick is free to rotate around a horizontal axis through its other end (see the following figure). (a) If the combination is released with the stick horizontal, what is the speed of the center of the disk when the stick is vertical? (b) What is the acceleration of the center of the disk at the instant the stick is released? (c) At the instant the stick passes through the vertical?

**Glossary**

Newton's second law for rotation

sum of the torques on a rotating system equals its moment of inertia times its angular acceleration

rotational dynamics

analysis of rotational motion using the net torque and moment of inertia to find the angular acceleration

Work and Power for Rotational Motion

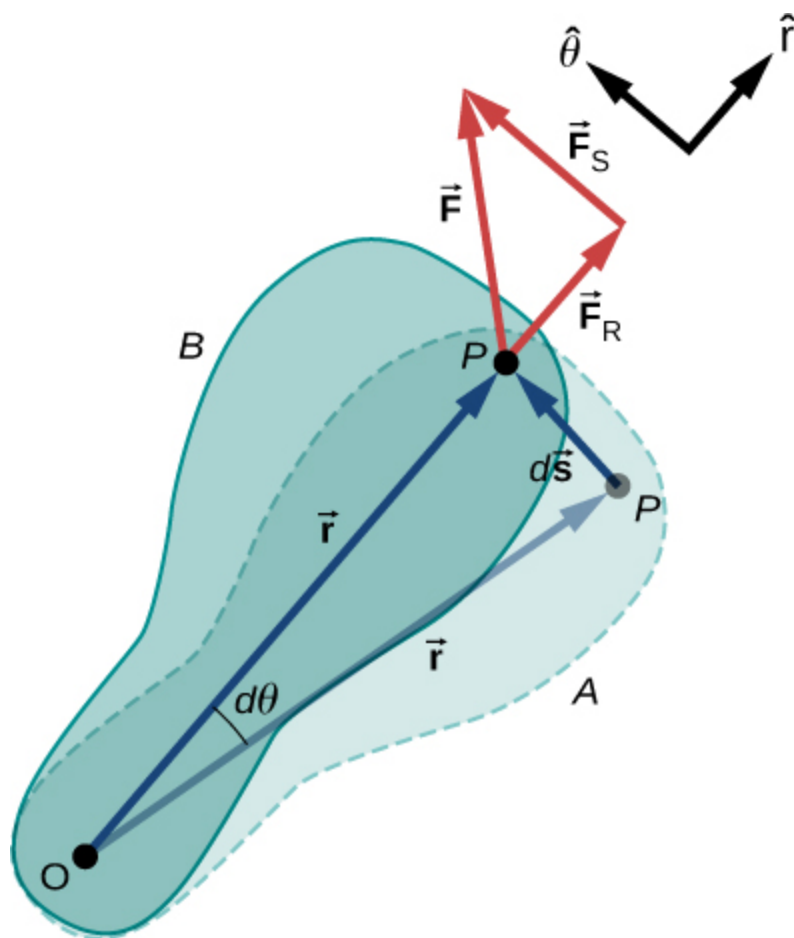
By the end of this section, you will be able to:

- Use the work-energy theorem to analyze rotation to find the work done on a system when it is rotated about a fixed axis for a finite angular displacement
- Solve for the angular velocity of a rotating rigid body using the work-energy theorem
- Find the power delivered to a rotating rigid body given the applied torque and angular velocity
- Summarize the rotational variables and equations and relate them to their translational counterparts

Thus far in the chapter, we have extensively addressed kinematics and dynamics for rotating rigid bodies around a fixed axis. In this final section, we define work and power within the context of rotation about a fixed axis, which has applications to both physics and engineering. The discussion of work and power makes our treatment of rotational motion almost complete, with the exception of rolling motion and angular momentum, which are discussed in [Angular Momentum](#). We begin this section with a treatment of the work-energy theorem for rotation.

Work for Rotational Motion

Now that we have determined how to calculate kinetic energy for rotating rigid bodies, we can proceed with a discussion of the work done on a rigid body rotating about a fixed axis. [\[link\]](#) shows a rigid body that has rotated through an angle $d\theta$ from A to B while under the influence of a force \vec{F} . The external force \vec{F} is applied to point P , whose position is \vec{r} , and the rigid body is constrained to rotate about a fixed axis that is perpendicular to the page and passes through O . The rotational axis is fixed, so the vector \vec{r} moves in a circle of radius r , and the vector $d\vec{s}$ is perpendicular to \vec{r} .



A rigid body rotates through an angle $d\theta$ from A to B by the action of an external force \vec{F} applied to point P .

From [\[link\]](#), we have
Equation:

$$\vec{s} = \vec{\theta} \times \vec{r}.$$

Thus,
Equation:

$$d\vec{s} = d(\vec{\theta} \times \vec{r}) = d\vec{\theta} \times \vec{r} + d\vec{r} \times \vec{\theta} = d\vec{\theta} \times \vec{r}.$$

Note that $d\vec{r}$ is zero because \vec{r} is fixed on the rigid body from the origin O to point P . Using the definition of work, we obtain

Equation:

$$W = \int \sum \vec{F} \cdot d\vec{s} = \int \sum \vec{F} \cdot (d\vec{\theta} \times \vec{r}) = \int d\vec{\theta} \cdot (\vec{r} \times \sum \vec{F})$$

where we used the identity $\vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{b} \cdot (\vec{c} \times \vec{a})$. Noting that $(\vec{r} \times \sum \vec{F}) = \sum \vec{\tau}$, we arrive at the expression for the **rotational work** done on a rigid body:

Equation:

$$W = \int \sum \vec{\tau} \cdot d\vec{\theta}.$$

The total work done on a rigid body is the sum of the torques integrated over the angle through which the body rotates. The incremental work is

Note:

Equation:

$$dW = \left(\sum_i \tau_i \right) d\theta$$

where we have taken the dot product in [\[link\]](#), leaving only torques along the axis of rotation. In a rigid body, all particles rotate through the same angle; thus the work of every external force is equal to the torque times the

common incremental angle $d\theta$. The quantity $\left(\sum_i \tau_i\right)$ is the net torque on the body due to external forces.

Similarly, we found the kinetic energy of a rigid body rotating around a fixed axis by summing the kinetic energy of each particle that makes up the rigid body. Since the work-energy theorem $W_i = \Delta K_i$ is valid for each particle, it is valid for the sum of the particles and the entire body.

Note:

Work-Energy Theorem for Rotation

The work-energy theorem for a rigid body rotating around a fixed axis is

Equation:

$$W_{AB} = K_B - K_A$$

where

Equation:

$$K = \frac{1}{2} I \omega^2$$

and the rotational work done by a net force rotating a body from point A to point B is

Equation:

$$W_{AB} = \int_{\theta_A}^{\theta_B} \left(\sum_i \tau_i \right) d\theta.$$

We give a strategy for using this equation when analyzing rotational motion.

Note:**Work-Energy Theorem for Rotational Motion**

1. Identify the forces on the body and draw a free-body diagram.
Calculate the torque for each force.
2. Calculate the work done during the body's rotation by every torque.
3. Apply the work-energy theorem by equating the net work done on the body to the change in rotational kinetic energy.

Let's look at two examples and use the work-energy theorem to analyze rotational motion.

Example:**Rotational Work and Energy**

A $12.0 \text{ N} \cdot \text{m}$ torque is applied to a flywheel that rotates about a fixed axis and has a moment of inertia of $30.0 \text{ kg} \cdot \text{m}^2$. If the flywheel is initially at rest, what is its angular velocity after it has turned through eight revolutions?

Strategy

We apply the work-energy theorem. We know from the problem description what the torque is and the angular displacement of the flywheel. Then we can solve for the final angular velocity.

Solution

The flywheel turns through eight revolutions, which is 16π radians. The work done by the torque, which is constant and therefore can come outside the integral in [\[link\]](#), is

Equation:

$$W_{AB} = \tau(\theta_B - \theta_A).$$

We apply the work-energy theorem:

Equation:

$$W_{AB} = \tau(\theta_B - \theta_A) = \frac{1}{2}I\omega_B^2 - \frac{1}{2}I\omega_A^2.$$

With

$\tau = 12.0 \text{ N} \cdot \text{m}$, $\theta_B - \theta_A = 16.0\pi \text{ rad}$, $I = 30.0 \text{ kg} \cdot \text{m}^2$, and $\omega_A = 0$, we have

Equation:

$$12.0 \text{ N} \cdot \text{m}(16.0\pi \text{ rad}) = \frac{1}{2}(30.0 \text{ kg} \cdot \text{m}^2)(\omega_B^2) - 0.$$

Therefore,

Equation:

$$\omega_B = 6.3 \text{ rad/s}.$$

This is the angular velocity of the flywheel after eight revolutions.

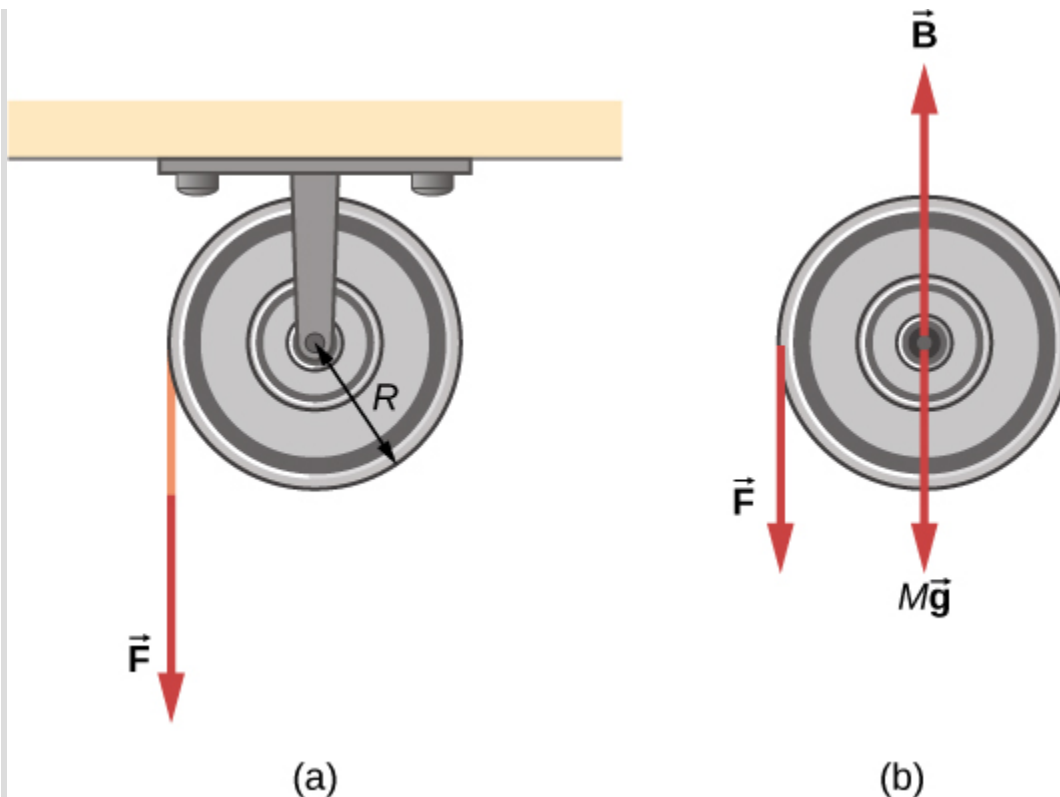
Significance

The work-energy theorem provides an efficient way to analyze rotational motion, connecting torque with rotational kinetic energy.

Example:

Rotational Work: A Pulley

A string wrapped around the pulley in [\[link\]](#) is pulled with a constant downward force \vec{F} of magnitude 50 N. The radius R and moment of inertia I of the pulley are 0.10 m and $2.5 \times 10^{-3} \text{ kg} \cdot \text{m}^2$, respectively. If the string does not slip, what is the angular velocity of the pulley after 1.0 m of string has unwound? Assume the pulley starts from rest.



(a) A string is wrapped around a pulley of radius R . (b) The free-body diagram.

Strategy

Looking at the free-body diagram, we see that neither \vec{B} , the force on the bearings of the pulley, nor $M\vec{g}$, the weight of the pulley, exerts a torque around the rotational axis, and therefore does no work on the pulley. As the pulley rotates through an angle θ , \vec{F} acts through a distance d such that $d = R\theta$.

Solution

Since the torque due to \vec{F} has magnitude $\tau = RF$, we have

Equation:

$$W = \tau\theta = (FR)\theta = Fd.$$

If the force on the string acts through a distance of 1.0 m, we have, from the work-energy theorem,

Equation:

$$\begin{aligned}W_{AB} &= K_B - K_A \\Fd &= \frac{1}{2}I\omega^2 - 0 \\(50.0 \text{ N})(1.0 \text{ m}) &= \frac{1}{2}(2.5 \times 10^{-3} \text{ kg}\cdot\text{m}^2)\omega^2.\end{aligned}$$

Solving for ω , we obtain

Equation:

$$\omega = 200.0 \text{ rad/s.}$$

Power for Rotational Motion

Power always comes up in the discussion of applications in engineering and physics. Power for rotational motion is equally as important as power in linear motion and can be derived in a similar way as in linear motion when the force is a constant. The linear power when the force is a constant is

$P = \vec{\mathbf{F}} \cdot \vec{\mathbf{v}}$. If the net torque is constant over the angular displacement, [\[link\]](#) simplifies and the net torque can be taken out of the integral. In the following discussion, we assume the net torque is constant. We can apply the definition of power derived in [Power](#) to rotational motion. From [Work and Kinetic Energy](#), the instantaneous power (or just power) is defined as the rate of doing work,

Equation:

$$P = \frac{dW}{dt}.$$

If we have a constant net torque, [\[link\]](#) becomes $W = \tau\theta$ and the power is

Equation:

$$P = \frac{dW}{dt} = \frac{d}{dt}(\tau\theta) = \tau \frac{d\theta}{dt}$$

or

Note:

Equation:

$$P = \tau\omega.$$

Example:

Torque on a Boat Propeller

A boat engine operating at $9.0 \times 10^4 \text{ W}$ is running at 300 rev/min. What is the torque on the propeller shaft?

Strategy

We are given the rotation rate in rev/min and the power consumption, so we can easily calculate the torque.

Solution

Equation:

$$300.0 \text{ rev/min} = 31.4 \text{ rad/s};$$

Equation:

$$\tau = \frac{P}{\omega} = \frac{9.0 \times 10^4 \text{ N} \cdot \text{m/s}}{31.4 \text{ rad/s}} = 2864.8 \text{ N} \cdot \text{m}.$$

Significance

It is important to note the radian is a dimensionless unit because its definition is the ratio of two lengths. It therefore does not appear in the solution.

Note:

Exercise:

Problem:

Check Your Understanding A constant torque of $500 \text{ kN} \cdot \text{m}$ is applied to a wind turbine to keep it rotating at 6 rad/s . What is the power required to keep the turbine rotating?

Solution:

3 MW

Rotational and Translational Relationships Summarized

The rotational quantities and their linear analog are summarized in three tables. [\[link\]](#) summarizes the rotational variables for circular motion about a fixed axis with their linear analogs and the connecting equation, except for the centripetal acceleration, which stands by itself. [\[link\]](#) summarizes the rotational and translational kinematic equations. [\[link\]](#) summarizes the rotational dynamics equations with their linear analogs.

Rotational	Translational	Relationship
θ	x	$\theta = \frac{s}{r}$
ω	v_t	$\omega = \frac{v_t}{r}$
α	a_t	$\alpha = \frac{a_t}{r}$
	a_c	$a_c = \frac{v_t^2}{r}$

Rotational and Translational Variables: Summary

Rotational	Translational
$\theta_f = \theta_0 + \bar{\omega}t$	$x = x_0 + \bar{v}t$
$\omega_f = \omega_0 + \alpha t$	$v_f = v_0 + at$
$\theta_f = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$	$x_f = x_0 + v_0 t + \frac{1}{2}at^2$
$\omega_f^2 = \omega_0^2 + 2\alpha(\Delta\theta)$	$v_f^2 = v_0^2 + 2a(\Delta x)$

Rotational and Translational Kinematic Equations: Summary

Rotational	Translational
$I = \sum_i m_i r_i^2$	m
$K = \frac{1}{2}I\omega^2$	$K = \frac{1}{2}mv^2$
$\sum_i \tau_i = I\alpha$	$\sum_i \vec{\mathbf{F}}_i = m\vec{\mathbf{a}}$

Rotational	Translational
$W_{AB} = \int_{\theta_A}^{\theta_B} \left(\sum_i \tau_i \right) d\theta$	$W = \int \vec{\mathbf{F}} \cdot d\vec{\mathbf{s}}$
$P = \tau\omega$	$P = \vec{\mathbf{F}} \cdot \vec{\mathbf{v}}$

Rotational and Translational Equations: Dynamics

Summary

- The incremental work dW in rotating a rigid body about a fixed axis is the sum of the torques about the axis times the incremental angle $d\theta$.
- The total work done to rotate a rigid body through an angle θ about a fixed axis is the sum of the torques integrated over the angular displacement. If the torque is a constant as a function of θ , then $W_{AB} = \tau(\theta_B - \theta_A)$.
- The work-energy theorem relates the rotational work done to the change in rotational kinetic energy: $W_{AB} = K_B - K_A$ where $K = \frac{1}{2}I\omega^2$.
- The power delivered to a system that is rotating about a fixed axis is the torque times the angular velocity, $P = \tau\omega$.

Key Equations

Angular position	$\theta = \frac{s}{r}$
Angular velocity	$\omega = \lim_{\Delta t \rightarrow 0} \frac{\Delta\theta}{\Delta t} = \frac{d\theta}{dt}$

Tangential speed	$v_t = r\omega$
Angular acceleration	$\alpha = \lim_{\Delta t \rightarrow 0} \frac{\Delta\omega}{\Delta t} = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2}$
Tangential acceleration	$a_t = r\alpha$
Average angular velocity	$\bar{\omega} = \frac{\omega_0 + \omega_f}{2}$
Angular displacement	$\theta_f = \theta_0 + \bar{\omega}t$
Angular velocity from constant angular acceleration	$\omega_f = \omega_0 + \alpha t$
Angular velocity from displacement and constant angular acceleration	$\theta_f = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$
Change in angular velocity	$\omega_f^2 = \omega_0^2 + 2\alpha(\Delta\theta)$
Total acceleration	$\vec{a} = \vec{a}_c + \vec{a}_t$
Rotational kinetic energy	$K = \frac{1}{2} \left(\sum_j m_j r_j^2 \right) \omega^2$
Moment of inertia	$I = \sum_j m_j r_j^2$
Rotational kinetic energy in terms of the moment of inertia of a rigid body	$K = \frac{1}{2} I \omega^2$
Moment of inertia of a continuous object	$I = \int r^2 dm$

Parallel-axis theorem	$I_{\text{parallel-axis}} = I_{\text{center of mass}} + md^2$
Moment of inertia of a compound object	$I_{\text{total}} = \sum_i I_i$
Torque vector	$\vec{\tau} = \vec{r} \times \vec{F}$
Magnitude of torque	$ \vec{\tau} = r_{\perp} F$
Total torque	$\vec{\tau}_{\text{net}} = \sum_i \vec{\tau}_i $
Newton's second law for rotation	$\sum_i \tau_i = I\alpha$
Incremental work done by a torque	$dW = \left(\sum_i \tau_i \right) d\theta$
Work-energy theorem	$W_{AB} = K_B - K_A$
Rotational work done by net force	$W_{AB} = \int_{\theta_A}^{\theta_B} \left(\sum_i \tau_i \right) d\theta$
Rotational power	$P = \tau\omega$

Problems

Exercise:

Problem:

A wind turbine rotates at 20 rev/min. If its power output is 2.0 MW, what is the torque produced on the turbine from the wind?

Solution:

$$\tau = \frac{P}{\omega} = \frac{2.0 \times 10^6 \text{ W}}{2.1 \text{ rad/s}} = 9.5 \times 10^5 \text{ N} \cdot \text{m}$$

Exercise:**Problem:**

A clay cylinder of radius 20 cm on a potter's wheel spins at a constant rate of 10 rev/s. The potter applies a force of 10 N to the clay with his hands where the coefficient of friction is 0.1 between his hands and the clay. What is the power that the potter has to deliver to the wheel to keep it rotating at this constant rate?

Exercise:**Problem:**

A uniform cylindrical grindstone has a mass of 10 kg and a radius of 12 cm. (a) What is the rotational kinetic energy of the grindstone when it is rotating at $1.5 \times 10^3 \text{ rev/min}$? (b) After the grindstone's motor is turned off, a knife blade is pressed against the outer edge of the grindstone with a perpendicular force of 5.0 N. The coefficient of kinetic friction between the grindstone and the blade is 0.80. Use the work energy theorem to determine how many turns the grindstone makes before it stops.

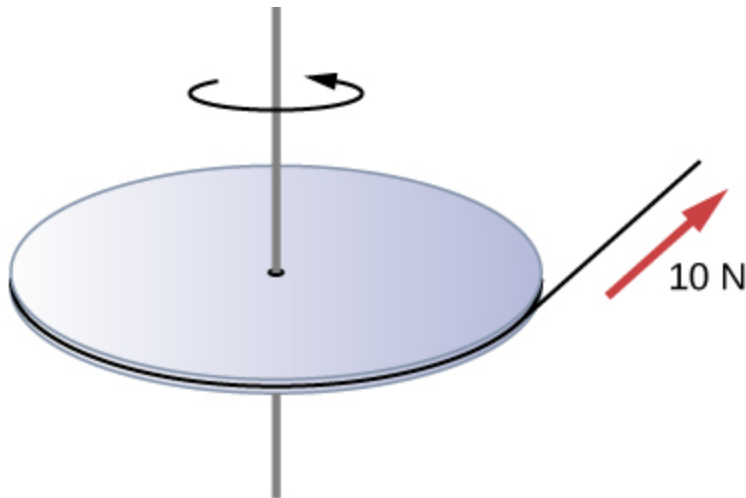
Solution:

- a. $K = 888.50 \text{ J}$;
- b. $\Delta\theta = 294.6 \text{ rev}$

Exercise:

Problem:

A uniform disk of mass 500 kg and radius 0.25 m is mounted on frictionless bearings so it can rotate freely around a vertical axis through its center (see the following figure). A cord is wrapped around the rim of the disk and pulled with a force of 10 N. (a) How much work has the force done at the instant the disk has completed three revolutions, starting from rest? (b) Determine the torque due to the force, then calculate the work done by this torque at the instant the disk has completed three revolutions? (c) What is the angular velocity at that instant? (d) What is the power output of the force at that instant?

**Exercise:****Problem:**

A propeller is accelerated from rest to an angular velocity of 1000 rev/min over a period of 6.0 seconds by a constant torque of $2.0 \times 10^3 \text{ N} \cdot \text{m}$. (a) What is the moment of inertia of the propeller? (b) What power is being provided to the propeller 3.0 s after it starts rotating?

Solution:

- a. $I = 114.6 \text{ kg} \cdot \text{m}^2$;
- b. $P = 104,700 \text{ W}$

Exercise:

Problem:

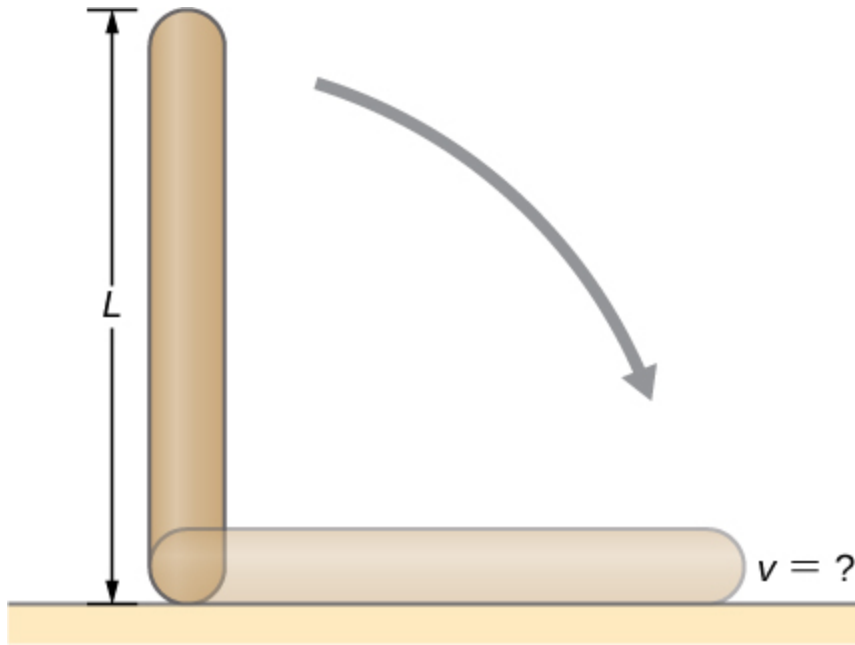
A sphere of mass 1.0 kg and radius 0.5 m is attached to the end of a massless rod of length 3.0 m . The rod rotates about an axis that is at the opposite end of the sphere (see below). The system rotates horizontally about the axis at a constant 400 rev/min . After rotating at this angular speed in a vacuum, air resistance is introduced and provides a force 0.15 N on the sphere opposite to the direction of motion. What is the power provided by air resistance to the system 100.0 s after air resistance is introduced?



Exercise:

Problem:

A uniform rod of length L and mass M is held vertically with one end resting on the floor as shown below. When the rod is released, it rotates around its lower end until it hits the floor. Assuming the lower end of the rod does not slip, what is the linear velocity of the upper end when it hits the floor?



Solution:

$$v = L\omega = \sqrt{3Lg}$$

Exercise:

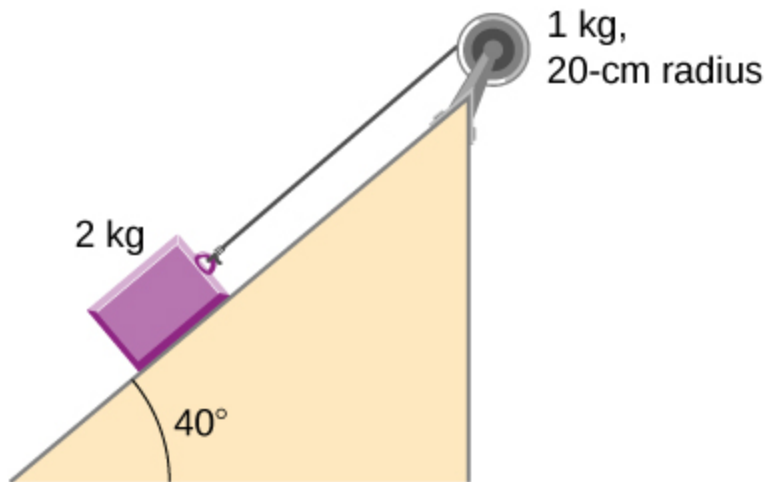
Problem:

An athlete in a gym applies a constant force of 50 N to the pedals of a bicycle at a rate of the pedals moving 60 rev/min. The length of the pedal arms is 30 cm. What is the power delivered to the bicycle by the athlete?

Exercise:

Problem:

A 2-kg block on a frictionless inclined plane at 40° has a cord attached to a pulley of mass 1 kg and radius 20 cm (see the following figure).
 (a) What is the acceleration of the block down the plane? (b) What is the work done by the cord on the pulley?



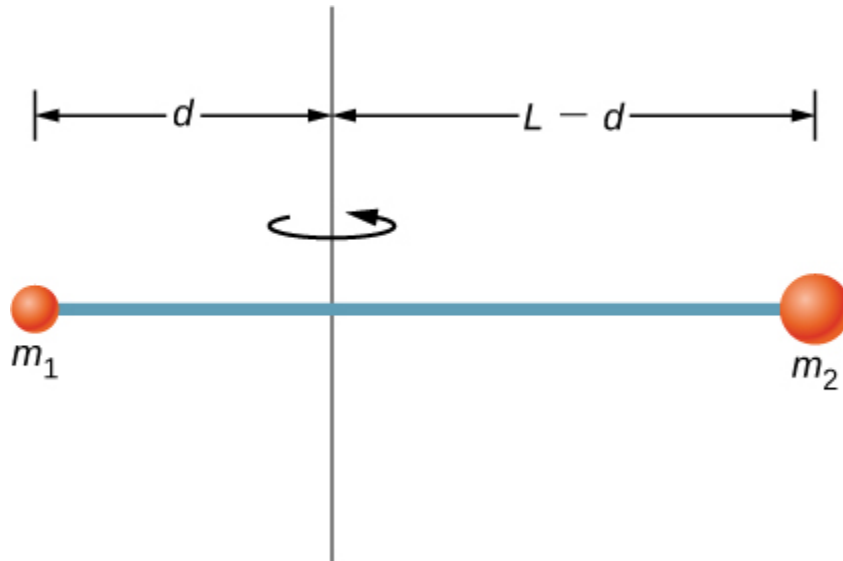
Solution:

a. $a = 5.0\text{ m/s}^2$; b. $W = 1.25\text{ N} \cdot \text{m}$

Exercise:

Problem:

Small bodies of mass m_1 and m_2 are attached to opposite ends of a thin rigid rod of length L and mass M . The rod is mounted so that it is free to rotate in a horizontal plane around a vertical axis (see below). What distance d from m_1 should the rotational axis be so that a minimum amount of work is required to set the rod rotating at an angular velocity ω ?



Additional Problems

Exercise:

Problem:

A cyclist is riding such that the wheels of the bicycle have a rotation rate of 3.0 rev/s . If the cyclist brakes such that the rotation rate of the wheels decrease at a rate of 0.3 rev/s^2 , how long does it take for the cyclist to come to a complete stop?

Solution:

$$\Delta t = 10.0 \text{ s}$$

Exercise:

Problem:

Calculate the angular velocity of the orbital motion of Earth around the Sun.

Exercise:

Problem:

A phonograph turntable rotating at $33 \frac{1}{3}$ rev/min slows down and stops in 1.0 min. (a) What is the turntable's angular acceleration assuming it is constant? (b) How many revolutions does the turntable make while stopping?

Solution:

a. 0.06 rad/s^2 ; b. $\theta = 105.0 \text{ rad}$

Exercise:**Problem:**

With the aid of a string, a gyroscope is accelerated from rest to 32 rad/s in 0.40 s under a constant angular acceleration. (a) What is its angular acceleration in rad/s^2 ? (b) How many revolutions does it go through in the process?

Exercise:**Problem:**

Suppose a piece of dust has fallen on a CD. If the spin rate of the CD is 500 rpm , and the piece of dust is 4.3 cm from the center, what is the total distance traveled by the dust in 3 minutes? (Ignore accelerations due to getting the CD rotating.)

Solution:

$s = 405.26 \text{ m}$

Exercise:

Problem:

A system of point particles is rotating about a fixed axis at 4 rev/s. The particles are fixed with respect to each other. The masses and distances to the axis of the point particles are $m_1 = 0.1 \text{ kg}$, $r_1 = 0.2 \text{ m}$, $m_2 = 0.05 \text{ kg}$, $r_2 = 0.4 \text{ m}$, $m_3 = 0.5 \text{ kg}$, $r_3 = 0.01 \text{ m}$. (a) What is the moment of inertia of the system? (b) What is the rotational kinetic energy of the system?

Exercise:**Problem:**

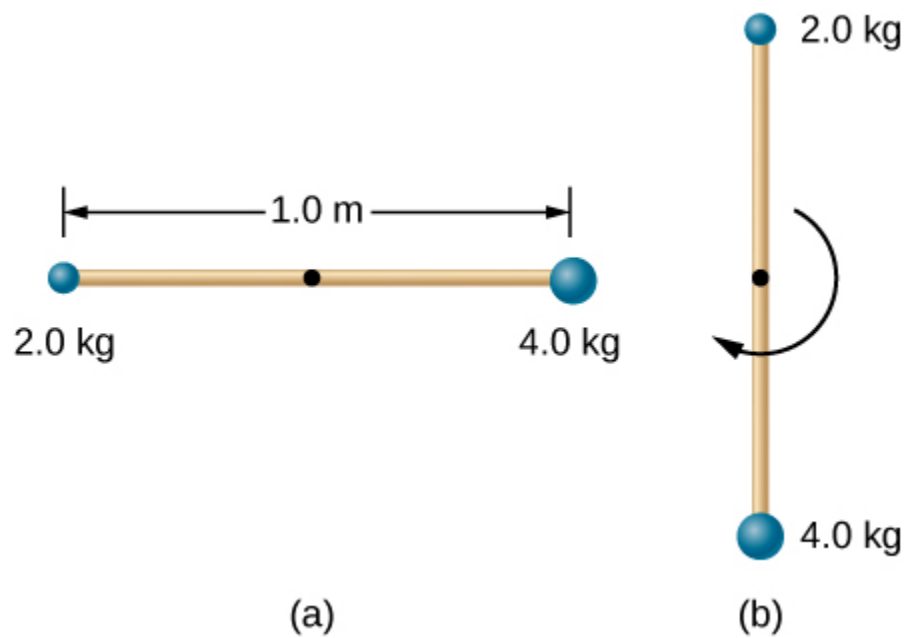
Calculate the moment of inertia of a skater given the following information. (a) The 60.0-kg skater is approximated as a cylinder that has a 0.110-m radius. (b) The skater with arms extended is approximated by a cylinder that is 52.5 kg, has a 0.110-m radius, and has two 0.900-m-long arms which are 3.75 kg each and extend straight out from the cylinder like rods rotated about their ends.

Solution:

- a. $I = 0.363 \text{ kg} \cdot \text{m}^2$;
- b. $I = 2.34 \text{ kg} \cdot \text{m}^2$

Exercise:**Problem:**

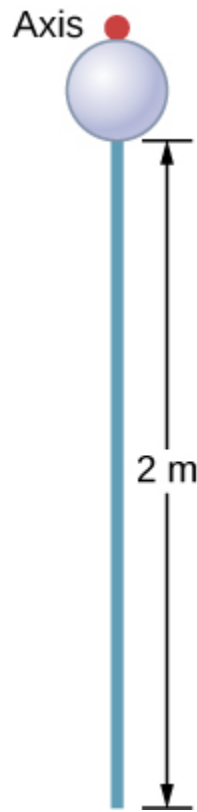
A stick of length 1.0 m and mass 6.0 kg is free to rotate about a horizontal axis through the center. Small bodies of masses 4.0 and 2.0 kg are attached to its two ends (see the following figure). The stick is released from the horizontal position. What is the angular velocity of the stick when it swings through the vertical?



Exercise:

Problem:

A pendulum consists of a rod of length 2 m and mass 3 kg with a solid sphere of mass 1 kg and radius 0.3 m attached at one end. The axis of rotation is as shown below. What is the angular velocity of the pendulum at its lowest point if it is released from rest at an angle of 30° ?



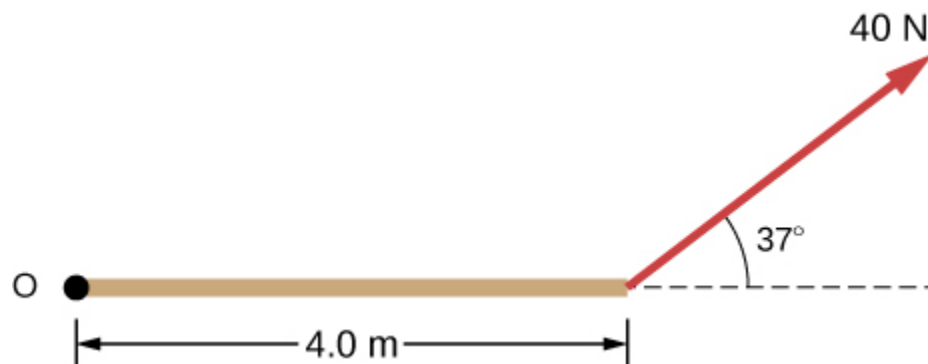
Solution:

$$\omega = \sqrt{\frac{6.68 \text{ J}}{4.4 \text{ kgm}^2}} = 1.23 \text{ rad/s}$$

Exercise:

Problem:

Calculate the torque of the 40-N force around the axis through O and perpendicular to the plane of the page as shown below.



Exercise:**Problem:**

Two children push on opposite sides of a door during play. Both push horizontally and perpendicular to the door. One child pushes with a force of 17.5 N at a distance of 0.600 m from the hinges, and the second child pushes at a distance of 0.450 m. What force must the second child exert to keep the door from moving? Assume friction is negligible.

Solution:

$$F = 23.3 \text{ N}$$

Exercise:**Problem:**

The force of $20\hat{j}\text{N}$ is applied at $\vec{r} = (4.0\hat{i} - 2.0\hat{j}) \text{ m}$. What is the torque of this force about the origin?

Exercise:**Problem:**

An automobile engine can produce $200 \text{ N}\cdot\text{m}$ of torque. Calculate the angular acceleration produced if 95.0% of this torque is applied to the drive shaft, axle, and rear wheels of a car, given the following information. The car is suspended so that the wheels can turn freely. Each wheel acts like a 15.0-kg disk that has a 0.180-m radius. The walls of each tire act like a 2.00-kg annular ring that has inside radius of 0.180 m and outside radius of 0.320 m. The tread of each tire acts like a 10.0-kg hoop of radius 0.330 m. The 14.0-kg axle acts like a rod that has a 2.00-cm radius. The 30.0-kg drive shaft acts like a rod that has a 3.20-cm radius.

Solution:

$$\alpha = \frac{190.0 \text{ N}\cdot\text{m}}{2.94 \text{ kg}\cdot\text{m}^2} = 64.4 \text{ rad/s}^2$$

Exercise:

Problem:

A grindstone with a mass of 50 kg and radius 0.8 m maintains a constant rotation rate of 4.0 rev/s by a motor while a knife is pressed against the edge with a force of 5.0 N. The coefficient of kinetic friction between the grindstone and the blade is 0.8. What is the power provided by the motor to keep the grindstone at the constant rotation rate?

Challenge Problems

Exercise:

Problem:

The angular acceleration of a rotating rigid body is given by $\alpha = (2.0 - 3.0t) \text{ rad/s}^2$. If the body starts rotating from rest at $t = 0$, (a) what is the angular velocity? (b) Angular position? (c) What angle does it rotate through in 10 s? (d) Where does the vector perpendicular to the axis of rotation indicating 0° at $t = 0$ lie at $t = 10$ s?

Solution:

a. $\omega = 2.0t - 1.5t^2$; b. $\theta = t^2 - 0.5t^3$; c. $\theta = -400.0 \text{ rad}$; d. the vector is at $-0.66(360^\circ) = -237.6^\circ$

Exercise:

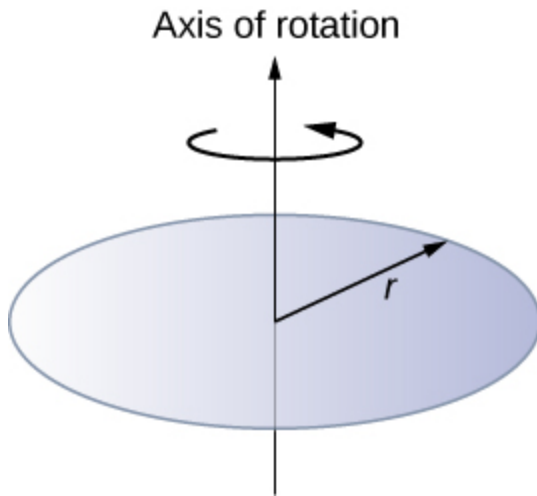
Problem:

Earth's day has increased by 0.002 s in the last century. If this increase in Earth's period is constant, how long will it take for Earth to come to rest?

Exercise:

Problem:

A disk of mass m , radius R , and area A has a surface mass density $\sigma = \frac{mr}{AR}$ (see the following figure). What is the moment of inertia of the disk about an axis through the center?



Solution:

$$I = \frac{2}{5}mR^2$$

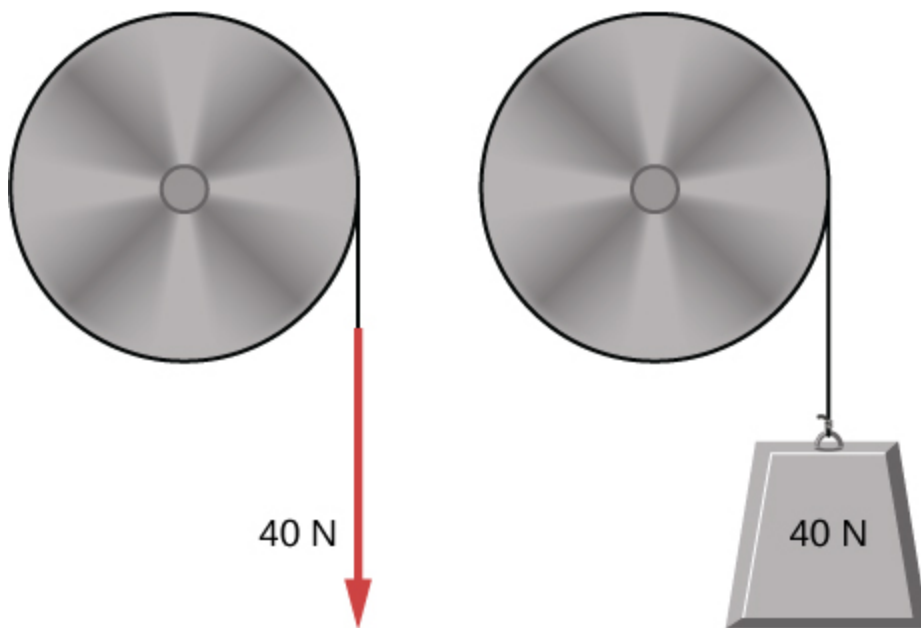
Exercise:**Problem:**

Zorch, an archenemy of Rotation Man, decides to slow Earth's rotation to once per 28.0 h by exerting an opposing force at and parallel to the equator. Rotation Man is not immediately concerned, because he knows Zorch can only exert a force of $4.00 \times 10^7 \text{ N}$ (a little greater than a Saturn V rocket's thrust). How long must Zorch push with this force to accomplish his goal? (This period gives Rotation Man time to devote to other villains.)

Exercise:

Problem:

A cord is wrapped around the rim of a solid cylinder of radius 0.25 m, and a constant force of 40 N is exerted on the cord shown, as shown in the following figure. The cylinder is mounted on frictionless bearings, and its moment of inertia is $6.0 \text{ kg} \cdot \text{m}^2$. (a) Use the work energy theorem to calculate the angular velocity of the cylinder after 5.0 m of cord have been removed. (b) If the 40-N force is replaced by a 40-N weight, what is the angular velocity of the cylinder after 5.0 m of cord have unwound?



Solution:

a. $\omega = 8.2 \text{ rad/s}$; b. $\omega = 8.0 \text{ rad/s}$

Glossary

rotational work

work done on a rigid body due to the sum of the torques integrated over the angle through which the body rotates

Introduction

class="introduction"

A helicopter
has its main
lift blades
rotating to
keep the
aircraft
airborne.

Due to
conservation
of angular
momentum,
the body of
the
helicopter
would want
to rotate in
the opposite
sense to the
blades, if it
were not for
the small
rotor on the
tail of the
aircraft,
which
provides
thrust to
stabilize it.



Angular momentum is the rotational counterpart of linear momentum. Any massive object that rotates about an axis carries angular momentum, including rotating flywheels, planets, stars, hurricanes, tornadoes, whirlpools, and so on. The helicopter shown in the chapter-opening picture can be used to illustrate the concept of angular momentum. The lift blades spin about a vertical axis through the main body and carry angular momentum. The body of the helicopter tends to rotate in the opposite sense in order to conserve angular momentum. The small rotors at the tail of the aircraft provide a counter thrust against the body to prevent this from happening, and the helicopter stabilizes itself. The concept of conservation of angular momentum is discussed later in this chapter. In the main part of this chapter, we explore the intricacies of angular momentum of rigid bodies such as a top, and also of point particles and systems of particles. But to be complete, we start with a discussion of rolling motion, which builds upon the concepts of the previous chapter.

Rolling Motion

By the end of this section, you will be able to:

- Describe the physics of rolling motion without slipping
- Explain how linear variables are related to angular variables for the case of rolling motion without slipping
- Find the linear and angular accelerations in rolling motion with and without slipping
- Calculate the static friction force associated with rolling motion without slipping
- Use energy conservation to analyze rolling motion

Rolling motion is that common combination of rotational and translational motion that we see everywhere, every day. Think about the different situations of wheels moving on a car along a highway, or wheels on a plane landing on a runway, or wheels on a robotic explorer on another planet. Understanding the forces and torques involved in **rolling motion** is a crucial factor in many different types of situations.

For analyzing rolling motion in this chapter, refer to [\[link\]](#) in [Fixed-Axis Rotation](#) to find moments of inertia of some common geometrical objects. You may also find it useful in other calculations involving rotation.

Rolling Motion without Slipping

People have observed rolling motion without slipping ever since the invention of the wheel. For example, we can look at the interaction of a car's tires and the surface of the road. If the driver depresses the accelerator to the floor, such that the tires spin without the car moving forward, there must be kinetic friction between the wheels and the surface of the road. If the driver depresses the accelerator slowly, causing the car to move forward, then the tires roll without slipping. It is surprising to most people that, in fact, the bottom of the wheel is at rest with respect to the ground, indicating there must be static friction between the tires and the road surface. In [\[link\]](#), the bicycle is in motion with the rider staying upright. The tires have contact with the road surface, and, even though they are rolling, the bottoms of the tires deform slightly, do not slip, and are at rest with respect to the road surface for a measurable amount of time. There must be static friction between the tire and the road surface for this to be so.



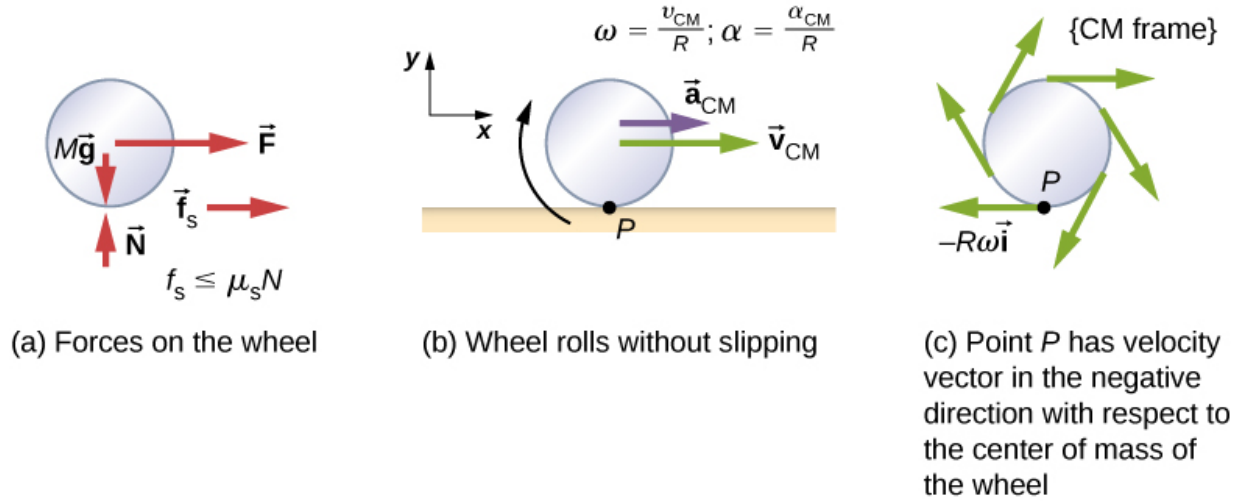
(a)



(b)

(a) The bicycle moves forward, and its tires do not slip. The bottom of the slightly deformed tire is at rest with respect to the road surface for a measurable amount of time. (b) This image shows that the top of a rolling wheel appears blurred by its motion, but the bottom of the wheel is instantaneously at rest. (credit a: modification of work by Nelson Lourenço; credit b: modification of work by Colin Rose)

To analyze rolling without slipping, we first derive the linear variables of velocity and acceleration of the center of mass of the wheel in terms of the angular variables that describe the wheel's motion. The situation is shown in [\[link\]](#).



- (a) A wheel is pulled across a horizontal surface by a force \vec{F} . The force of static friction \vec{f}_s , $|\vec{f}_s| \leq \mu_s N$ is large enough to keep it from slipping. (b) The linear velocity and acceleration vectors of the center of mass and the relevant expressions for ω and α . Point P is at rest relative to the surface. (c) Relative to the center of mass (CM) frame, point P has linear velocity $-R\omega\hat{i}$.

From [\[link\]](#)(a), we see the force vectors involved in preventing the wheel from slipping. In (b), point P that touches the surface is at rest relative to the surface. Relative to the center of mass, point P has velocity $-R\omega\hat{i}$, where R is the radius of the wheel and ω is the wheel's angular velocity about its axis. Since the wheel is rolling, the velocity of P with respect to the surface is its velocity with respect to the center of mass plus the velocity of the center of mass with respect to the surface:

Equation:

$$\vec{v}_P = -R\omega\hat{i} + v_{CM}\hat{i}.$$

Since the velocity of P relative to the surface is zero, $v_P = 0$, this says that

Note:

Equation:

$$v_{\text{CM}} = R\omega.$$

Thus, the velocity of the wheel's center of mass is its radius times the angular velocity about its axis. We show the correspondence of the linear variable on the left side of the equation with the angular variable on the right side of the equation. This is done below for the linear acceleration.

If we differentiate [\[link\]](#) on the left side of the equation, we obtain an expression for the linear acceleration of the center of mass. On the right side of the equation, R is a constant and since $\alpha = \frac{d\omega}{dt}$, we have

Note:

Equation:

$$a_{\text{CM}} = R\alpha.$$

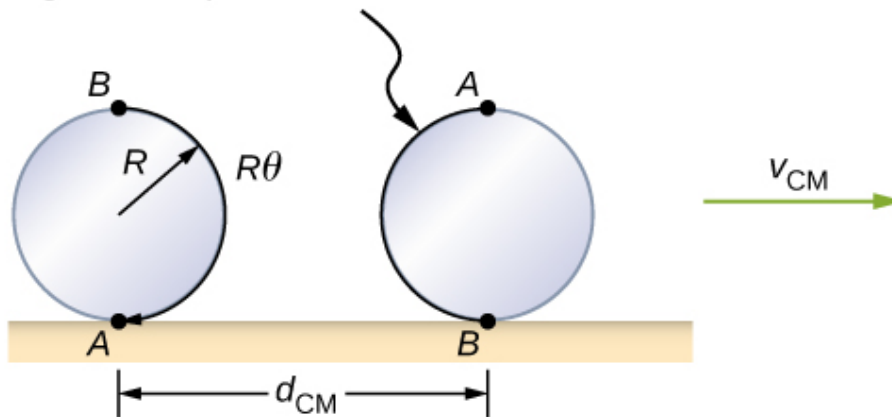
Furthermore, we can find the distance the wheel travels in terms of angular variables by referring to [\[link\]](#). As the wheel rolls from point A to point B , its outer surface maps onto the ground by exactly the distance travelled, which is d_{CM} . We see from [\[link\]](#) that the length of the outer surface that maps onto the ground is the arc length $R\theta$. Equating the two distances, we obtain

Note:

Equation:

$$d_{\text{CM}} = R\theta.$$

Arc length AB maps onto wheel's surface



As the wheel rolls on the surface, the arc length $R\theta$ from A to B maps onto the surface, corresponding to the distance d_{CM} that the center of mass has moved.

Example:

Rolling Down an Inclined Plane

A solid cylinder rolls down an inclined plane without slipping, starting from rest. It has mass m and radius r . (a) What is its acceleration? (b) What condition must the coefficient of static friction μ_s satisfy so the cylinder does not slip?

Strategy

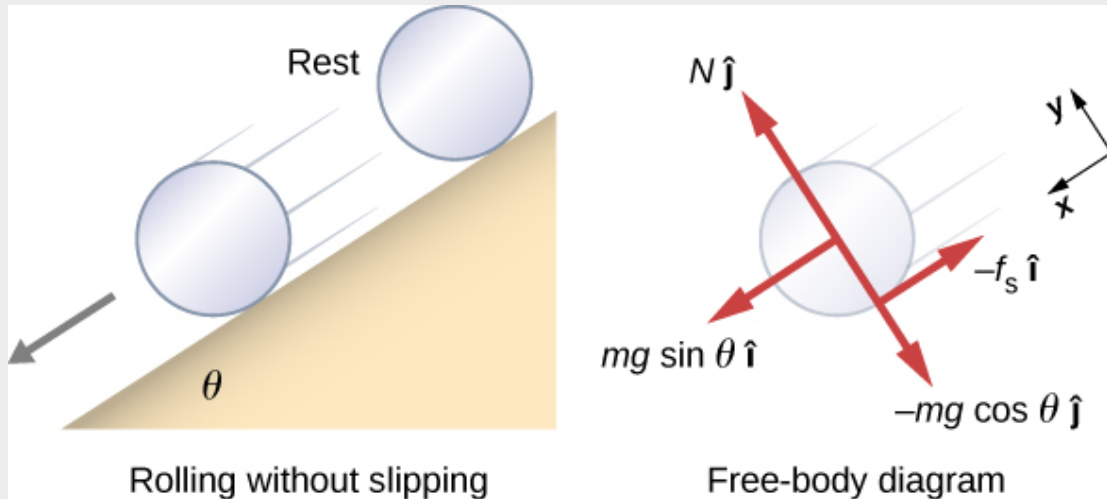
Draw a sketch and free-body diagram, and choose a coordinate system. We put x in the direction down the plane and y upward perpendicular to the plane. Identify the forces involved. These are the normal force, the force of gravity, and the force due to friction. Write down Newton's laws in the x - and y -directions, and Newton's law for rotation, and then solve for the acceleration and force due to friction.

Solution

- The free-body diagram and sketch are shown in [\[link\]](#), including the normal force, components of the weight, and the static friction force. There is barely enough friction to keep the cylinder rolling without slipping. Since there is no slipping, the magnitude of the friction force is less than or equal to $\mu_s N$. Writing down Newton's laws in the x - and y -directions, we have

Equation:

$$\sum F_x = ma_x; \quad \sum F_y = ma_y.$$



A solid cylinder rolls down an inclined plane without slipping from rest. The coordinate system has x in the direction down the inclined plane and y perpendicular to the plane. The free-body diagram is shown with the normal force, the static friction force, and the components of the weight $m\vec{g}$. Friction makes the cylinder roll down the plane rather than slip.

Substituting in from the free-body diagram,

Equation:

$$\begin{aligned} mg \sin \theta - f_s &= m(a_{\text{CM}})_x, \\ N - mg \cos \theta &= 0 \end{aligned}$$

we can then solve for the linear acceleration of the center of mass from these equations:

Equation:

$$a_{\text{CM}} = g \sin \theta - \frac{f_s}{m}$$

However, it is useful to express the linear acceleration in terms of the moment of inertia. For this, we write down Newton's second law for rotation,

Equation:

$$\sum \tau_{\text{CM}} = I_{\text{CM}}\alpha.$$

The torques are calculated about the axis through the center of mass of the cylinder. The only nonzero torque is provided by the friction force. We have

Equation:

$$f_s r = I_{\text{CM}}\alpha.$$

Finally, the linear acceleration is related to the angular acceleration by **Equation:**

$$(a_{\text{CM}})_x = r\alpha.$$

These equations can be used to solve for a_{CM} , α , and f_s in terms of the moment of inertia, where we have dropped the x -subscript. We rewrite a_{CM} in terms of the vertical component of gravity and the friction force, and make the following substitutions.

Equation:

$$f_s = \frac{I_{\text{CM}}\alpha}{r} = \frac{I_{\text{CM}}a_{\text{CM}}}{r^2}$$

From this we obtain

Equation:

$$\begin{aligned} a_{\text{CM}} &= g \sin \theta - \frac{I_{\text{CM}}a_{\text{CM}}}{mr^2}, \\ &= \frac{mg \sin \theta}{m + (I_{\text{CM}}/r^2)}. \end{aligned}$$

Note that this result is independent of the coefficient of static friction, μ_s . Since we have a solid cylinder, from [\[link\]](#), we have $I_{\text{CM}} = mr^2/2$ and **Equation:**

$$a_{\text{CM}} = \frac{mg \sin \theta}{m + (mr^2/2r^2)} = \frac{2}{3}g \sin \theta.$$

Therefore, we have **Equation:**

$$\alpha = \frac{a_{\text{CM}}}{r} = \frac{2}{3r}g \sin \theta.$$

b. Because slipping does not occur, $f_s \leq \mu_s N$. Solving for the friction force, **Equation:**

$$f_s = I_{\text{CM}} \frac{\alpha}{r} = I_{\text{CM}} \frac{(a_{\text{CM}})}{r^2} = \frac{I_{\text{CM}}}{r^2} \left(\frac{mg \sin \theta}{m + (I_{\text{CM}}/r^2)} \right) = \frac{mg I_{\text{CM}} \sin \theta}{mr^2 + I_{\text{CM}}}.$$

Substituting this expression into the condition for no slipping, and noting that $N = mg \cos \theta$, we have

Equation:

$$\frac{mg I_{\text{CM}} \sin \theta}{mr^2 + I_{\text{CM}}} \leq \mu_s mg \cos \theta$$

or

Equation:

$$\mu_s \geq \frac{\tan \theta}{1 + (mr^2/I_{\text{CM}})}.$$

For the solid cylinder, this becomes

Equation:

$$\mu_s \geq \frac{\tan \theta}{1 + (2mr^2/mr^2)} = \frac{1}{3}\tan \theta.$$

Significance

- The linear acceleration is linearly proportional to $\sin \theta$. Thus, the greater the angle of the incline, the greater the linear acceleration, as would be expected. The angular acceleration, however, is linearly proportional to $\sin \theta$ and inversely proportional to the radius of the cylinder. Thus, the larger the radius, the smaller the angular acceleration.
- For no slipping to occur, the coefficient of static friction must be greater than or equal to $(1/3)\tan \theta$. Thus, the greater the angle of incline, the greater the coefficient of static friction must be to prevent the cylinder from slipping.

Note:

Exercise:

Problem:

Check Your Understanding A hollow cylinder is on an incline at an angle of 60° . The coefficient of static friction on the surface is $\mu_s = 0.6$. (a) Does the cylinder roll without slipping? (b) Will a solid cylinder roll without slipping?

Solution:

a. $\mu_s \geq \frac{\tan \theta}{1 + (mr^2/I_{CM})}$; inserting the angle and noting that for a hollow cylinder $I_{CM} = mr^2$, we have $\mu_s \geq \frac{\tan 60^\circ}{1 + (mr^2/mr^2)} = \frac{1}{2}\tan 60^\circ = 0.87$; we are given a value of 0.6 for the coefficient of static friction, which is less than 0.87, so the condition isn't satisfied and the hollow cylinder will slip; b. The solid cylinder obeys the condition $\mu_s \geq \frac{1}{3}\tan \theta = \frac{1}{3}\tan 60^\circ = 0.58$. The value of 0.6 for μ_s satisfies this condition, so the solid cylinder will not slip.

It is worthwhile to repeat the equation derived in this example for the acceleration of an object rolling without slipping:

Note:

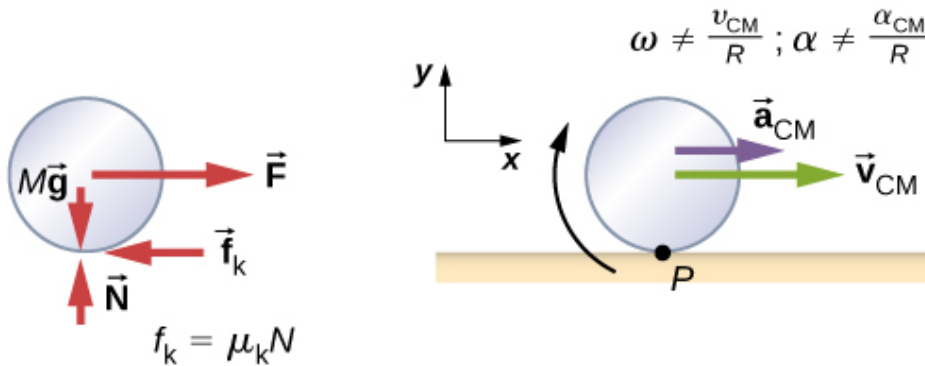
Equation:

$$a_{\text{CM}} = \frac{mg \sin \theta}{m + (I_{\text{CM}}/r^2)}.$$

This is a very useful equation for solving problems involving rolling without slipping. Note that the acceleration is less than that for an object sliding down a frictionless plane with no rotation. The acceleration will also be different for two rotating objects with different rotational inertias.

Rolling Motion with Slipping

In the case of rolling motion with slipping, we must use the coefficient of kinetic friction, which gives rise to the kinetic friction force since static friction is not present. The situation is shown in [\[link\]](#). In the case of slipping, $v_{\text{CM}} - R\omega \neq 0$, because point P on the wheel is not at rest on the surface, and $v_P \neq 0$. Thus, $\omega \neq \frac{v_{\text{CM}}}{R}$, $\alpha \neq \frac{a_{\text{CM}}}{R}$.



(a) Forces on wheel

(b) Wheel is rolling and slipping

(a) Kinetic friction arises between the wheel and the surface because the wheel is slipping. (b) The simple relationships between the linear and angular variables are no longer valid.

Example:

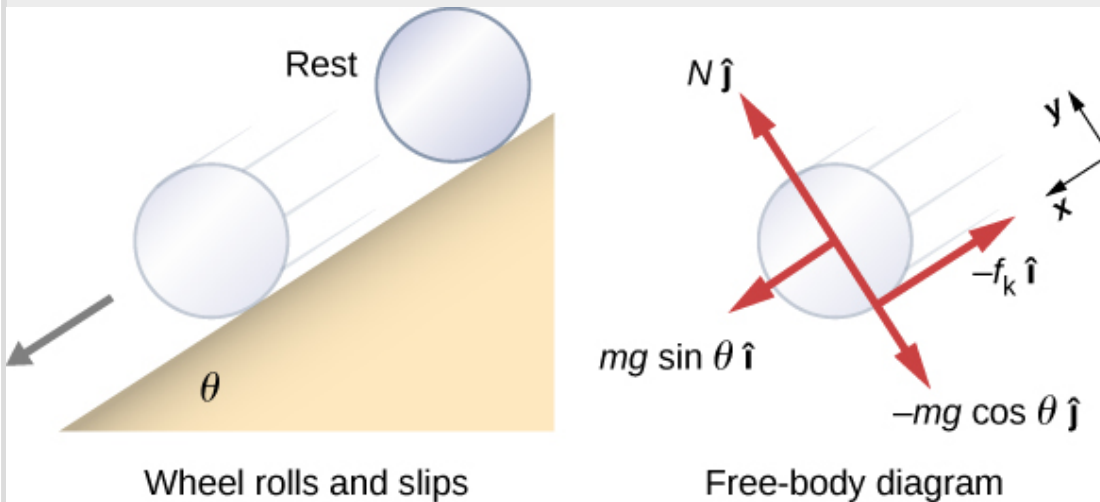
Rolling Down an Inclined Plane with Slipping

A solid cylinder rolls down an inclined plane from rest and undergoes slipping ([\[link\]](#)). It has mass m and radius r . (a) What is its linear acceleration? (b) What is its angular acceleration about an axis through the center of mass?

Strategy

Draw a sketch and free-body diagram showing the forces involved. The free-body diagram is similar to the no-slipping case except for the friction force, which is kinetic instead of static. Use Newton's second law to solve for the acceleration in the x -direction. Use Newton's second law of rotation to solve for the angular acceleration.

Solution



A solid cylinder rolls down an inclined plane from rest and undergoes slipping. The coordinate system has x in the direction down the inclined plane and y upward perpendicular to the plane. The free-body diagram shows the normal force, kinetic friction force, and the components of the weight $m\vec{g}$.

The sum of the forces in the y -direction is zero, so the friction force is now $f_k = \mu_k N = \mu_k mg \cos \theta$.

Newton's second law in the x -direction becomes

Equation:

$$\sum F_x = ma_x,$$

Equation:

$$mg \sin \theta - \mu_k mg \cos \theta = m(a_{\text{CM}})_x,$$

or

Equation:

$$(a_{\text{CM}})_x = g(\sin \theta - \mu_k \cos \theta).$$

The friction force provides the only torque about the axis through the center of mass, so Newton's second law of rotation becomes

Equation:

$$\sum \tau_{\text{CM}} = I_{\text{CM}}\alpha,$$

Equation:

$$f_k r = I_{\text{CM}}\alpha = \frac{1}{2}mr^2\alpha.$$

Solving for α , we have

Equation:

$$\alpha = \frac{2f_k}{mr} = \frac{2\mu_k g \cos \theta}{r}.$$

Significance

We write the linear and angular accelerations in terms of the coefficient of kinetic friction. The linear acceleration is the same as that found for an object sliding down an inclined plane with kinetic friction. The angular acceleration about the axis of rotation is linearly proportional to the normal force, which depends on the cosine of the angle of inclination. As $\theta \rightarrow 90^\circ$, this force goes to zero, and, thus, the angular acceleration goes to zero.

Conservation of Mechanical Energy in Rolling Motion

In the preceding chapter, we introduced rotational kinetic energy. Any rolling object carries rotational kinetic energy, as well as translational kinetic energy and potential energy if the system requires. Including the gravitational potential energy, the total mechanical energy of an object rolling is

Equation:

$$E_T = \frac{1}{2}mv_{\text{CM}}^2 + \frac{1}{2}I_{\text{CM}}\omega^2 + mgh.$$

In the absence of any nonconservative forces that would take energy out of the system in the form of heat, the total energy of a rolling object without slipping is conserved and is constant throughout the motion. Examples where energy is not conserved are a rolling object that is slipping, production of heat as a result of kinetic friction, and a rolling object encountering air resistance.

You may ask why a rolling object that is not slipping conserves energy, since the static friction force is nonconservative. The answer can be found by referring back to [\[link\]](#). Point P in contact with the surface is at rest with respect to the surface. Therefore, its infinitesimal displacement $d\vec{r}$ with respect to the surface is zero, and the incremental work done by the static friction force is zero. We can apply energy conservation to our study of rolling motion to bring out some interesting results.

Example:

Curiosity Rover

The *Curiosity* rover, shown in [\[link\]](#), was deployed on Mars on August 6, 2012. The wheels of the rover have a radius of 25 cm. Suppose astronauts arrive on Mars in the year 2050 and find the now-inoperative *Curiosity* on the side of a basin. While they are dismantling the rover, an astronaut accidentally loses a grip on one of the wheels, which rolls without slipping down into the bottom of the basin 25 meters below. If the wheel has a mass of 5 kg, what is its velocity at the bottom of the basin?



The NASA Mars Science Laboratory rover *Curiosity* during testing on June 3, 2011. The location is inside the Spacecraft Assembly Facility at NASA's Jet Propulsion Laboratory in Pasadena, California. (credit: NASA/JPL-Caltech)

Strategy

We use mechanical energy conservation to analyze the problem. At the top of the hill, the wheel is at rest and has only potential energy. At the bottom of the

basin, the wheel has rotational and translational kinetic energy, which must be equal to the initial potential energy by energy conservation. Since the wheel is rolling without slipping, we use the relation $v_{\text{CM}} = r\omega$ to relate the translational variables to the rotational variables in the energy conservation equation. We then solve for the velocity. From [\[link\]](#), we see that a hollow cylinder is a good approximation for the wheel, so we can use this moment of inertia to simplify the calculation.

Solution

Energy at the top of the basin equals energy at the bottom:

Equation:

$$mgh = \frac{1}{2}mv_{\text{CM}}^2 + \frac{1}{2}I_{\text{CM}}\omega^2.$$

The known quantities are $I_{\text{CM}} = mr^2$, $r = 0.25$ m, and $h = 25.0$ m.

We rewrite the energy conservation equation eliminating ω by using $\omega = \frac{v_{\text{CM}}}{r}$.

We have

Equation:

$$mgh = \frac{1}{2}mv_{\text{CM}}^2 + \frac{1}{2}mr^2 \frac{v_{\text{CM}}^2}{r^2}$$

or

Equation:

$$gh = \frac{1}{2}v_{\text{CM}}^2 + \frac{1}{2}v_{\text{CM}}^2 \Rightarrow v_{\text{CM}} = \sqrt{gh}.$$

On Mars, the acceleration of gravity is 3.71 m/s^2 , which gives the magnitude of the velocity at the bottom of the basin as

Equation:

$$v_{\text{CM}} = \sqrt{(3.71 \text{ m/s}^2)25.0 \text{ m}} = 9.63 \text{ m/s}.$$

Significance

This is a fairly accurate result considering that Mars has very little atmosphere, and the loss of energy due to air resistance would be minimal. The result also assumes that the terrain is smooth, such that the wheel wouldn't encounter rocks and bumps along the way.

Also, in this example, the kinetic energy, or energy of motion, is equally shared between linear and rotational motion. If we look at the moments of inertia in [\[link\]](#), we see that the hollow cylinder has the largest moment of inertia for a given radius and mass. If the wheels of the rover were solid and approximated by solid cylinders, for example, there would be more kinetic energy in linear motion than in rotational motion. This would give the wheel a larger linear velocity than the hollow cylinder approximation. Thus, the solid cylinder would reach the bottom of the basin faster than the hollow cylinder.

Summary

- In rolling motion without slipping, a static friction force is present between the rolling object and the surface. The relations $v_{\text{CM}} = R\omega$, $a_{\text{CM}} = R\alpha$, and $d_{\text{CM}} = R\theta$ all apply, such that the linear velocity, acceleration, and distance of the center of mass are the angular variables multiplied by the radius of the object.
- In rolling motion with slipping, a kinetic friction force arises between the rolling object and the surface. In this case, $v_{\text{CM}} \neq R\omega$, $a_{\text{CM}} \neq R\alpha$, and $d_{\text{CM}} \neq R\theta$.
- Energy conservation can be used to analyze rolling motion. Energy is conserved in rolling motion without slipping. Energy is not conserved in rolling motion with slipping due to the heat generated by kinetic friction.

Conceptual Questions

Exercise:

Problem:

Can a round object released from rest at the top of a frictionless incline undergo rolling motion?

Solution:

No, the static friction force is zero.

Exercise:

Problem:

A cylindrical can of radius R is rolling across a horizontal surface without slipping. (a) After one complete revolution of the can, what is the distance that its center of mass has moved? (b) Would this distance be greater or smaller if slipping occurred?

Exercise:**Problem:**

A wheel is released from the top on an incline. Is the wheel most likely to slip if the incline is steep or gently sloped?

Solution:

The wheel is more likely to slip on a steep incline since the coefficient of static friction must increase with the angle to keep rolling motion without slipping.

Exercise:**Problem:**

Which rolls down an inclined plane faster, a hollow cylinder or a solid sphere? Both have the same mass and radius.

Exercise:**Problem:**

A hollow sphere and a hollow cylinder of the same radius and mass roll up an incline without slipping and have the same initial center of mass velocity. Which object reaches a greater height before stopping?

Solution:

The cylinder reaches a greater height. By [\[link\]](#), its acceleration in the direction down the incline would be less.

Problems

Exercise:**Problem:**

What is the angular velocity of a 75.0-cm-diameter tire on an automobile traveling at 90.0 km/h?

Solution:

$$v_{\text{CM}} = R\omega \Rightarrow \omega = 66.7 \text{ rad/s}$$

Exercise:**Problem:**

A boy rides his bicycle 2.00 km. The wheels have radius 30.0 cm. What is the total angle the tires rotate through during his trip?

Exercise:**Problem:**

If the boy on the bicycle in the preceding problem accelerates from rest to a speed of 10.0 m/s in 10.0 s, what is the angular acceleration of the tires?

Solution:

$$\alpha = 3.3 \text{ rad/s}^2$$

Exercise:**Problem:**

Formula One race cars have 66-cm-diameter tires. If a Formula One averages a speed of 300 km/h during a race, what is the angular displacement in revolutions of the wheels if the race car maintains this speed for 1.5 hours?

Exercise:**Problem:**

A marble rolls down an incline at 30° from rest. (a) What is its acceleration? (b) How far does it go in 3.0 s?

Solution:

$$I_{\text{CM}} = \frac{2}{5}mr^2, a_{\text{CM}} = 3.5 \text{ m/s}^2; x = 15.75 \text{ m}$$

Exercise:**Problem:**

Repeat the preceding problem replacing the marble with a solid cylinder. Explain the new result.

Exercise:**Problem:**

A rigid body with a cylindrical cross-section is released from the top of a 30° incline. It rolls 10.0 m to the bottom in 2.60 s. Find the moment of inertia of the body in terms of its mass m and radius r .

Solution:

positive is down the incline plane;

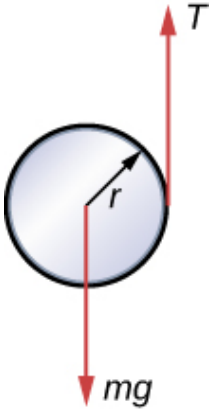
$$a_{\text{CM}} = \frac{mg \sin \theta}{m + (I_{\text{CM}}/r^2)} \Rightarrow I_{\text{CM}} = r^2 \left[\frac{mg \sin 30}{a_{\text{CM}}} - m \right],$$

$$x - x_0 = v_0 t - \frac{1}{2} a_{\text{CM}} t^2 \Rightarrow a_{\text{CM}} = 2.96 \text{ m/s}^2,$$

$$I_{\text{CM}} = 0.66 mr^2$$

Exercise:**Problem:**

A yo-yo can be thought of a solid cylinder of mass m and radius r that has a light string wrapped around its circumference (see below). One end of the string is held fixed in space. If the cylinder falls as the string unwinds without slipping, what is the acceleration of the cylinder?



Exercise:

Problem:

A solid cylinder of radius 10.0 cm rolls down an incline with slipping. The angle of the incline is 30° . The coefficient of kinetic friction on the surface is 0.400. What is the angular acceleration of the solid cylinder? What is the linear acceleration?

Solution:

$$\alpha = 67.9 \text{ rad/s}^2,$$
$$(a_{\text{CM}})_x = 1.5 \text{ m/s}^2$$

Exercise:

Problem:

A bowling ball rolls up a ramp 0.5 m high without slipping to storage. It has an initial velocity of its center of mass of 3.0 m/s. (a) What is its velocity at the top of the ramp? (b) If the ramp is 1 m high does it make it to the top?

Exercise:

Problem:

A 40.0-kg solid cylinder is rolling across a horizontal surface at a speed of 6.0 m/s. How much work is required to stop it?

Solution:

$$W = -1080.0 \text{ J}$$

Exercise:

Problem:

A 40.0-kg solid sphere is rolling across a horizontal surface with a speed of 6.0 m/s. How much work is required to stop it? Compare results with the preceding problem.

Exercise:

Problem:

A solid cylinder rolls up an incline at an angle of 20° . If it starts at the bottom with a speed of 10 m/s, how far up the incline does it travel?

Solution:

Mechanical energy at the bottom equals mechanical energy at the top;

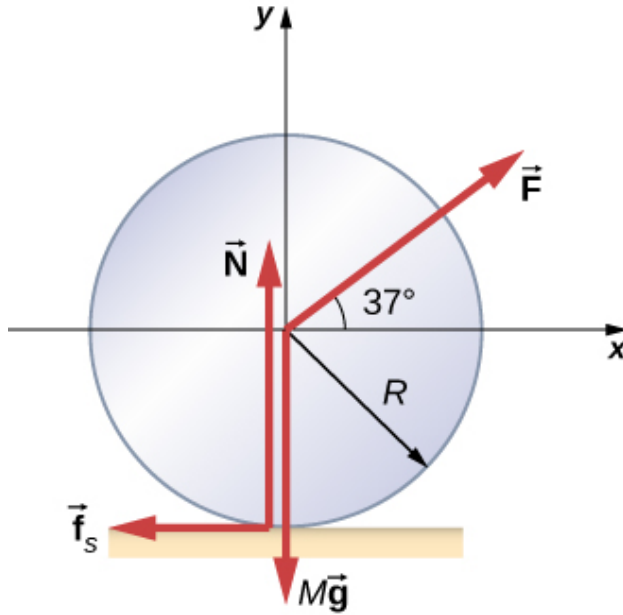
$$\frac{1}{2}mv_0^2 + \frac{1}{2}\left(\frac{1}{2}mr^2\right)\left(\frac{v_0}{r}\right)^2 = mgh \Rightarrow h = \frac{1}{g}\left(\frac{1}{2} + \frac{1}{4}\right)v_0^2,$$

$h = 7.7 \text{ m}$, so the distance up the incline is 22.5 m.

Exercise:

Problem:

A solid cylindrical wheel of mass M and radius R is pulled by a force $\vec{\mathbf{F}}$ applied to the center of the wheel at 37° to the horizontal (see the following figure). If the wheel is to roll without slipping, what is the maximum value of $|\vec{\mathbf{F}}|$? The coefficients of static and kinetic friction are $\mu_S = 0.40$ and $\mu_k = 0.30$.



Exercise:

Problem:

A hollow cylinder that is rolling without slipping is given a velocity of 5.0 m/s and rolls up an incline to a vertical height of 1.0 m. If a hollow sphere of the same mass and radius is given the same initial velocity, how high vertically does it roll up the incline?

Solution:

Use energy conservation

$$\frac{1}{2}mv_0^2 + \frac{1}{2}I_{\text{Cyl}}\omega_0^2 = mgh_{\text{Cyl}},$$

$$\frac{1}{2}mv_0^2 + \frac{1}{2}I_{\text{Sph}}\omega_0^2 = mgh_{\text{Sph}}.$$

Subtracting the two equations, eliminating the initial translational energy, we have

$$\frac{1}{2}I_{\text{Cyl}}\omega_0^2 - \frac{1}{2}I_{\text{Sph}}\omega_0^2 = mg(h_{\text{Cyl}} - h_{\text{Sph}}),$$

$$\frac{1}{2}mr^2\left(\frac{v_0}{r}\right)^2 - \frac{1}{2}\frac{2}{3}mr^2\left(\frac{v_0}{r}\right)^2 = mg(h_{\text{Cyl}} - h_{\text{Sph}}),$$

$$\frac{1}{2}v_0^2 - \frac{1}{2}\frac{2}{3}v_0^2 = g(h_{\text{Cyl}} - h_{\text{Sph}}),$$

$$h_{\text{Cyl}} - h_{\text{Sph}} = \frac{1}{g}\left(\frac{1}{2} - \frac{1}{3}\right)v_0^2 = \frac{1}{9.8 \text{ m/s}^2}\left(\frac{1}{6}\right)(5.0 \text{ m/s})^2 = 0.43 \text{ m}.$$

Thus, the hollow sphere, with the smaller moment of inertia, rolls up to a lower height of $1.0 - 0.43 = 0.57 \text{ m}$.

Glossary

rolling motion

combination of rotational and translational motion with or without slipping

Angular Momentum

By the end of this section, you will be able to:

- Describe the vector nature of angular momentum
- Find the total angular momentum and torque about a designated origin of a system of particles
- Calculate the angular momentum of a rigid body rotating about a fixed axis
- Calculate the torque on a rigid body rotating about a fixed axis
- Use conservation of angular momentum in the analysis of objects that change their rotation rate

Why does Earth keep on spinning? What started it spinning to begin with? Why doesn't Earth's gravitational attraction not bring the Moon crashing in toward Earth? And how does an ice skater manage to spin faster and faster simply by pulling her arms in? Why does she not have to exert a torque to spin faster?

Questions like these have answers based in angular momentum, the rotational analog to linear momentum. In this chapter, we first define and then explore angular momentum from a variety of viewpoints. First, however, we investigate the angular momentum of a single particle. This allows us to develop angular momentum for a system of particles and for a rigid body that is cylindrically symmetric.

Angular Momentum of a Single Particle

[\[link\]](#) shows a particle at a position \vec{r} with linear momentum $\vec{p} = m\vec{v}$ with respect to the origin. Even if the particle is not rotating about the origin, we can still define an angular momentum in terms of the position vector and the linear momentum.

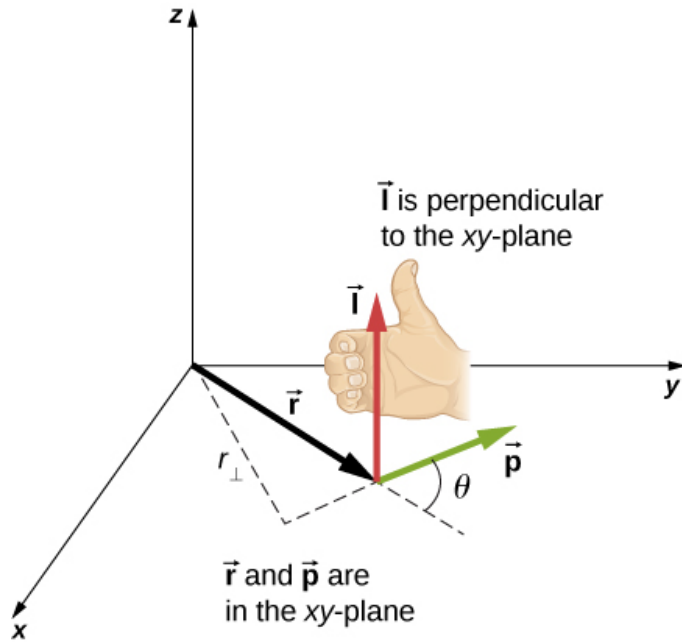
Note:

Angular Momentum of a Particle

The **angular momentum** \vec{l} of a particle is defined as the cross-product of \vec{r} and \vec{p} , and is perpendicular to the plane containing \vec{r} and \vec{p} :

Equation:

$$\vec{l} = \vec{r} \times \vec{p}.$$



In three-dimensional space, the position vector \vec{r} locates a particle in the xy -plane with linear momentum \vec{p} . The angular momentum with respect to the origin is $\vec{L} = \vec{r} \times \vec{p}$, which is in the z -direction. The direction of \vec{L} is given by the right-hand rule, as shown.

The intent of choosing the direction of the angular momentum to be perpendicular to the plane containing \vec{r} and \vec{p} is similar to choosing the direction of torque to be perpendicular to the plane of \vec{r} and \vec{F} , as discussed in [Fixed-Axis Rotation](#). The magnitude of the angular momentum is found from the definition of the cross-product,

Equation:

$$l = rp \sin \theta,$$

where θ is the angle between \vec{r} and \vec{p} . The units of angular momentum are $\text{kg} \cdot \text{m}^2/\text{s}$.

As with the definition of torque, we can define a lever arm r_{\perp} that is the perpendicular distance from the momentum vector \vec{p} to the origin, $r_{\perp} = r \sin \theta$. With this definition, the magnitude of the angular momentum becomes

Equation:

$$l = r_{\perp} p = r_{\perp} mv.$$

We see that if the direction of $\vec{\mathbf{p}}$ is such that it passes through the origin, then $\theta = 0$, and the angular momentum is zero because the lever arm is zero. In this respect, the magnitude of the angular momentum depends on the choice of origin.

If we take the time derivative of the angular momentum, we arrive at an expression for the torque on the particle:

Equation:

$$\frac{d\vec{\mathbf{l}}}{dt} = \frac{d\vec{\mathbf{r}}}{dt} \times \vec{\mathbf{p}} + \vec{\mathbf{r}} \times \frac{d\vec{\mathbf{p}}}{dt} = \vec{\mathbf{v}} \times m\vec{\mathbf{v}} + \vec{\mathbf{r}} \times \frac{d\vec{\mathbf{p}}}{dt} = \vec{\mathbf{r}} \times \frac{d\vec{\mathbf{p}}}{dt}.$$

Here we have used the definition of $\vec{\mathbf{p}}$ and the fact that a vector crossed into itself is zero. From Newton's second law, $\frac{d\vec{\mathbf{p}}}{dt} = \sum \vec{\mathbf{F}}$, the net force acting on the particle, and the definition of the net torque, we can write

Note:

Equation:

$$\frac{d\vec{\mathbf{l}}}{dt} = \sum \vec{\boldsymbol{\tau}}.$$

Note the similarity with the linear result of Newton's second law, $\frac{d\vec{\mathbf{p}}}{dt} = \sum \vec{\mathbf{F}}$. The following problem-solving strategy can serve as a guideline for calculating the angular momentum of a particle.

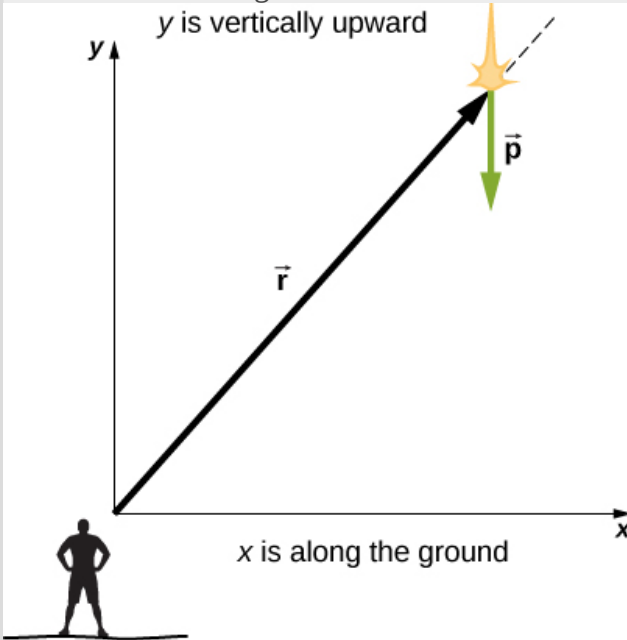
Note:

Angular Momentum of a Particle

1. Choose a coordinate system about which the angular momentum is to be calculated.
2. Write down the radius vector to the point particle in unit vector notation.
3. Write the linear momentum vector of the particle in unit vector notation.
4. Take the cross product $\vec{\mathbf{l}} = \vec{\mathbf{r}} \times \vec{\mathbf{p}}$ and use the right-hand rule to establish the direction of the angular momentum vector.
5. See if there is a time dependence in the expression of the angular momentum vector. If there is, then a torque exists about the origin, and use $\frac{d\vec{\mathbf{l}}}{dt} = \sum \vec{\boldsymbol{\tau}}$ to calculate the torque. If there is no time dependence in the expression for the angular momentum, then the net torque is zero.

Example:**Angular Momentum and Torque on a Meteor**

A meteor enters Earth's atmosphere ([link](#)) and is observed by someone on the ground before it burns up in the atmosphere. The vector $\vec{r} = 25 \text{ km}\hat{i} + 25 \text{ km}\hat{j}$ gives the position of the meteor with respect to the observer. At the instant the observer sees the meteor, it has linear momentum $\vec{p} = 15.0 \text{ kg}(-2.0 \text{ km/s}\hat{j})$, and it is accelerating at a constant $2.0 \text{ m/s}^2(-\hat{j})$ along its path, which for our purposes can be taken as a straight line. (a) What is the angular momentum of the meteor about the origin, which is at the location of the observer? (b) What is the torque on the meteor about the origin?



An observer on the ground sees a meteor at position \vec{r} with linear momentum \vec{p} .

Strategy

We resolve the acceleration into x - and y -components and use the kinematic equations to express the velocity as a function of acceleration and time. We insert these expressions into the linear momentum and then calculate the angular momentum using the cross-product. Since the position and momentum vectors are in the xy -plane, we expect the angular momentum vector to be along the z -axis. To find the torque, we take the time derivative of the angular momentum.

Solution

The meteor is entering Earth's atmosphere at an angle of 90.0° below the horizontal, so the components of the acceleration in the x - and y -directions are

Equation:

$$a_x = 0, \quad a_y = -2.0 \text{ m/s}^2.$$

We write the velocities using the kinematic equations.

Equation:

$$v_x = 0, \quad v_y = -2.0 \times 10^3 \text{ m/s} - (2.0 \text{ m/s}^2)t.$$

a. The angular momentum is

Equation:

$$\begin{aligned}\vec{\mathbf{I}} &= \vec{\mathbf{r}} \times \vec{\mathbf{p}} = (25.0 \text{ km}\hat{\mathbf{i}} + 25.0 \text{ km}\hat{\mathbf{j}}) \times 15.0 \text{ kg}(0\hat{\mathbf{i}} + v_y\hat{\mathbf{j}}) \\ &= 15.0 \text{ kg}[25.0 \text{ km}(v_y)\hat{\mathbf{k}}] \\ &= 15.0 \text{ kg}[2.50 \times 10^4 \text{ m}(-2.0 \times 10^3 \text{ m/s} - (2.0 \text{ m/s}^2)t)\hat{\mathbf{k}}].\end{aligned}$$

At $t = 0$, the angular momentum of the meteor about the origin is

Equation:

$$\vec{\mathbf{I}}_0 = 15.0 \text{ kg}[2.50 \times 10^4 \text{ m}(-2.0 \times 10^3 \text{ m/s})\hat{\mathbf{k}}] = 7.50 \times 10^8 \text{ kg} \cdot \text{m}^2/\text{s}(-\hat{\mathbf{k}}).$$

This is the instant that the observer sees the meteor.

b. To find the torque, we take the time derivative of the angular momentum. Taking the time derivative of $\vec{\mathbf{I}}$ as a function of time, which is the second equation immediately above, we have

Equation:

$$\frac{d\vec{\mathbf{I}}}{dt} = -15.0 \text{ kg}(2.50 \times 10^4 \text{ m})(2.0 \text{ m/s}^2)\hat{\mathbf{k}}.$$

Then, since $\frac{d\vec{\mathbf{I}}}{dt} = \sum \vec{\boldsymbol{\tau}}$, we have

Equation:

$$\sum \vec{\boldsymbol{\tau}} = -7.5 \times 10^5 \text{ N} \cdot \text{m}\hat{\mathbf{k}}.$$

The units of torque are given as newton-meters, not to be confused with joules. As a check, we note that the lever arm is the x -component of the vector $\vec{\mathbf{r}}$ in [\[link\]](#) since it is perpendicular to the force acting on the meteor, which is along its path. By Newton's second law, this force is

Equation:

$$\vec{\mathbf{F}} = ma(-\hat{\mathbf{j}}) = 15.0 \text{ kg}(2.0 \text{ m/s}^2)(-\hat{\mathbf{j}}) = 30.0 \text{ kg} \cdot \text{m/s}^2(-\hat{\mathbf{j}}).$$

The lever arm is

Equation:

$$\vec{\mathbf{r}}_{\perp} = 2.5 \times 10^4 \text{ m}\hat{\mathbf{i}}.$$

Thus, the torque is

Equation:

$$\begin{aligned}\sum \vec{\tau} &= \vec{r}_{\perp} \times \vec{F} = (2.5 \times 10^4 \text{ m } \hat{i}) \times (-30.0 \text{ kg} \cdot \text{m/s}^2 \hat{j}), \\ &= 7.5 \times 10^5 \text{ N} \cdot \text{m}(-\hat{k}).\end{aligned}$$

Significance

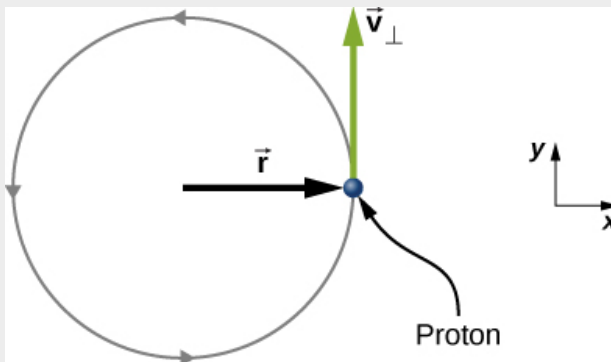
Since the meteor is accelerating downward toward Earth, its radius and velocity vector are changing. Therefore, since $\vec{L} = \vec{r} \times \vec{p}$, the angular momentum is changing as a function of time. The torque on the meteor about the origin, however, is constant, because the lever arm \vec{r}_{\perp} and the force on the meteor are constants. This example is important in that it illustrates that the angular momentum depends on the choice of origin about which it is calculated. The methods used in this example are also important in developing angular momentum for a system of particles and for a rigid body.

Note:

Exercise:

Problem:

Check Your Understanding A proton spiraling around a magnetic field executes circular motion in the plane of the paper, as shown below. The circular path has a radius of 0.4 m and the proton has velocity $4.0 \times 10^6 \text{ m/s}$. What is the angular momentum of the proton about the origin?



Solution:

From the figure, we see that the cross product of the radius vector with the momentum vector gives a vector directed out of the page. Inserting the radius and momentum into the expression for the angular momentum, we have

$$\vec{L} = \vec{r} \times \vec{p} = (0.4 \text{ m } \hat{i}) \times (1.67 \times 10^{-27} \text{ kg}(4.0 \times 10^6 \text{ m/s})\hat{j}) = 2.7 \times 10^{-21} \text{ kg} \cdot \text{m}^2/\text{s}\hat{k}$$

Angular Momentum of a System of Particles

The angular momentum of a system of particles is important in many scientific disciplines, one being astronomy. Consider a spiral galaxy, a rotating island of stars like our own Milky Way. The individual stars can be treated as point particles, each of which has its own angular momentum. The vector sum of the individual angular momenta give the total angular momentum of the galaxy. In this section, we develop the tools with which we can calculate the total angular momentum of a system of particles.

In the preceding section, we introduced the angular momentum of a single particle about a designated origin. The expression for this angular momentum is $\vec{\mathbf{L}} = \vec{\mathbf{r}} \times \vec{\mathbf{p}}$, where the vector $\vec{\mathbf{r}}$ is from the origin to the particle, and $\vec{\mathbf{p}}$ is the particle's linear momentum. If we have a system of N particles, each with position vector from the origin given by $\vec{\mathbf{r}}_i$ and each having momentum $\vec{\mathbf{p}}_i$, then the total angular momentum of the system of particles about the origin is the vector sum of the individual angular momenta about the origin. That is,

Note:

Equation:

$$\vec{\mathbf{L}} = \vec{\mathbf{L}}_1 + \vec{\mathbf{L}}_2 + \cdots + \vec{\mathbf{L}}_N.$$

Similarly, if particle i is subject to a net torque $\vec{\tau}_i$ about the origin, then we can find the net torque about the origin due to the system of particles by differentiating [\[link\]](#):

Equation:

$$\frac{d\vec{\mathbf{L}}}{dt} = \sum_i \frac{d\vec{\mathbf{L}}_i}{dt} = \sum_i \vec{\tau}_i.$$

The sum of the individual torques produces a net external torque on the system, which we designate $\sum \vec{\tau}$. Thus,

Note:

Equation:

$$\frac{d\vec{\mathbf{L}}}{dt} = \sum \vec{\tau}.$$

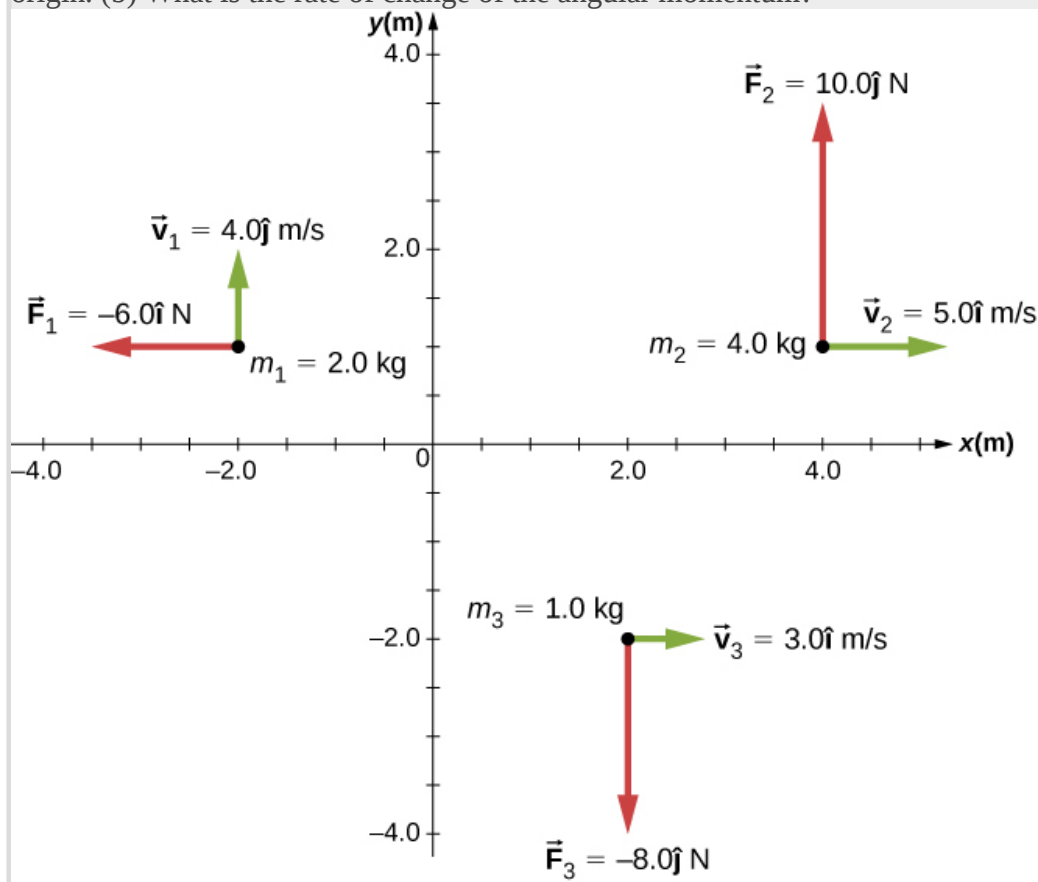
[\[link\]](#) states that *the rate of change of the total angular momentum of a system is equal to the net external torque acting on the system when both quantities are measured with respect to a given*

origin. [link](#) can be applied to any system that has net angular momentum, including rigid bodies, as discussed in the next section.

Example:

Angular Momentum of Three Particles

Referring to [link](#)(a), determine the total angular momentum due to the three particles about the origin. (b) What is the rate of change of the angular momentum?



Three particles in the xy-plane with different position and momentum vectors.

Strategy

Write down the position and momentum vectors for the three particles. Calculate the individual angular momenta and add them as vectors to find the total angular momentum. Then do the same for the torques.

Solution

a. Particle 1: $\vec{r}_1 = -2.0\hat{i} + 1.0\hat{j}$, $\vec{p}_1 = 2.0\text{ kg}(4.0\text{ m/s}\hat{j}) = 8.0\text{ kg} \cdot \text{m/s}\hat{j}$,

Equation:

$$\vec{L}_1 = \vec{r}_1 \times \vec{p}_1 = -16.0 \text{ kg} \cdot \text{m}^2/\text{s} \hat{\mathbf{k}}.$$

Particle 2: $\vec{r}_2 = 4.0 \text{ m} \hat{\mathbf{i}} + 1.0 \text{ m} \hat{\mathbf{j}}$, $\vec{p}_2 = 4.0 \text{ kg}(5.0 \text{ m/s} \hat{\mathbf{i}}) = 20.0 \text{ kg} \cdot \text{m/s} \hat{\mathbf{i}}$,

Equation:

$$\vec{L}_2 = \vec{r}_2 \times \vec{p}_2 = -20.0 \text{ kg} \cdot \text{m}^2/\text{s} \hat{\mathbf{k}}.$$

Particle 3: $\vec{r}_3 = 2.0 \text{ m} \hat{\mathbf{i}} - 2.0 \text{ m} \hat{\mathbf{j}}$, $\vec{p}_3 = 1.0 \text{ kg}(3.0 \text{ m/s} \hat{\mathbf{i}}) = 3.0 \text{ kg} \cdot \text{m/s} \hat{\mathbf{i}}$,

Equation:

$$\vec{L}_3 = \vec{r}_3 \times \vec{p}_3 = 6.0 \text{ kg} \cdot \text{m}^2/\text{s} \hat{\mathbf{k}}.$$

We add the individual angular momenta to find the total about the origin:

Equation:

$$\vec{L}_T = \vec{L}_1 + \vec{L}_2 + \vec{L}_3 = -30 \text{ kg} \cdot \text{m}^2/\text{s} \hat{\mathbf{k}}.$$

b. The individual forces and lever arms are

Equation:

$$\vec{r}_{1\perp} = 1.0 \text{ m} \hat{\mathbf{j}}, \quad \vec{F}_1 = -6.0 \text{ N} \hat{\mathbf{i}}, \quad \vec{\tau}_1 = 6.0 \text{ N} \cdot \text{m} \hat{\mathbf{k}}$$

$$\vec{r}_{2\perp} = 4.0 \text{ m} \hat{\mathbf{i}}, \quad \vec{F}_2 = 10.0 \text{ N} \hat{\mathbf{j}}, \quad \vec{\tau}_2 = 40.0 \text{ N} \cdot \text{m} \hat{\mathbf{k}}$$

$$\vec{r}_{3\perp} = 2.0 \text{ m} \hat{\mathbf{i}}, \quad \vec{F}_3 = -8.0 \text{ N} \hat{\mathbf{j}}, \quad \vec{\tau}_3 = -16.0 \text{ N} \cdot \text{m} \hat{\mathbf{k}}.$$

Therefore:

Equation:

$$\sum_i \vec{\tau}_i = \vec{\tau}_1 + \vec{\tau}_2 + \vec{\tau}_3 = 30 \text{ N} \cdot \text{m} \hat{\mathbf{k}}.$$

Significance

This example illustrates the superposition principle for angular momentum and torque of a system of particles. Care must be taken when evaluating the radius vectors \vec{r}_i of the particles to calculate the angular momenta, and the lever arms, $\vec{r}_{i\perp}$ to calculate the torques, as they are completely different quantities.

Angular Momentum of a Rigid Body

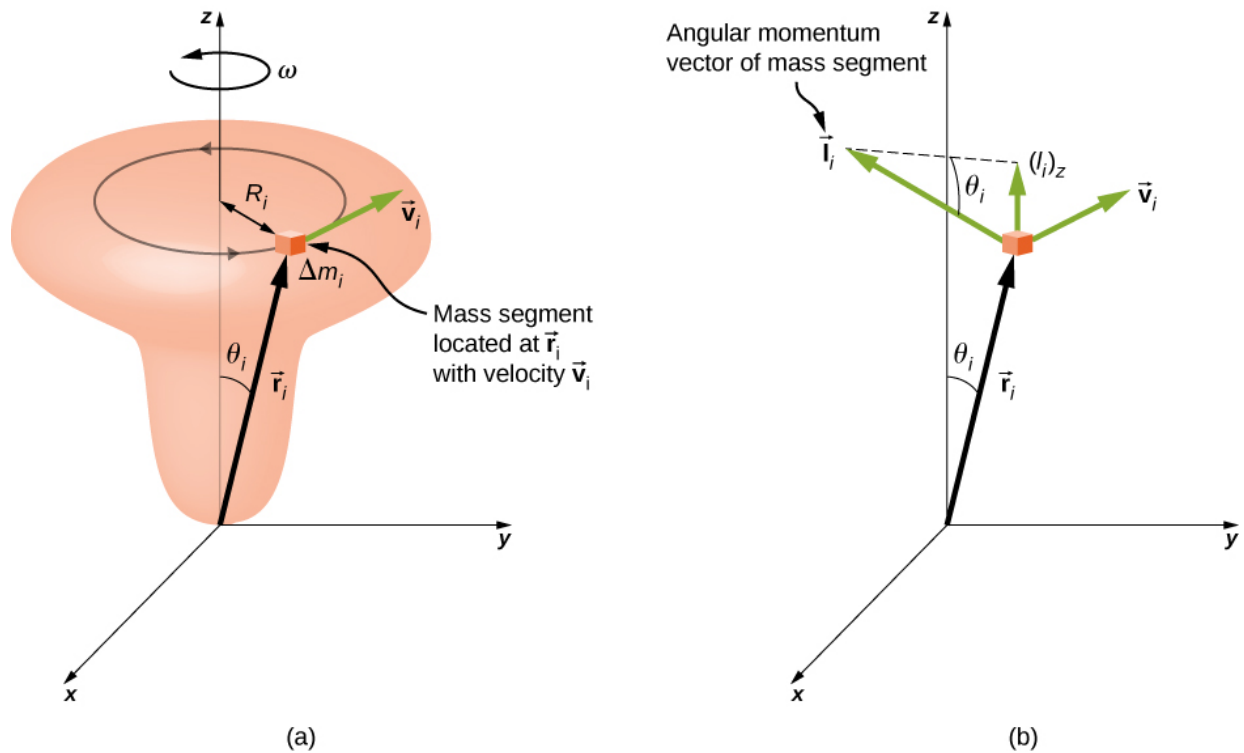
We have investigated the angular momentum of a single particle, which we generalized to a system of particles. Now we can use the principles discussed in the previous section to develop the concept of the angular momentum of a rigid body. Celestial objects such as planets have angular momentum due to their spin and orbits around stars. In engineering, anything that rotates about an axis carries angular momentum, such as flywheels, propellers, and rotating parts in

engines. Knowledge of the angular momenta of these objects is crucial to the design of the system in which they are a part.

To develop the angular momentum of a rigid body, we model a rigid body as being made up of small mass segments, Δm_i . In [\[link\]](#), a rigid body is constrained to rotate about the z-axis with angular velocity ω . All mass segments that make up the rigid body undergo circular motion about the z-axis with the same angular velocity. Part (a) of the figure shows mass segment Δm_i with position vector \vec{r}_i from the origin and radius R_i to the z-axis. The magnitude of its tangential velocity is $v_i = R_i\omega$. Because the vectors \vec{v}_i and \vec{r}_i are perpendicular to each other, the magnitude of the angular momentum of this mass segment is

Equation:

$$l_i = r_i(\Delta m v_i)\sin 90^\circ.$$



(a) A rigid body is constrained to rotate around the z-axis. The rigid body is symmetrical about the z-axis. A mass segment Δm_i is located at position \vec{r}_i , which makes angle θ_i with respect to the z-axis. The circular motion of an infinitesimal mass segment is shown. (b) \vec{l}_i is the angular momentum of the mass segment and has a component along the z-axis $(\vec{l}_i)_z$.

Using the right-hand rule, the angular momentum vector points in the direction shown in part (b). The sum of the angular momenta of all the mass segments contains components both along and

perpendicular to the axis of rotation. Every mass segment has a perpendicular component of the angular momentum that will be cancelled by the perpendicular component of an identical mass segment on the opposite side of the rigid body, because it is cylindrically symmetric. Thus, the component along the axis of rotation is the only component that gives a nonzero value when summed over all the mass segments. From part (b), the component of \vec{l}_i along the axis of rotation is

Equation:

$$\begin{aligned}(l_i)_z &= l_i \sin \theta_i = (r_i \Delta m_i v_i) \sin \theta_i, \\ &= (r_i \sin \theta_i) (\Delta m_i v_i) = R_i \Delta m_i v_i.\end{aligned}$$

The net angular momentum of the rigid body along the axis of rotation is

Equation:

$$L = \sum_i (\vec{l}_i)_z = \sum_i R_i \Delta m_i v_i = \sum_i R_i \Delta m_i (R_i \omega) = \omega \sum_i \Delta m_i (R_i)^2.$$

The summation $\sum_i \Delta m_i (R_i)^2$ is simply the moment of inertia I of the rigid body about the axis of rotation. For a thin hoop rotating about an axis perpendicular to the plane of the hoop, all of the R_i 's are equal to R so the summation reduces to $R^2 \sum_i \Delta m_i = mR^2$, which is the moment of inertia for a thin hoop found in [\[link\]](#). Thus, the magnitude of the angular momentum along the axis of rotation of a rigid body rotating with angular velocity ω about the axis is

Note:

Equation:

$$L = I\omega.$$

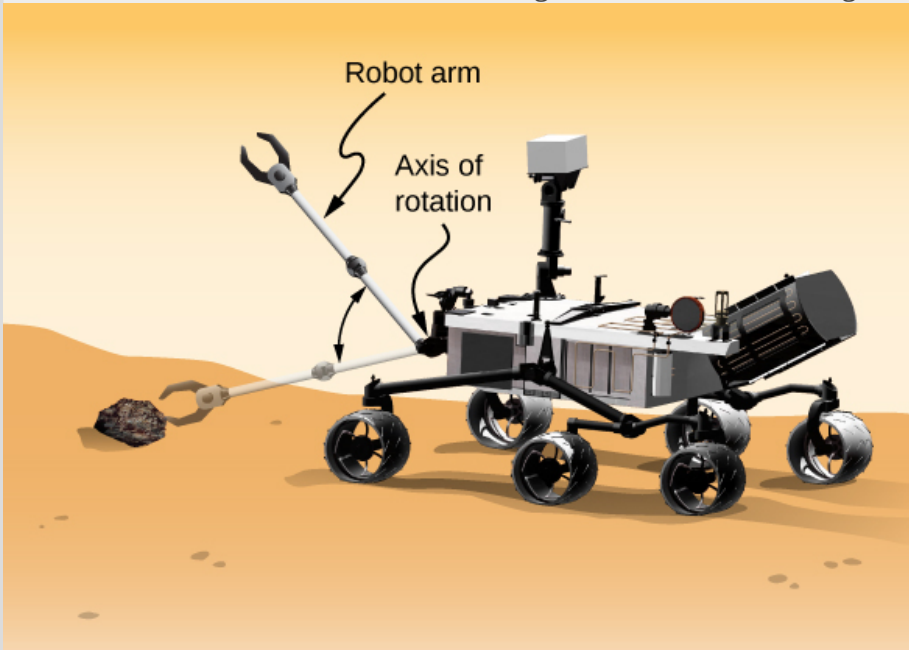
This equation is analogous to the magnitude of the linear momentum $p = mv$. The direction of the angular momentum vector is directed along the axis of rotation given by the right-hand rule.

Example:

Angular Momentum of a Robot Arm

A robot arm on a Mars rover like *Curiosity* shown in [\[link\]](#) is 1.0 m long and has forceps at the free end to pick up rocks. The mass of the arm is 2.0 kg and the mass of the forceps is 1.0 kg. See [\[link\]](#). The robot arm and forceps move from rest to $\omega = 0.1\pi$ rad/s in 0.1 s. It rotates down and picks up a Mars rock that has mass 1.5 kg. The axis of rotation is the point where the robot arm connects to the rover. (a) What is the angular momentum of the robot arm by itself about the axis

of rotation after 0.1 s when the arm has stopped accelerating? (b) What is the angular momentum of the robot arm when it has the Mars rock in its forceps and is rotating upwards? (c) When the arm does not have a rock in the forceps, what is the torque about the point where the arm connects to the rover when it is accelerating from rest to its final angular velocity?



A robot arm on a Mars rover swings down and picks up a Mars rock. (credit: modification of work by NASA/JPL-Caltech)

Strategy

We use [\[link\]](#) to find angular momentum in the various configurations. When the arm is rotating downward, the right-hand rule gives the angular momentum vector directed out of the page, which we will call the positive z-direction. When the arm is rotating upward, the right-hand rule gives the direction of the angular momentum vector into the page or in the negative z-direction. The moment of inertia is the sum of the individual moments of inertia. The arm can be approximated with a solid rod, and the forceps and Mars rock can be approximated as point masses located at a distance of 1 m from the origin. For part (c), we use Newton's second law of motion for rotation to find the torque on the robot arm.

Solution

- a. Writing down the individual moments of inertia, we have

$$\text{Robot arm: } I_R = \frac{1}{3} m_R r^2 = \frac{1}{3} (2.00 \text{ kg})(1.00 \text{ m})^2 = \frac{2}{3} \text{ kg} \cdot \text{m}^2.$$

$$\text{Forceps: } I_F = m_F r^2 = (1.0 \text{ kg})(1.0 \text{ m})^2 = 1.0 \text{ kg} \cdot \text{m}^2.$$

$$\text{Mars rock: } I_{MR} = m_{MR} r^2 = (1.5 \text{ kg})(1.0 \text{ m})^2 = 1.5 \text{ kg} \cdot \text{m}^2.$$

Therefore, without the Mars rock, the total moment of inertia is

Equation:

$$I_{\text{Total}} = I_R + I_F = 1.67 \text{ kg} \cdot \text{m}^2$$

and the magnitude of the angular momentum is

Equation:

$$L = I\omega = 1.67 \text{ kg} \cdot \text{m}^2(0.1\pi \text{ rad/s}) = 0.17\pi \text{ kg} \cdot \text{m}^2/\text{s}.$$

The angular momentum vector is directed out of the page in the $\hat{\mathbf{k}}$ direction since the robot arm is rotating counterclockwise.

- b. We must include the Mars rock in the calculation of the moment of inertia, so we have

Equation:

$$I_{\text{Total}} = I_{\text{R}} + I_{\text{F}} + I_{\text{MR}} = 3.17 \text{ kg} \cdot \text{m}^2$$

and

Equation:

$$L = I\omega = 3.17 \text{ kg} \cdot \text{m}^2(0.1\pi \text{ rad/s}) = 0.32\pi \text{ kg} \cdot \text{m}^2/\text{s}.$$

Now the angular momentum vector is directed into the page in the $-\hat{\mathbf{k}}$ direction, by the right-hand rule, since the robot arm is now rotating clockwise.

- c. We find the torque when the arm does not have the rock by taking the derivative of the angular momentum using [\[link\]](#) $\frac{d\vec{L}}{dt} = \sum \vec{\tau}$. But since $L = I\omega$, and understanding that the direction of the angular momentum and torque vectors are along the axis of rotation, we can suppress the vector notation and find

Equation:

$$\frac{dL}{dt} = \frac{d(I\omega)}{dt} = I \frac{d\omega}{dt} = I\alpha = \sum \tau,$$

which is Newton's second law for rotation. Since $\alpha = \frac{0.1\pi \text{ rad/s}}{0.1 \text{ s}} = \pi \text{ rad/s}^2$, we can calculate the net torque:

Equation:

$$\sum \tau = I\alpha = 1.67 \text{ kg} \cdot \text{m}^2(\pi \text{ rad/s}^2) = 1.67\pi \text{ N} \cdot \text{m}.$$

Significance

The angular momentum in (a) is less than that of (b) due to the fact that the moment of inertia in (b) is greater than (a), while the angular velocity is the same.

Note:

Exercise:

Problem:

Check Your Understanding Which has greater angular momentum: a solid sphere of mass m rotating at a constant angular frequency ω_0 about the z-axis, or a solid cylinder of same mass and rotation rate about the z-axis?

Solution:

$I_{\text{sphere}} = \frac{2}{5}mr^2$, $I_{\text{cylinder}} = \frac{1}{2}mr^2$; Taking the ratio of the angular momenta, we have:

$$\frac{L_{\text{cylinder}}}{L_{\text{sphere}}} = \frac{I_{\text{cylinder}}\omega_0}{I_{\text{sphere}}\omega_0} = \frac{\frac{1}{2}mr^2}{\frac{2}{5}mr^2} = \frac{5}{4}. \text{ Thus, the cylinder has 25\% more angular momentum.}$$

This is because the cylinder has more mass distributed farther from the axis of rotation.

Note:

Visit the [University of Colorado's Interactive Simulation of Angular Momentum](#) to learn more about angular momentum.

Summary

- The angular momentum $\vec{L} = \vec{r} \times \vec{p}$ of a single particle about a designated origin is the vector product of the position vector in the given coordinate system and the particle's linear momentum.
- The angular momentum $\vec{L} = \sum_i \vec{L}_i$ of a system of particles about a designated origin is the vector sum of the individual momenta of the particles that make up the system.
- The net torque on a system about a given origin is the time derivative of the angular momentum about that origin: $\frac{d\vec{L}}{dt} = \sum \vec{\tau}$.
- A rigid rotating body has angular momentum $L = I\omega$ directed along the axis of rotation. The time derivative of the angular momentum $\frac{dL}{dt} = \sum \tau$ gives the net torque on a rigid body and is directed along the axis of rotation.

Conceptual Questions**Exercise:****Problem:**

Can you assign an angular momentum to a particle without first defining a reference point?

Exercise:

Problem:

For a particle traveling in a straight line, are there any points about which the angular momentum is zero? Assume the line intersects the origin.

Solution:

All points on the straight line will give zero angular momentum, because a vector crossed into a parallel vector is zero.

Exercise:**Problem:**

Under what conditions does a rigid body have angular momentum but not linear momentum?

Exercise:**Problem:**

If a particle is moving with respect to a chosen origin it has linear momentum. What conditions must exist for this particle's angular momentum to be zero about the chosen origin?

Solution:

The particle must be moving on a straight line that passes through the chosen origin.

Exercise:**Problem:**

If you know the velocity of a particle, can you say anything about the particle's angular momentum?

Problems**Exercise:****Problem:**

A 0.2-kg particle is travelling along the line $y = 2.0$ m with a velocity 5.0 m/s. What is the angular momentum of the particle about the origin?

Exercise:**Problem:**

A bird flies overhead from where you stand at an altitude of 300.0 m and at a speed horizontal to the ground of 20.0 m/s. The bird has a mass of 2.0 kg. The radius vector to the bird makes an angle θ with respect to the ground. The radius vector to the bird and its momentum vector lie in the xy -plane. What is the bird's angular momentum about the point where you are standing?

Solution:

The magnitude of the cross product of the radius to the bird and its momentum vector yields $rp \sin \theta$, which gives $r \sin \theta$ as the altitude of the bird h . The direction of the angular momentum is perpendicular to the radius and momentum vectors, which we choose arbitrarily as $\hat{\mathbf{k}}$, which is in the plane of the ground:

$$\vec{\mathbf{L}} = \vec{\mathbf{r}} \times \vec{\mathbf{p}} = hmv\hat{\mathbf{k}} = (300.0 \text{ m})(2.0 \text{ kg})(20.0 \text{ m/s})\hat{\mathbf{k}} = 12,000.0 \text{ kg} \cdot \text{m}^2/\text{s}\hat{\mathbf{k}}$$

Exercise:**Problem:**

A Formula One race car with mass 750.0 kg is speeding through a course in Monaco and enters a circular turn at 220.0 km/h in the counterclockwise direction about the origin of the circle. At another part of the course, the car enters a second circular turn at 180 km/h also in the counterclockwise direction. If the radius of curvature of the first turn is 130.0 m and that of the second is 100.0 m, compare the angular momenta of the race car in each turn taken about the origin of the circular turn.

Exercise:**Problem:**

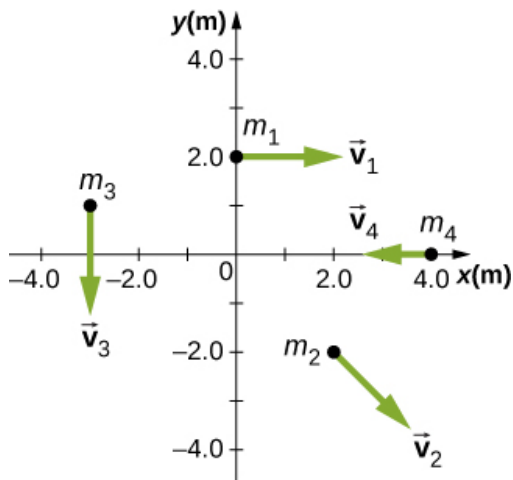
A particle of mass 5.0 kg has position vector $\vec{\mathbf{r}} = (2.0\hat{\mathbf{i}} - 3.0\hat{\mathbf{j}})\text{m}$ at a particular instant of time when its velocity is $\vec{\mathbf{v}} = (3.0\hat{\mathbf{i}})\text{m/s}$ with respect to the origin. (a) What is the angular momentum of the particle? (b) If a force $\vec{\mathbf{F}} = 5.0\hat{\mathbf{j}} \text{ N}$ acts on the particle at this instant, what is the torque about the origin?

Solution:

- a. $\vec{\mathbf{L}} = 45.0 \text{ kg} \cdot \text{m}^2/\text{s}\hat{\mathbf{k}};$
- b. $\vec{\boldsymbol{\tau}} = 10.0 \text{ N} \cdot \text{m}\hat{\mathbf{k}}$

Exercise:**Problem:**

Use the right-hand rule to determine the directions of the angular momenta about the origin of the particles as shown below. The z-axis is out of the page.



Exercise:

Problem:

Suppose the particles in the preceding problem have masses $m_1 = 0.10 \text{ kg}$, $m_2 = 0.20 \text{ kg}$, $m_3 = 0.30 \text{ kg}$, $m_4 = 0.40 \text{ kg}$. The velocities of the particles are $v_1 = 2.0\hat{i} \text{ m/s}$, $v_2 = (3.0\hat{i} - 3.0\hat{j}) \text{ m/s}$, $v_3 = -1.5\hat{j} \text{ m/s}$, $v_4 = -4.0\hat{i} \text{ m/s}$. (a) Calculate the angular momentum of each particle about the origin. (b) What is the total angular momentum of the four-particle system about the origin?

Solution:

a. $\vec{l}_1 = -0.4 \text{ kg} \cdot \text{m}^2/\text{s}\hat{k}$, $\vec{l}_2 = \vec{l}_4 = 0$,
 $\vec{l}_3 = 1.35 \text{ kg} \cdot \text{m}^2/\text{s}\hat{k}$; b. $\vec{L} = 0.95 \text{ kg} \cdot \text{m}^2/\text{s}\hat{k}$

Exercise:

Problem:

Two particles of equal mass travel with the same speed in opposite directions along parallel lines separated by a distance d . Show that the angular momentum of this two-particle system is the same no matter what point is used as the reference for calculating the angular momentum.

Exercise:

Problem:

An airplane of mass $4.0 \times 10^4 \text{ kg}$ flies horizontally at an altitude of 10 km with a constant speed of 250 m/s relative to Earth. (a) What is the magnitude of the airplane's angular momentum relative to a ground observer directly below the plane? (b) Does the angular momentum change as the airplane flies along a constant altitude?

Solution:

a. $L = 1.0 \times 10^{11} \text{ kg} \cdot \text{m}^2/\text{s}$; b. No, the angular momentum stays the same since the cross-product involves only the perpendicular distance from the plane to the ground no matter where it is along its path.

Exercise:

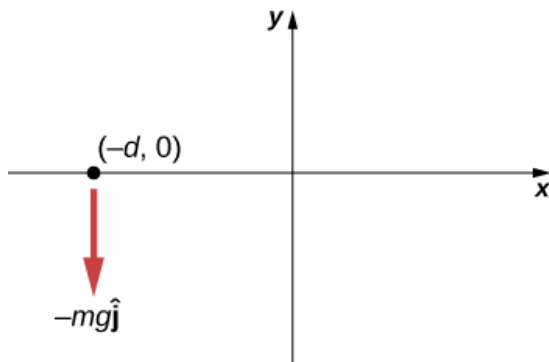
Problem:

At a particular instant, a 1.0-kg particle's position is $\vec{r} = (2.0\hat{i} - 4.0\hat{j} + 6.0\hat{k})\text{m}$, its velocity is $\vec{v} = (-1.0\hat{i} + 4.0\hat{j} + 1.0\hat{k})\text{m/s}$, and the force on it is $\vec{F} = (10.0\hat{i} + 15.0\hat{j})\text{N}$. (a) What is the angular momentum of the particle about the origin? (b) What is the torque on the particle about the origin? (c) What is the time rate of change of the particle's angular momentum at this instant?

Exercise:

Problem:

A particle of mass m is dropped at the point $(-d, 0)$ and falls vertically in Earth's gravitational field $-g\hat{j}$. (a) What is the expression for the angular momentum of the particle around the z-axis, which points directly out of the page as shown below? (b) Calculate the torque on the particle around the z-axis. (c) Is the torque equal to the time rate of change of the angular momentum?



Solution:

- a. $\vec{v} = -gt\hat{j}$, $\vec{r}_{\perp} = -d\hat{i}$, $\vec{l} = mdgt\hat{k}$;
 b. $\vec{F} = -mg\hat{j}$, $\sum \vec{\tau} = dmg\hat{k}$; c. yes

Exercise:

Problem:

(a) Calculate the angular momentum of Earth in its orbit around the Sun. (b) Compare this angular momentum with the angular momentum of Earth about its axis.

Exercise:

Problem:

A boulder of mass 20 kg and radius 20 cm rolls down a hill 15 m high from rest. What is its angular momentum when it is half way down the hill? (b) At the bottom?

Solution:

$$\text{a. } mgh = \frac{1}{2}m(r\omega)^2 + \frac{1}{2}\frac{2}{5}mr^2\omega^2;$$

$$\omega = 51.2 \text{ rad/s};$$

$$L = 16.4 \text{ kg} \cdot \text{m}^2/\text{s};$$

$$\text{b. } \omega = 72.5 \text{ rad/s};$$

$$L = 23.2 \text{ kg} \cdot \text{m}^2/\text{s}$$

Exercise:**Problem:**

A satellite is spinning at 6.0 rev/s. The satellite consists of a main body in the shape of a sphere of radius 2.0 m and mass 10,000 kg, and two antennas projecting out from the center of mass of the main body that can be approximated with rods of length 3.0 m each and mass 10 kg. The antenna's lie in the plane of rotation. What is the angular momentum of the satellite?

Exercise:**Problem:**

A propeller consists of two blades each 3.0 m in length and mass 120 kg each. The propeller can be approximated by a single rod rotating about its center of mass. The propeller starts from rest and rotates up to 1200 rpm in 30 seconds at a constant rate. (a) What is the angular momentum of the propeller at $t = 10 \text{ s}$; $t = 20 \text{ s}$? (b) What is the torque on the propeller?

Solution:

$$\text{a. } I = 720.0 \text{ kg} \cdot \text{m}^2; \alpha = 4.20 \text{ rad/s}^2;$$

$$\omega(10 \text{ s}) = 42.0 \text{ rad/s}; L = 3.02 \times 10^4 \text{ kg} \cdot \text{m}^2/\text{s};$$

$$\omega(20 \text{ s}) = 84.0 \text{ rad/s};$$

$$\text{b. } \tau = 3.03 \times 10^3 \text{ N} \cdot \text{m}$$

Exercise:**Problem:**

A pulsar is a rapidly rotating neutron star. The Crab nebula pulsar in the constellation Taurus has a period of $33.5 \times 10^{-3} \text{ s}$, radius 10.0 km, and mass $2.8 \times 10^{30} \text{ kg}$. The pulsar's rotational period will increase over time due to the release of electromagnetic radiation, which doesn't change its radius but reduces its rotational energy. (a) What is the angular momentum of the pulsar? (b) Suppose the angular velocity decreases at a rate of 10^{-14} rad/s^2 . What is the torque on the pulsar?

Exercise:

Problem:

The blades of a wind turbine are 30 m in length and rotate at a maximum rotation rate of 20 rev/min. (a) If the blades are 6000 kg each and the rotor assembly has three blades, calculate the angular momentum of the turbine at this rotation rate. (b) What is the torque required to rotate the blades up to the maximum rotation rate in 5 minutes?

Solution:

- a. $L = 1.131 \times 10^7 \text{ kg} \cdot \text{m}^2/\text{s}$;
b. $\tau = 3.77 \times 10^4 \text{ N} \cdot \text{m}$

Exercise:**Problem:**

A roller coaster has mass 3000.0 kg and needs to make it safely through a vertical circular loop of radius 50.0 m. What is the minimum angular momentum of the coaster at the bottom of the loop to make it safely through? Neglect friction on the track. Take the coaster to be a point particle.

Exercise:**Problem:**

A mountain biker takes a jump in a race and goes airborne. The mountain bike is travelling at 10.0 m/s before it goes airborne. If the mass of the front wheel on the bike is 750 g and has radius 35 cm, what is the angular momentum of the spinning wheel in the air the moment the bike leaves the ground?

Solution:

$$\omega = 28.6 \text{ rad/s} \Rightarrow L = 2.6 \text{ kg} \cdot \text{m}^2/\text{s}$$

Glossary

angular momentum

rotational analog of linear momentum, found by taking the product of moment of inertia and angular velocity

Conservation of Angular Momentum

By the end of this section, you will be able to:

- Apply conservation of angular momentum to determine the angular velocity of a rotating system in which the moment of inertia is changing
- Explain how the rotational kinetic energy changes when a system undergoes changes in both moment of inertia and angular velocity

So far, we have looked at the angular momentum of systems consisting of point particles and rigid bodies. We have also analyzed the torques involved, using the expression that relates the external net torque to the change in angular momentum, [\[link\]](#). Examples of systems that obey this equation include a freely spinning bicycle tire that slows over time due to torque arising from friction, or the slowing of Earth's rotation over millions of years due to frictional forces exerted on tidal deformations.

However, suppose there is no net external torque on the system, $\sum \vec{\tau} = 0$. In this case, [\[link\]](#) becomes the **law of conservation of angular momentum**.

Note:

Law of Conservation of Angular Momentum

The angular momentum of a system of particles around a point in a fixed inertial reference frame is conserved if there is no net external torque around that point:

Equation:

$$\frac{d\vec{L}}{dt} = 0$$

or

Equation:

$$\vec{L} = \vec{l}_1 + \vec{l}_2 + \cdots + \vec{l}_N = \text{constant}.$$

Note that the *total* angular momentum \vec{L} is conserved. Any of the individual angular momenta can change as long as their sum remains constant. This law is analogous to linear momentum being conserved when the external force on a system is zero.

As an example of conservation of angular momentum, [\[link\]](#) shows an ice skater executing a spin. The net torque on her is very close to zero because there is relatively little friction between her skates and the ice. Also, the friction is exerted very close to the pivot point. Both $|\vec{F}|$ and $|\vec{r}|$ are small, so $|\vec{\tau}|$ is negligible. Consequently, she can spin for quite some time. She can also increase her rate of spin by pulling her arms and legs in. Why does pulling her arms and legs in increase her rate of spin? The answer is that her angular momentum is constant, so that

Equation:

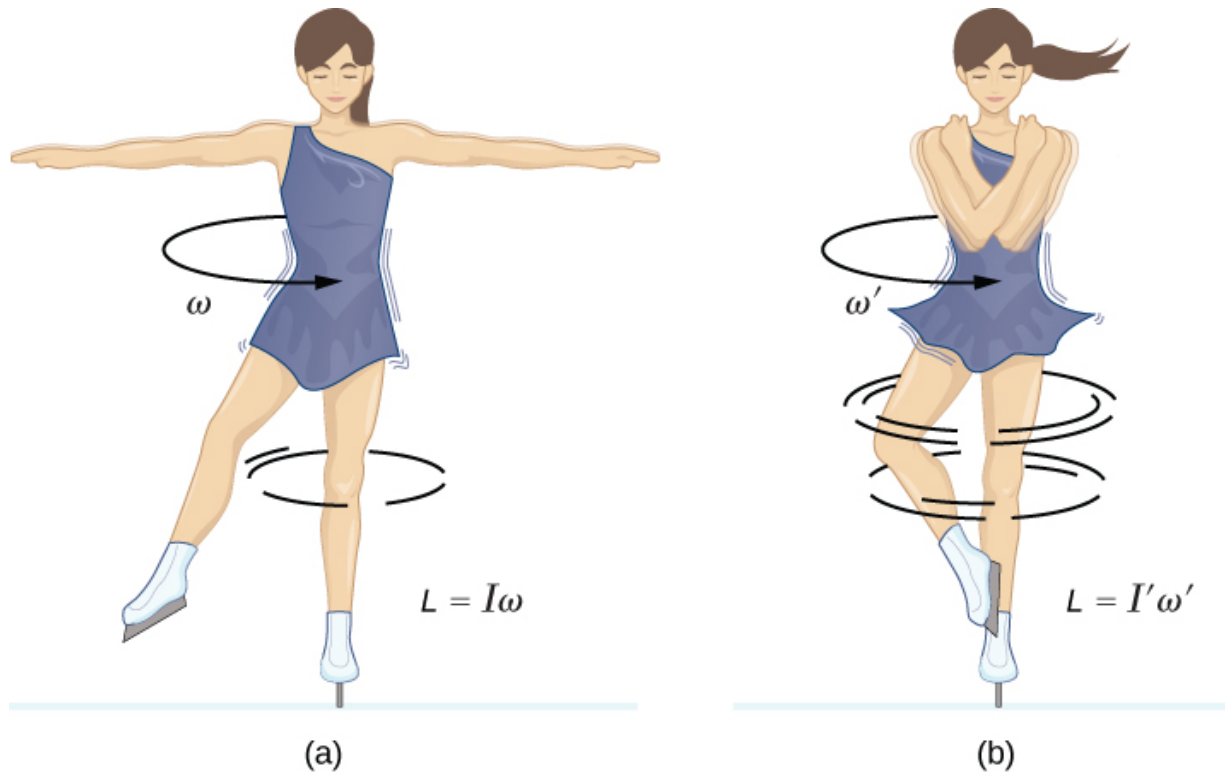
$$L' = L$$

or

Equation:

$$I'\omega' = I\omega,$$

where the primed quantities refer to conditions after she has pulled in her arms and reduced her moment of inertia. Because I' is smaller, the angular velocity ω' must increase to keep the angular momentum constant.



(a) An ice skater is spinning on the tip of her skate with her arms extended. Her angular momentum is conserved because the net torque on her is negligibly small. (b) Her rate of spin increases greatly when she pulls in her arms, decreasing her moment of inertia. The work she does to pull in her arms results in an increase in rotational kinetic energy.

It is interesting to see how the rotational kinetic energy of the skater changes when she pulls her arms in. Her initial rotational energy is

Equation:

$$K_{\text{Rot}} = \frac{1}{2}I\omega^2,$$

whereas her final rotational energy is

Equation:

$$K'_{\text{Rot}} = \frac{1}{2}I'(\omega')^2.$$

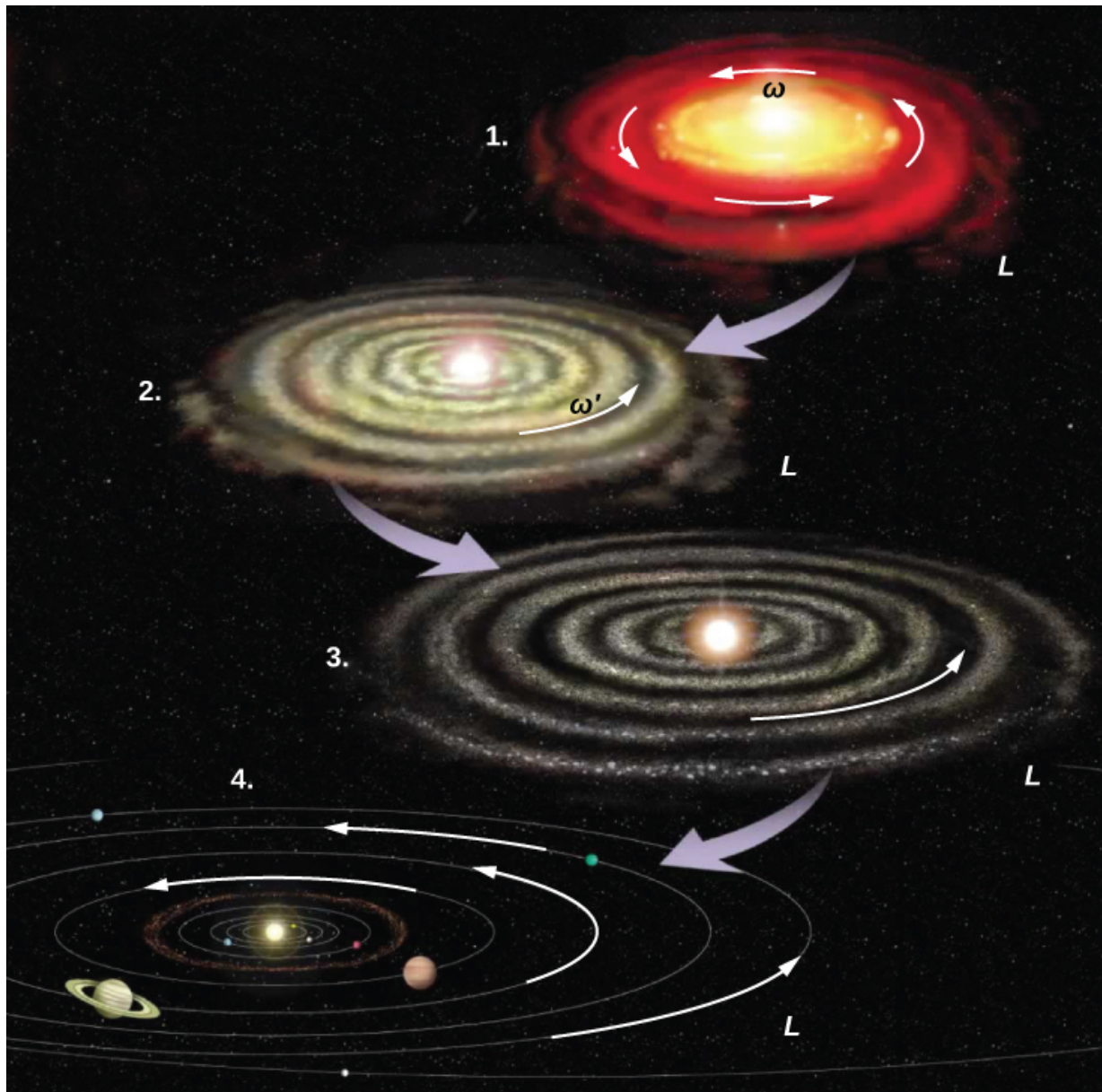
Since $I'\omega' = I\omega$, we can substitute for ω' and find

Equation:

$$K'_{\text{Rot}} = \frac{1}{2}I'(\omega')^2 = \frac{1}{2}I'\left(\frac{I}{I'}\omega\right)^2 = \frac{1}{2}I\omega^2\left(\frac{I}{I'}\right) = K_{\text{Rot}}\left(\frac{I}{I'}\right).$$

Because her moment of inertia has decreased, $I' < I$, her final rotational kinetic energy has increased. The source of this additional rotational kinetic energy is the work required to pull her arms inward. Note that the skater's arms do not move in a perfect circle—they spiral inward. This work causes an increase in the rotational kinetic energy, while her angular momentum remains constant. Since she is in a frictionless environment, no energy escapes the system. Thus, if she were to extend her arms to their original positions, she would rotate at her original angular velocity and her kinetic energy would return to its original value.

The solar system is another example of how conservation of angular momentum works in our universe. Our solar system was born from a huge cloud of gas and dust that initially had rotational energy. Gravitational forces caused the cloud to contract, and the rotation rate increased as a result of conservation of angular momentum ([\[link\]](#)).

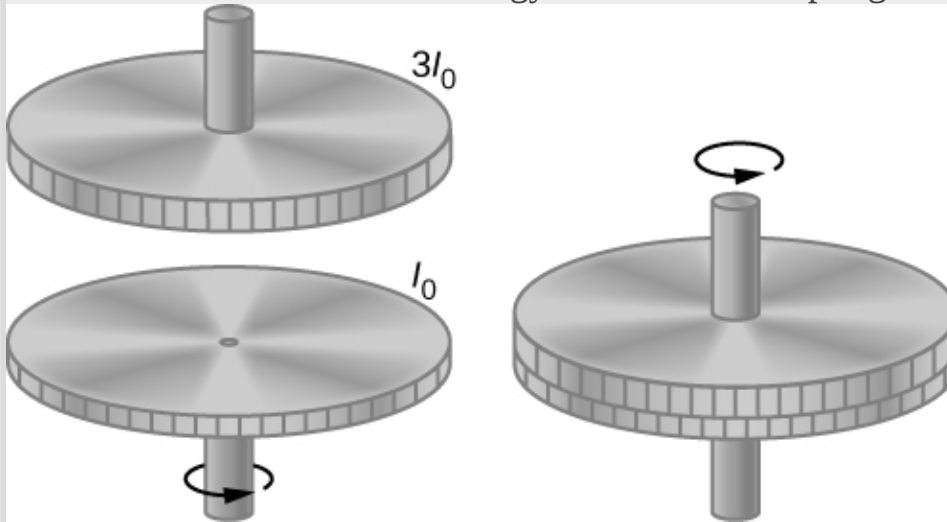


The solar system coalesced from a cloud of gas and dust that was originally rotating. The orbital motions and spins of the planets are in the same direction as the original spin and conserve the angular momentum of the parent cloud. (credit: modification of work by NASA)

We continue our discussion with an example that has applications to engineering.

Example:**Coupled Flywheels**

A flywheel rotates without friction at an angular velocity $\omega_0 = 600 \text{ rev/min}$ on a frictionless, vertical shaft of negligible rotational inertia. A second flywheel, which is at rest and has a moment of inertia three times that of the rotating flywheel, is dropped onto it ([link](#)). Because friction exists between the surfaces, the flywheels very quickly reach the same rotational velocity, after which they spin together. (a) Use the law of conservation of angular momentum to determine the angular velocity ω of the combination. (b) What fraction of the initial kinetic energy is lost in the coupling of the flywheels?



Two flywheels are coupled and rotate together.

Strategy

Part (a) is straightforward to solve for the angular velocity of the coupled system. We use the result of (a) to compare the initial and final kinetic energies of the system in part (b).

Solution

a. No external torques act on the system. The force due to friction produces an internal torque, which does not affect the angular momentum of the system. Therefore conservation of angular momentum gives

Equation:

$$I_0\omega_0 = (I_0 + 3I_0)\omega,$$
$$\omega = \frac{1}{4}\omega_0 = 150 \text{ rev/min} = 15.7 \text{ rad/s}.$$

b. Before contact, only one flywheel is rotating. The rotational kinetic energy of this flywheel is the initial rotational kinetic energy of the system, $\frac{1}{2} I_0 \omega_0^2$.

The final kinetic energy is $\frac{1}{2} (4I_0) \omega^2 = \frac{1}{2} (4I_0) \left(\frac{\omega_0}{4} \right)^2 = \frac{1}{8} I_0 \omega_0^2$.

Therefore, the ratio of the final kinetic energy to the initial kinetic energy is

Equation:

$$\frac{\frac{1}{8} I_0 \omega_0^2}{\frac{1}{2} I_0 \omega_0^2} = \frac{1}{4}.$$

Thus, 3/4 of the initial kinetic energy is lost to the coupling of the two flywheels.

Significance

Since the rotational inertia of the system increased, the angular velocity decreased, as expected from the law of conservation of angular momentum. In this example, we see that the final kinetic energy of the system has decreased, as energy is lost to the coupling of the flywheels. Compare this to the example of the skater in [\[link\]](#) doing work to bring her arms inward and adding rotational kinetic energy.

Note:

Exercise:

Problem:

Check Your Understanding A merry-go-round at a playground is rotating at 4.0 rev/min. Three children jump on and increase the moment of inertia of the merry-go-round/children rotating system by 25 %. What is the new rotation rate?

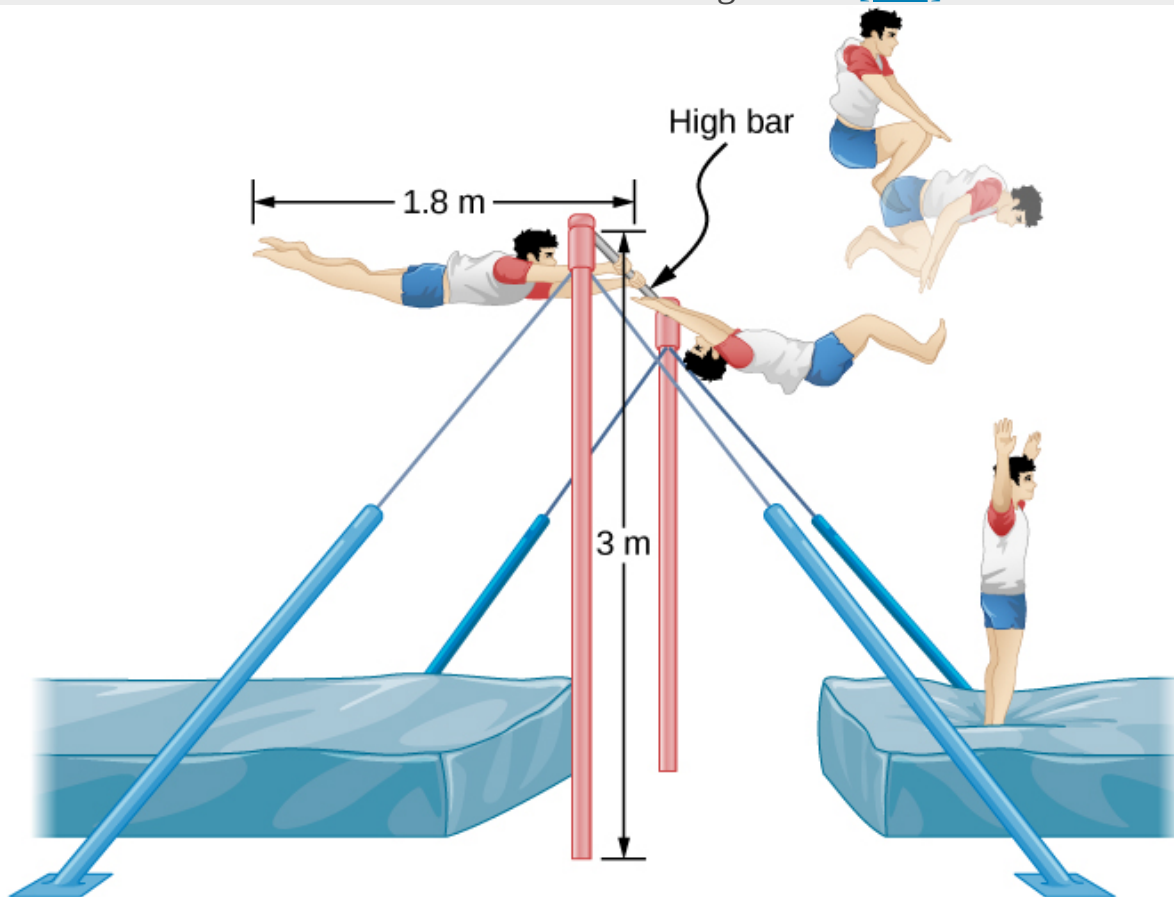
Solution:

Using conservation of angular momentum, we have

$$I(4.0 \text{ rev/min}) = 1.25I\omega_f, \quad \omega_f = \frac{1.0}{1.25}(4.0 \text{ rev/min}) = 3.2 \text{ rev/min}$$

Example:**Dismount from a High Bar**

An 80.0-kg gymnast dismounts from a high bar. He starts the dismount at full extension, then tucks to complete a number of revolutions before landing. His moment of inertia when fully extended can be approximated as a rod of length 1.8 m and when in the tuck a rod of half that length. If his rotation rate at full extension is 1.0 rev/s and he enters the tuck when his center of mass is at 3.0 m height moving horizontally to the floor, how many revolutions can he execute if he comes out of the tuck at 1.8 m height? See [\[link\]](#).



A gymnast dismounts from a high bar and executes a number of revolutions in the tucked position before landing upright.

Strategy

Using conservation of angular momentum, we can find his rotation rate when in the tuck. Using the equations of kinematics, we can find the time interval from a height of 3.0 m to 1.8 m. Since he is moving horizontally with respect

to the ground, the equations of free fall simplify. This will allow the number of revolutions that can be executed to be calculated. Since we are using a ratio, we can keep the units as rev/s and don't need to convert to radians/s.

Solution

The moment of inertia at full extension is

$$I_0 = \frac{1}{12}mL^2 = \frac{1}{12}80.0 \text{ kg}(1.8 \text{ m})^2 = 21.6 \text{ kg} \cdot \text{m}^2.$$

The moment of inertia in the tuck is

$$I_f = \frac{1}{12}mL_f^2 = \frac{1}{12}80.0 \text{ kg}(0.9 \text{ m})^2 = 5.4 \text{ kg} \cdot \text{m}^2.$$

Conservation of angular momentum:

$$I_f\omega_f = I_0\omega_0 \Rightarrow \omega_f = \frac{I_0\omega_0}{I_f} = \frac{21.6 \text{ kg} \cdot \text{m}^2(1.0 \text{ rev/s})}{5.4 \text{ kg} \cdot \text{m}^2} = 4.0 \text{ rev/s}.$$

$$\text{Time interval in the tuck: } t = \sqrt{\frac{2h}{g}} = \sqrt{\frac{2(3.0-1.8)\text{m}}{9.8 \text{ m/s}^2}} = 0.5 \text{ s}.$$

In 0.5 s, he will be able to execute two revolutions at 4.0 rev/s.

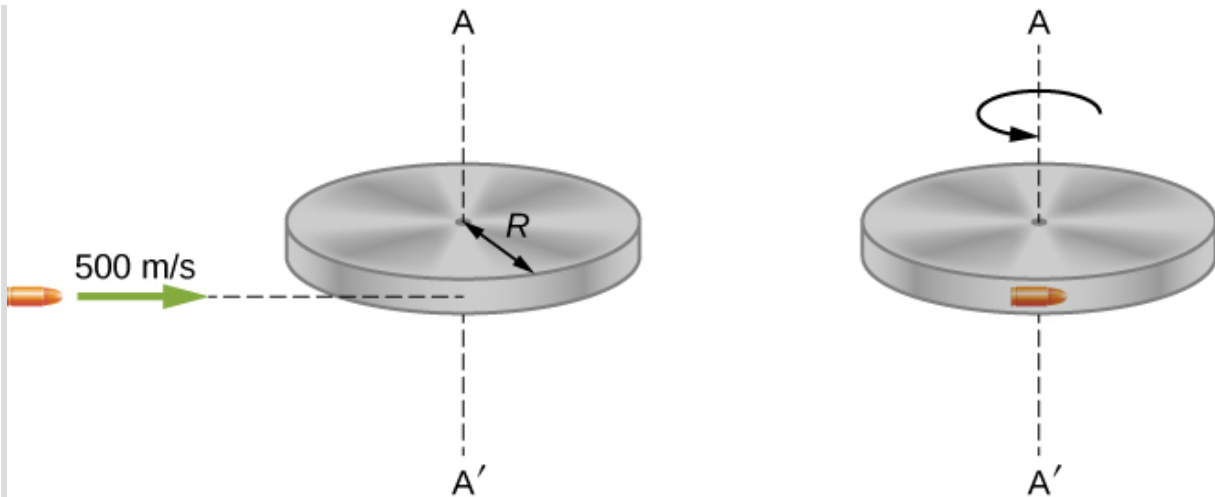
Significance

Note that the number of revolutions he can complete will depend on how long he is in the air. In the problem, he is exiting the high bar horizontally to the ground. He could also exit at an angle with respect to the ground, giving him more or less time in the air depending on the angle, positive or negative, with respect to the ground. Gymnasts must take this into account when they are executing their dismounts.

Example:

Conservation of Angular Momentum of a Collision

A bullet of mass $m = 2.0 \text{ g}$ is moving horizontally with a speed of 500.0 m/s . The bullet strikes and becomes embedded in the edge of a solid disk of mass $M = 3.2 \text{ kg}$ and radius $R = 0.5 \text{ m}$. The cylinder is free to rotate around its axis and is initially at rest ([\[link\]](#)). What is the angular velocity of the disk immediately after the bullet is embedded?



A bullet is fired horizontally and becomes embedded in the edge of a disk that is free to rotate about its vertical axis.

Strategy

For the system of the bullet and the cylinder, no external torque acts along the vertical axis through the center of the disk. Thus, the angular momentum along this axis is conserved. The initial angular momentum of the bullet is mvR , which is taken about the rotational axis of the disk the moment before the collision. The initial angular momentum of the cylinder is zero. Thus, the net angular momentum of the system is mvR . Since angular momentum is conserved, the initial angular momentum of the system is equal to the angular momentum of the bullet embedded in the disk immediately after impact.

Solution

The initial angular momentum of the system is

Equation:

$$L_i = mvR.$$

The moment of inertia of the system with the bullet embedded in the disk is

Equation:

$$I = mR^2 + \frac{1}{2}MR^2 = \left(m + \frac{M}{2}\right)R^2.$$

The final angular momentum of the system is

Equation:

$$L_f = I\omega_f.$$

Thus, by conservation of angular momentum, $L_i = L_f$ and

Equation:

$$mvR = \left(m + \frac{M}{2}\right)R^2\omega_f.$$

Solving for ω_f ,

Equation:

$$\omega_f = \frac{mvR}{(m + M/2)R^2} = \frac{(2.0 \times 10^{-3} \text{ kg})(500.0 \text{ m/s})}{(2.0 \times 10^{-3} \text{ kg} + 1.6 \text{ kg})(0.50 \text{ m})} = 1.2 \text{ rad/s}.$$

Significance

The system is composed of both a point particle and a rigid body. Care must be taken when formulating the angular momentum before and after the collision. Just before impact the angular momentum of the bullet is taken about the rotational axis of the disk.

Summary

- In the absence of external torques, a system's total angular momentum is conserved. This is the rotational counterpart to linear momentum being conserved when the external force on a system is zero.
- For a rigid body that changes its angular momentum in the absence of a net external torque, conservation of angular momentum gives $I_f\omega_f = I_i\omega_i$. This equation says that the angular velocity is inversely proportional to the moment of inertia. Thus, if the moment of inertia decreases, the angular velocity must increase to conserve angular momentum.
- Systems containing both point particles and rigid bodies can be analyzed using conservation of angular momentum. The angular momentum of all bodies in the system must be taken about a common axis.

Conceptual Questions

Exercise:**Problem:**

What is the purpose of the small propeller at the back of a helicopter that rotates in the plane perpendicular to the large propeller?

Solution:

Without the small propeller, the body of the helicopter would rotate in the opposite sense to the large propeller in order to conserve angular momentum. The small propeller exerts a thrust at a distance R from the center of mass of the aircraft to prevent this from happening.

Exercise:**Problem:**

Suppose a child walks from the outer edge of a rotating merry-go-round to the inside. Does the angular velocity of the merry-go-round increase, decrease, or remain the same? Explain your answer. Assume the merry-go-round is spinning without friction.

Exercise:**Problem:**

As the rope of a tethered ball winds around a pole, what happens to the angular velocity of the ball?

Solution:

The angular velocity increases because the moment of inertia is decreasing.

Exercise:**Problem:**

Suppose the polar ice sheets broke free and floated toward Earth's equator without melting. What would happen to Earth's angular velocity?

Exercise:

Problem: Explain why stars spin faster when they collapse.

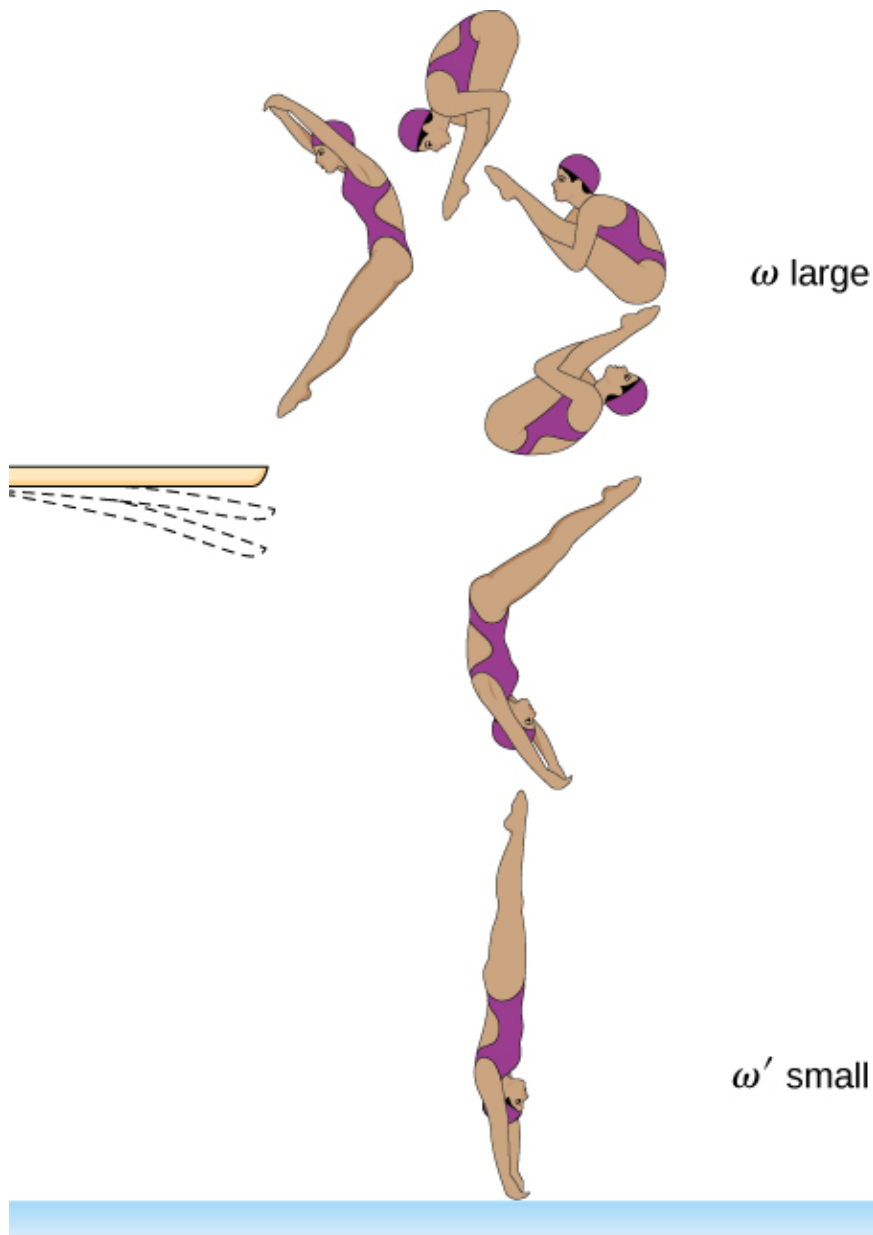
Solution:

More mass is concentrated near the rotational axis, which decreases the moment of inertia causing the star to increase its angular velocity.

Exercise:

Problem:

Competitive divers pull their limbs in and curl up their bodies when they do flips. Just before entering the water, they fully extend their limbs to enter straight down (see below). Explain the effect of both actions on their angular velocities. Also explain the effect on their angular momentum.



Problems

Exercise:

Problem:

A disk of mass 2.0 kg and radius 60 cm with a small mass of 0.05 kg attached at the edge is rotating at 2.0 rev/s. The small mass, while attached to the disk, slides gradually to the center of the disk. What is the disk's final rotation rate?

Exercise:**Problem:**

The Sun's mass is 2.0×10^{30} kg, its radius is 7.0×10^5 km, and it has a rotational period of approximately 28 days. If the Sun should collapse into a white dwarf of radius 3.5×10^3 km, what would its period be if no mass were ejected and a sphere of uniform density can model the Sun both before and after?

Solution:

$$\begin{aligned} L_f &= \frac{2}{5} M_S (3.5 \times 10^3 \text{ km})^2 \frac{2\pi}{T_f}, \\ (7.0 \times 10^5 \text{ km})^2 \frac{2\pi}{28 \text{ days}} &= (3.5 \times 10^3 \text{ km})^2 \frac{2\pi}{T_f} \Rightarrow T_f \\ &= 28 \text{ days} \frac{(3.5 \times 10^3 \text{ km})^2}{(7.0 \times 10^5 \text{ km})^2} = 7.0 \times 10^{-4} \text{ day} = 60.5 \text{ s} \end{aligned}$$

Exercise:**Problem:**

A cylinder with rotational inertia $I_1 = 2.0 \text{ kg} \cdot \text{m}^2$ rotates clockwise about a vertical axis through its center with angular speed $\omega_1 = 5.0 \text{ rad/s}$. A second cylinder with rotational inertia $I_2 = 1.0 \text{ kg} \cdot \text{m}^2$ rotates counterclockwise about the same axis with angular speed $\omega_2 = 8.0 \text{ rad/s}$. If the cylinders couple so they have the same rotational axis what is the angular speed of the combination? What percentage of the original kinetic energy is lost to friction?

Exercise:

Problem:

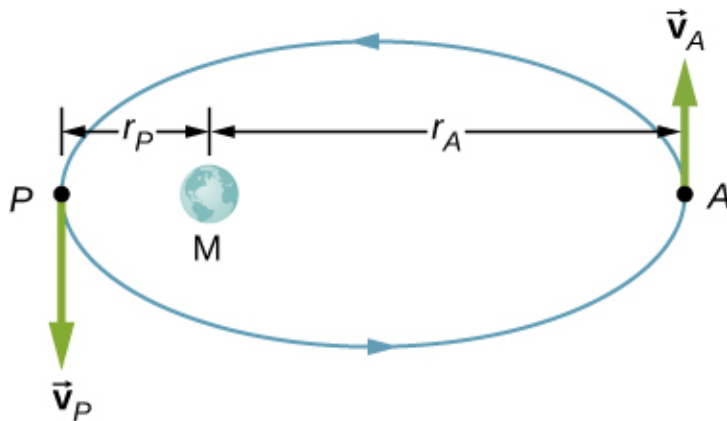
A diver off the high board imparts an initial rotation with his body fully extended before going into a tuck and executing three back somersaults before hitting the water. If his moment of inertia before the tuck is $16.9 \text{ kg} \cdot \text{m}^2$ and after the tuck during the somersaults is $4.2 \text{ kg} \cdot \text{m}^2$, what rotation rate must he impart to his body directly off the board and before the tuck if he takes 1.4 s to execute the somersaults before hitting the water?

Solution:

$$f_f = 2.1 \text{ rev/s} \Rightarrow f_0 = 0.5 \text{ rev/s}$$

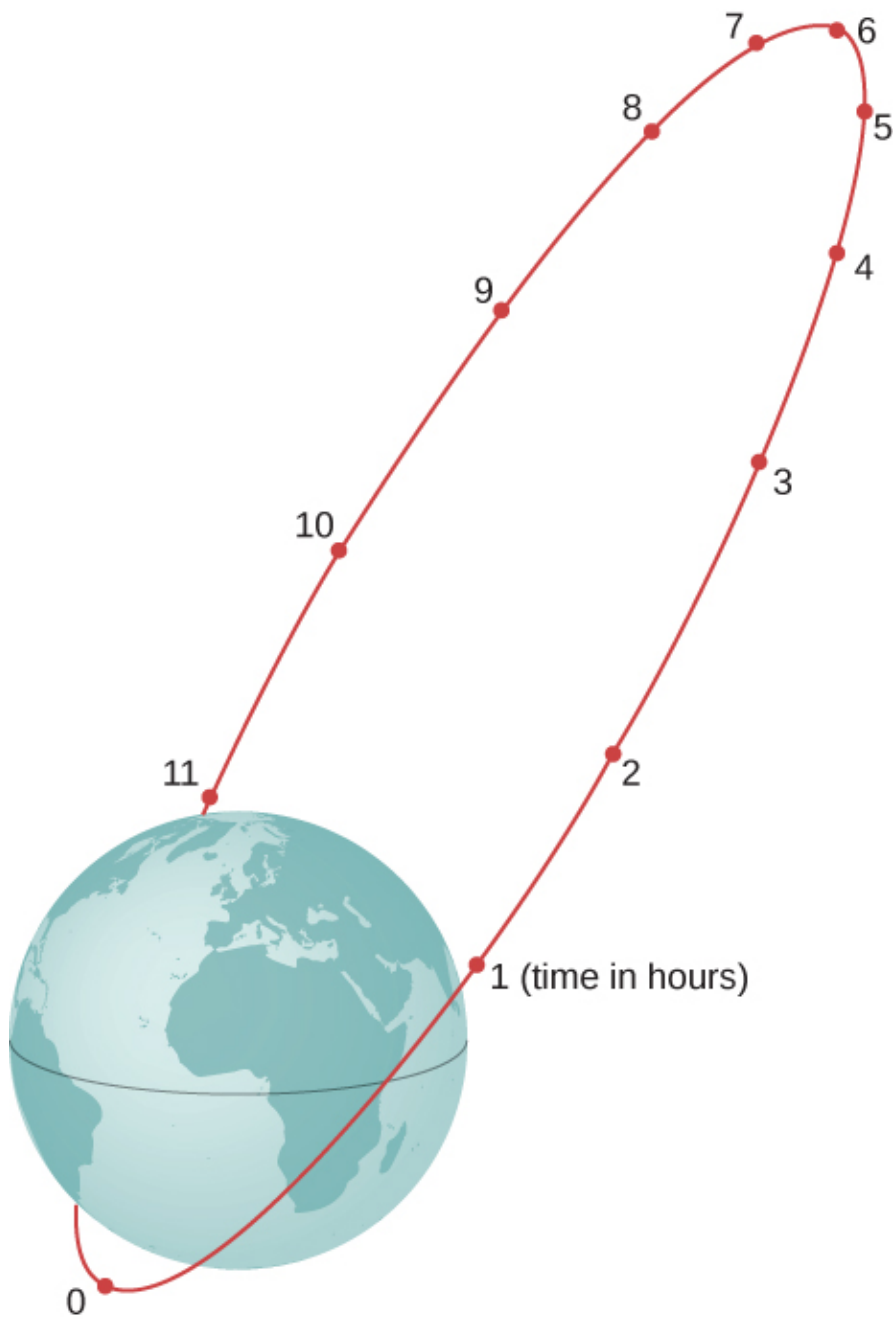
Exercise:**Problem:**

An Earth satellite has its apogee at 2500 km above the surface of Earth and perigee at 500 km above the surface of Earth. At apogee its speed is 6260 m/s. What is its speed at perigee? Earth's radius is 6370 km (see below).

**Exercise:**

Problem:

A Molniya orbit is a highly eccentric orbit of a communication satellite so as to provide continuous communications coverage for Scandinavian countries and adjacent Russia. The orbit is positioned so that these countries have the satellite in view for extended periods in time (see below). If a satellite in such an orbit has an apogee at 40,000.0 km as measured from the center of Earth and a velocity of 3.0 km/s, what would be its velocity at perigee measured at 200.0 km altitude?



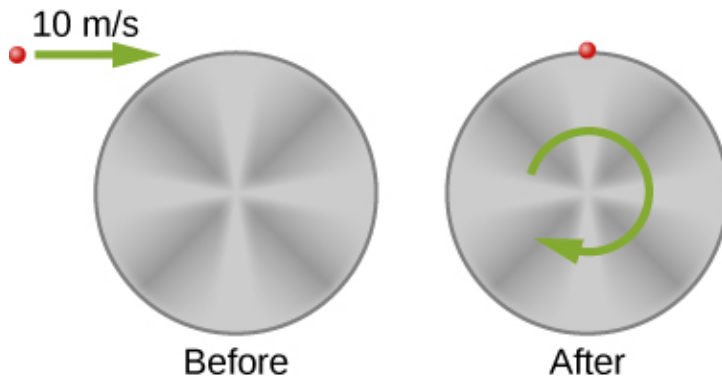
Solution:

$$r_P m v_P = r_A m v_A \Rightarrow v_P = 18.3 \text{ km/s}$$

Exercise:

Problem:

Shown below is a small particle of mass 20 g that is moving at a speed of 10.0 m/s when it collides and sticks to the edge of a uniform solid cylinder. The cylinder is free to rotate about its axis through its center and is perpendicular to the page. The cylinder has a mass of 0.5 kg and a radius of 10 cm, and is initially at rest. (a) What is the angular velocity of the system after the collision? (b) How much kinetic energy is lost in the collision?

**Exercise:****Problem:**

A bug of mass 0.020 kg is at rest on the edge of a solid cylindrical disk ($M = 0.10$ kg, $R = 0.10$ m) rotating in a horizontal plane around the vertical axis through its center. The disk is rotating at 10.0 rad/s. The bug crawls to the center of the disk. (a) What is the new angular velocity of the disk? (b) What is the change in the kinetic energy of the system? (c) If the bug crawls back to the outer edge of the disk, what is the angular velocity of the disk then? (d) What is the new kinetic energy of the system? (e) What is the cause of the increase and decrease of kinetic energy?

Solution:

$$\begin{aligned} \text{a. } I_{\text{disk}} &= 5.0 \times 10^{-4} \text{ kg} \cdot \text{m}^2, \\ I_{\text{bug}} &= 2.0 \times 10^{-4} \text{ kg} \cdot \text{m}^2, \end{aligned}$$

$$(I_{\text{disk}} + I_{\text{bug}})\omega_1 = I_{\text{disk}}\omega_2,$$

$$\omega_2 = 14.0 \text{ rad/s}$$

$$\text{b. } \Delta K = 0.014 \text{ J;}$$

$$\text{c. } \omega_3 = 10.0 \text{ rad/s back to the original value;}$$

$$\text{d. } \frac{1}{2}(I_{\text{disk}} + I_{\text{bug}})\omega_3^2 = 0.035 \text{ J back to the original value;}$$

$$\text{e. work of the bug crawling on the disk}$$

Exercise:

Problem:

A uniform rod of mass 200 g and length 100 cm is free to rotate in a horizontal plane around a fixed vertical axis through its center, perpendicular to its length. Two small beads, each of mass 20 g, are mounted in grooves along the rod. Initially, the two beads are held by catches on opposite sides of the rod's center, 10 cm from the axis of rotation. With the beads in this position, the rod is rotating with an angular velocity of 10.0 rad/s. When the catches are released, the beads slide outward along the rod. (a) What is the rod's angular velocity when the beads reach the ends of the rod? (b) What is the rod's angular velocity if the beads fly off the rod?

Exercise:

Problem:

A merry-go-round has a radius of 2.0 m and a moment of inertia $300 \text{ kg} \cdot \text{m}^2$. A boy of mass 50 kg runs tangent to the rim at a speed of 4.0 m/s and jumps on. If the merry-go-round is initially at rest, what is the angular velocity after the boy jumps on?

Solution:

$$L_i = 400.0 \text{ kg} \cdot \text{m}^2/\text{s},$$

$$L_f = 500.0 \text{ kg} \cdot \text{m}^2\omega,$$

$$\omega = 0.80 \text{ rad/s}$$

Exercise:

Problem:

A playground merry-go-round has a mass of 120 kg and a radius of 1.80 m and it is rotating with an angular velocity of 0.500 rev/s. What is its angular velocity after a 22.0-kg child gets onto it by grabbing its outer edge? The child is initially at rest.

Exercise:**Problem:**

Three children are riding on the edge of a merry-go-round that is 100 kg, has a 1.60-m radius, and is spinning at 20.0 rpm. The children have masses of 22.0, 28.0, and 33.0 kg. If the child who has a mass of 28.0 kg moves to the center of the merry-go-round, what is the new angular velocity in rpm?

Solution:

$$I_0 = 340.48 \text{ kg} \cdot \text{m}^2,$$

$$I_f = 268.8 \text{ kg} \cdot \text{m}^2,$$

$$\omega_f = 25.33 \text{ rpm}$$

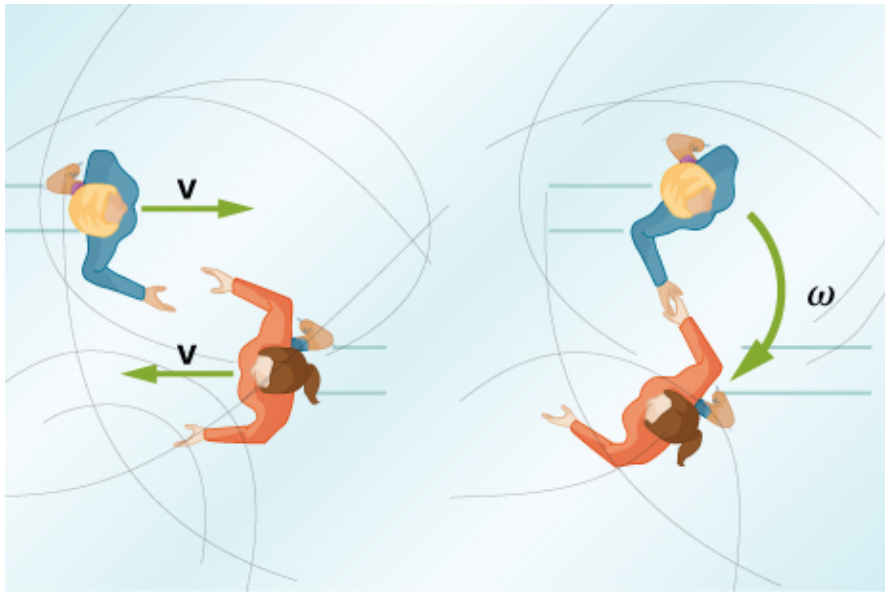
Exercise:**Problem:**

(a) Calculate the angular momentum of an ice skater spinning at 6.00 rev/s given his moment of inertia is $0.400 \text{ kg} \cdot \text{m}^2$. (b) He reduces his rate of spin (his angular velocity) by extending his arms and increasing his moment of inertia. Find the value of his moment of inertia if his angular velocity decreases to 1.25 rev/s. (c) Suppose instead he keeps his arms in and allows friction of the ice to slow him to 3.00 rev/s. What average torque was exerted if this takes 15.0 s?

Exercise:

Problem:

Twin skaters approach one another as shown below and lock hands. (a) Calculate their final angular velocity, given each had an initial speed of 2.50 m/s relative to the ice. Each has a mass of 70.0 kg, and each has a center of mass located 0.800 m from their locked hands. You may approximate their moments of inertia to be that of point masses at this radius. (b) Compare the initial kinetic energy and final kinetic energy.



(a)

(b)

Solution:

a. $L = 280 \text{ kg} \cdot \text{m}^2/\text{s}$,

$I_f = 89.6 \text{ kg} \cdot \text{m}^2$,

$\omega_f = 3.125 \text{ rad/s}$; b. $K_i = 437.5 \text{ J}$,

$K_f = 437.5 \text{ J}$

Exercise:

Problem:

A baseball catcher extends his arm straight up to catch a fast ball with a speed of 40 m/s. The baseball is 0.145 kg and the catcher's arm length is 0.5 m and mass 4.0 kg. (a) What is the angular velocity of the arm immediately after catching the ball as measured from the arm socket? (b) What is the torque applied if the catcher stops the rotation of his arm 0.3 s after catching the ball?

Exercise:**Problem:**

In 2015, in Warsaw, Poland, Olivia Oliver of Nova Scotia broke the world record for being the fastest spinner on ice skates. She achieved a record 342 rev/min, beating the existing Guinness World Record by 34 rotations. If an ice skater extends her arms at that rotation rate, what would be her new rotation rate? Assume she can be approximated by a 45-kg rod that is 1.7 m tall with a radius of 15 cm in the record spin. With her arms stretched take the approximation of a rod of length 130 cm with 10% of her body mass aligned perpendicular to the spin axis. Neglect frictional forces.

Solution:

Moment of inertia in the record spin: $I_0 = 0.5 \text{ kg} \cdot \text{m}^2$,

$I_f = 1.1 \text{ kg} \cdot \text{m}^2$,

$$\omega_f = \frac{I_0}{I_f} \omega_0 \Rightarrow f_f = 155.5 \text{ rev/min}$$

Exercise:**Problem:**

A satellite in a geosynchronous circular orbit is 42,164.0 km from the center of Earth. A small asteroid collides with the satellite sending it into an elliptical orbit of apogee 45,000.0 km. What is the speed of the satellite at apogee? Assume its angular momentum is conserved.

Exercise:

Problem:

A gymnast does cartwheels along the floor and then launches herself into the air and executes several flips in a tuck while she is airborne. If her moment of inertia when executing the cartwheels is $13.5 \text{ kg} \cdot \text{m}^2$ and her spin rate is 0.5 rev/s , how many revolutions does she do in the air if her moment of inertia in the tuck is $3.4 \text{ kg} \cdot \text{m}^2$ and she has 2.0 s to do the flips in the air?

Solution:

Her spin rate in the air is: $f_f = 2.0 \text{ rev/s}$;
She can do four flips in the air.

Exercise:**Problem:**

The centrifuge at NASA Ames Research Center has a radius of 8.8 m and can produce forces on its payload of 20 gs or 20 times the force of gravity on Earth. (a) What is the angular momentum of a 20-kg payload that experiences 10 gs in the centrifuge? (b) If the driver motor was turned off in (a) and the payload lost 10 kg , what would be its new spin rate, taking into account there are no frictional forces present?

Exercise:**Problem:**

A ride at a carnival has four spokes to which pods are attached that can hold two people. The spokes are each 15 m long and are attached to a central axis. Each spoke has mass 200.0 kg , and the pods each have mass 100.0 kg . If the ride spins at 0.2 rev/s with each pod containing two 50.0-kg children, what is the new spin rate if all the children jump off the ride?

Solution:

Moment of inertia with all children aboard:

$$I_0 = 2.4 \times 10^5 \text{ kg} \cdot \text{m}^2;$$

$$I_f = 1.5 \times 10^5 \text{ kg} \cdot \text{m}^2;$$

$$f_f = 0.3 \text{ rev/s}$$

Exercise:**Problem:**

An ice skater is preparing for a jump with turns and has his arms extended. His moment of inertia is $1.8 \text{ kg} \cdot \text{m}^2$ while his arms are extended, and he is spinning at 0.5 rev/s . If he launches himself into the air at 9.0 m/s at an angle of 45° with respect to the ice, how many revolutions can he execute while airborne if his moment of inertia in the air is $0.5 \text{ kg} \cdot \text{m}^2$?

Exercise:**Problem:**

A space station consists of a giant rotating hollow cylinder of mass 10^6 kg including people on the station and a radius of 100.00 m . It is rotating in space at 3.30 rev/min in order to produce artificial gravity. If 100 people of an average mass of 65.00 kg spacewalk to an awaiting spaceship, what is the new rotation rate when all the people are off the station?

Solution:

$$\begin{aligned}I_0 &= 1.00 \times 10^{10} \text{ kg} \cdot \text{m}^2, \\I_f &= 9.94 \times 10^9 \text{ kg} \cdot \text{m}^2, \\f_f &= 3.32 \text{ rev/min}\end{aligned}$$

Exercise:**Problem:**

Neptune has a mass of $1.0 \times 10^{26} \text{ kg}$ and is $4.5 \times 10^9 \text{ km}$ from the Sun with an orbital period of 165 years. Planetesimals in the outer primordial solar system 4.5 billion years ago coalesced into Neptune over hundreds of millions of years. If the primordial disk that evolved into our present day solar system had a radius of 10^{11} km and if the matter that made up these planetesimals that later became Neptune was spread out evenly on the edges of it, what was the orbital period of the outer edges of the primordial disk?

Glossary

law of conservation of angular momentum

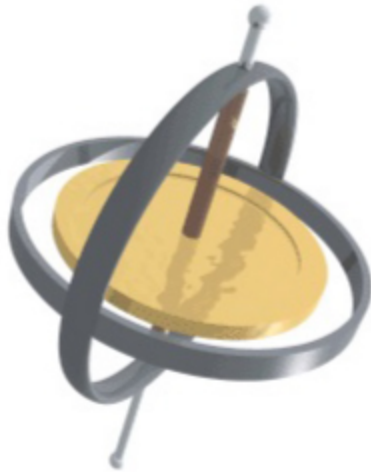
angular momentum is conserved, that is, the initial angular momentum is equal to the final angular momentum when no external torque is applied to the system

Precession of a Gyroscope

By the end of this section, you will be able to:

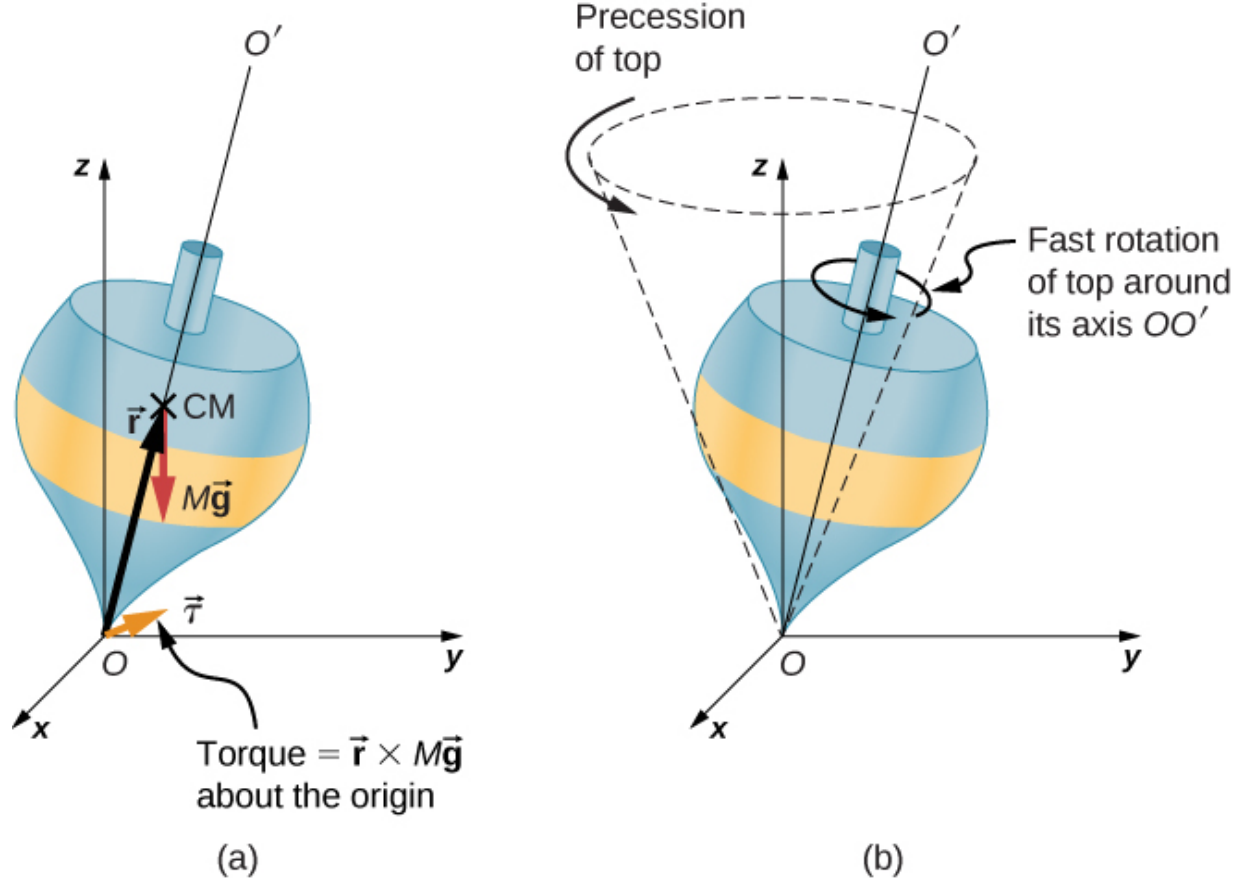
- Describe the physical processes underlying the phenomenon of precession
- Calculate the precessional angular velocity of a gyroscope

[\[link\]](#) shows a gyroscope, defined as a spinning disk in which the axis of rotation is free to assume any orientation. When spinning, the orientation of the spin axis is unaffected by the orientation of the body that encloses it. The body or vehicle enclosing the gyroscope can be moved from place to place and the orientation of the spin axis will remain the same. This makes gyroscopes very useful in navigation, especially where magnetic compasses can't be used, such as in piloted and unpiloted spacecrafts, intercontinental ballistic missiles, unmanned aerial vehicles, and satellites like the Hubble Space Telescope.



A gyroscope consists of a spinning disk about an axis that is free to assume any orientation.

We illustrate the **precession** of a gyroscope with an example of a top in the next two figures. If the top is placed on a flat surface near the surface of Earth at an angle to the vertical and is not spinning, it will fall over, due to the force of gravity producing a torque acting on its center of mass. This is shown in [\[link\]](#)(a). However, if the top is spinning on its axis, rather than topple over due to this torque, it precesses about the vertical, shown in part (b) of the figure. This is due to the torque on the center of mass, which provides the change in angular momentum.



(a) If the top is not spinning, there is a torque $\vec{r} \times M\vec{g}$ about the origin, and the top falls over. (b) If the top is spinning about its axis OO' , it doesn't fall over but precesses about the z -axis.

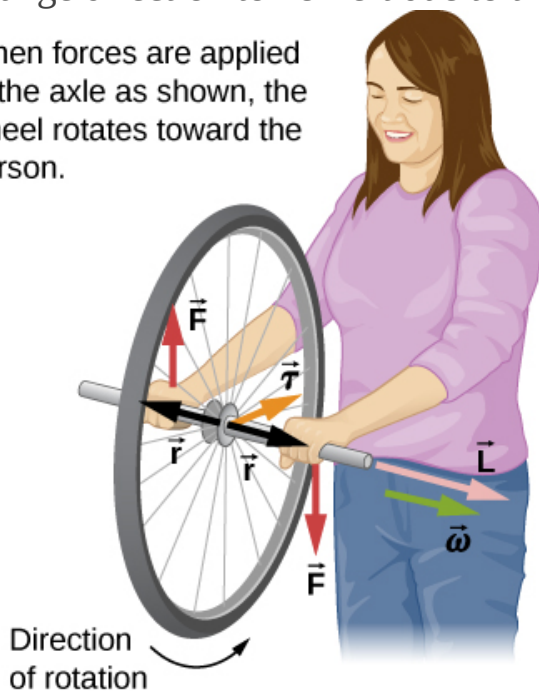
[\[link\]](#) shows the forces acting on a spinning top. The torque produced is perpendicular to the angular momentum vector. This changes the direction

of the angular momentum vector \vec{L} according to $d\vec{L} = \vec{\tau}dt$, but not its magnitude. The top *precesses* around a vertical axis, since the torque is always horizontal and perpendicular to \vec{L} . If the top is *not* spinning, it acquires angular momentum in the direction of the torque, and it rotates around a horizontal axis, falling over just as we would expect.
 [missing_resource: CNX_UPhysics_11_04_Precess.jpg]

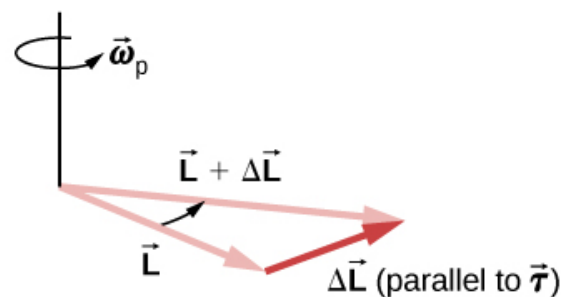
The force of gravity acting on the center of mass produces a torque $\vec{\tau}$ in the direction perpendicular to \vec{L} . The magnitude of \vec{L} doesn't change but its direction does, and the top precesses about the z-axis.

We can experience this phenomenon first hand by holding a spinning bicycle wheel and trying to rotate it about an axis perpendicular to the spin axis. As shown in [\[link\]](#), the person applies forces perpendicular to the spin axis in an attempt to rotate the wheel, but instead, the wheel axis starts to change direction to her left due to the applied torque.

When forces are applied to the axle as shown, the wheel rotates toward the person.



(a)



(b)

(a) A person holding the spinning bike wheel lifts it with her right hand and pushes down with her left hand in an attempt to rotate the wheel. This action creates a torque directly toward her. This torque causes a change in angular momentum $\Delta\vec{L}$ in exactly the same direction. (b) A vector diagram depicting how $\Delta\vec{L}$ and \vec{L} add, producing a new angular momentum pointing more toward the person. The wheel moves toward the person, perpendicular to the forces she exerts on it.

We all know how easy it is for a bicycle to tip over when sitting on it at rest. But when riding the bicycle at a good pace, tipping it over involves changing the angular momentum vector of the spinning wheels.

Note:

View the video on [gyroscope precession](#) for a complete demonstration of precession of the bicycle wheel.

Also, when a spinning disk is put in a box such as a Blu-Ray player, try to move it. It is easy to translate the box in a given direction but difficult to rotate it about an axis perpendicular to the axis of the spinning disk, since we are putting a torque on the box that will cause the angular momentum vector of the spinning disk to precess.

We can calculate the precession rate of the top in [\[link\]](#). From [\[link\]](#), we see that the magnitude of the torque is

Equation:

$$\tau = rMg \sin \theta.$$

Thus,

Equation:

$$dL = rMg \sin \theta dt.$$

The angle the top precesses through in time dt is

Equation:

$$d\phi = \frac{dL}{L \sin \theta} = \frac{rMg \sin \theta}{L \sin \theta} dt = \frac{rMg}{L} dt.$$

The precession angular velocity is $\omega_P = \frac{d\phi}{dt}$ and from this equation we see that

Equation:

$$\omega_P = \frac{rMg}{L}. \text{ or, since } L = I\omega,$$

Note:

Equation:

$$\omega_P = \frac{rMg}{I\omega}.$$

In this derivation, we assumed that $\omega_P \ll \omega$, that is, that the precession angular velocity is much less than the angular velocity of the gyroscope disk. The precession angular velocity adds a small component to the angular momentum along the z-axis. This is seen in a slight bob up and down as the gyroscope precesses, referred to as nutation.

Earth itself acts like a gigantic gyroscope. Its angular momentum is along its axis and currently points at Polaris, the North Star. But Earth is slowly

precessing (once in about 26,000 years) due to the torque of the Sun and the Moon on its nonspherical shape.

Example:**Period of Precession**

A gyroscope spins with its tip on the ground and is spinning with negligible frictional resistance. The disk of the gyroscope has mass 0.3 kg and is spinning at 20 rev/s. Its center of mass is 5.0 cm from the pivot and the radius of the disk is 5.0 cm. What is the precessional period of the gyroscope?

Strategy

We use [\[link\]](#) to find the precessional angular velocity of the gyroscope. This allows us to find the period of precession.

Solution

The moment of inertia of the disk is

Equation:

$$I = \frac{1}{2}mr^2 = \frac{1}{2}(0.30 \text{ kg})(0.05 \text{ m})^2 = 3.75 \times 10^{-4} \text{ kg} \cdot \text{m}^2.$$

The angular velocity of the disk is

Equation:

$$20.0 \text{ rev/s} = 20.0(2\pi) \text{ rad/s} = 125.66 \text{ rad/s}.$$

We can now substitute in [\[link\]](#). The precessional angular velocity is

Equation:

$$\omega_P = \frac{rMg}{I\omega} = \frac{(0.05 \text{ m})(0.3 \text{ kg})(9.8 \text{ m/s}^2)}{(3.75 \times 10^{-4} \text{ kg} \cdot \text{m}^2)(125.66 \text{ rad/s})} = 3.12 \text{ rad/s}.$$

The precessional period of the gyroscope is

Equation:

$$T_P = \frac{2\pi}{3.12 \text{ rad/s}} = 2.0 \text{ s}.$$

Significance

The precessional angular frequency of the gyroscope, 3.12 rad/s, or about 0.5 rev/s, is much less than the angular velocity 20 rev/s of the gyroscope disk. Therefore, we don't expect a large component of the angular momentum to arise due to precession, and [\[link\]](#) is a good approximation of the precessional angular velocity.

Note:

Exercise:

Problem:

Check Your Understanding A top has a precession frequency of 5.0 rad/s on Earth. What is its precession frequency on the Moon?

Solution:

The Moon's gravity is 1/6 that of Earth's. By examining [\[link\]](#), we see that the top's precession frequency is linearly proportional to the acceleration of gravity. All other quantities, mass, moment of inertia, and spin rate are the same on the Moon. Thus, the precession frequency on the Moon is

$$\omega_P(\text{Moon}) = \frac{1}{6} \omega_P(\text{Earth}) = \frac{1}{6} (5.0 \text{ rad/s}) = 0.83 \text{ rad/s}.$$

Summary

- When a gyroscope is set on a pivot near the surface of Earth, it precesses around a vertical axis, since the torque is always horizontal and perpendicular to \vec{L} . If the gyroscope is not spinning, it acquires angular momentum in the direction of the torque, and it rotates about a horizontal axis, falling over just as we would expect.

- The precessional angular velocity is given by $\omega_P = \frac{rMg}{I\omega}$, where r is the distance from the pivot to the center of mass of the gyroscope, I is the moment of inertia of the gyroscope's spinning disk, M is its mass, and ω is the angular frequency of the gyroscope disk.

Key Equations

Velocity of center of mass of rolling object	$v_{\text{CM}} = R\omega$
Acceleration of center of mass of rolling object	$a_{\text{CM}} = R\alpha$
Displacement of center of mass of rolling object	$d_{\text{CM}} = R\theta$
Acceleration of an object rolling without slipping	$a_{\text{CM}} = \frac{mg \sin \theta}{m + (I_{\text{CM}}/r^2)}$
Angular momentum	$\vec{\mathbf{l}} = \vec{\mathbf{r}} \times \vec{\mathbf{p}}$
Derivative of angular momentum equals torque	$\frac{d\vec{\mathbf{l}}}{dt} = \sum \vec{\boldsymbol{\tau}}$
Angular momentum of a system of particles	$\vec{\mathbf{L}} = \vec{\mathbf{l}}_1 + \vec{\mathbf{l}}_2 + \cdots + \vec{\mathbf{l}}_N$
For a system of particles, derivative of angular momentum equals torque	$\frac{d\vec{\mathbf{L}}}{dt} = \sum \vec{\boldsymbol{\tau}}$

Angular momentum of a rotating rigid body	$L = I\omega$
Conservation of angular momentum	$\frac{d\vec{L}}{dt} = 0$
Conservation of angular momentum	$\vec{L} = \vec{l}_1 + \vec{l}_2 + \cdots + \vec{l}_N = \text{constant}$
Precessional angular velocity	$\omega_P = \frac{rMg}{I\omega}$

Conceptual Questions

Exercise:

Problem:

Gyroscopes used in guidance systems to indicate directions in space must have an angular momentum that does not change in direction. When placed in the vehicle, they are put in a compartment that is separated from the main fuselage, such that changes in the orientation of the fuselage does not affect the orientation of the gyroscope. If the space vehicle is subjected to large forces and accelerations how can the direction of the gyroscopes angular momentum be constant at all times?

Solution:

A torque is needed in the direction perpendicular to the angular momentum vector in order to change its direction. These forces on the space vehicle are external to the container in which the gyroscope is mounted and do not impart torques to the gyroscope's rotating disk.

Exercise:

Problem:

Earth precesses about its vertical axis with a period of 26,000 years. Discuss whether [\[link\]](#) can be used to calculate the precessional angular velocity of Earth.

Problems**Exercise:****Problem:**

A gyroscope has a 0.5-kg disk that spins at 40 rev/s. The center of mass of the disk is 10 cm from a pivot which is also the radius of the disk. What is the precession angular velocity?

Solution:

$$I = 2.5 \times 10^{-3} \text{ kg} \cdot \text{m}^2,$$
$$\omega_P = 0.78 \text{ rad/s}$$

Exercise:**Problem:**

The precession angular velocity of a gyroscope is 1.0 rad/s. If the mass of the rotating disk is 0.4 kg and its radius is 30 cm, as well as the distance from the center of mass to the pivot, what is the rotation rate in rev/s of the disk?

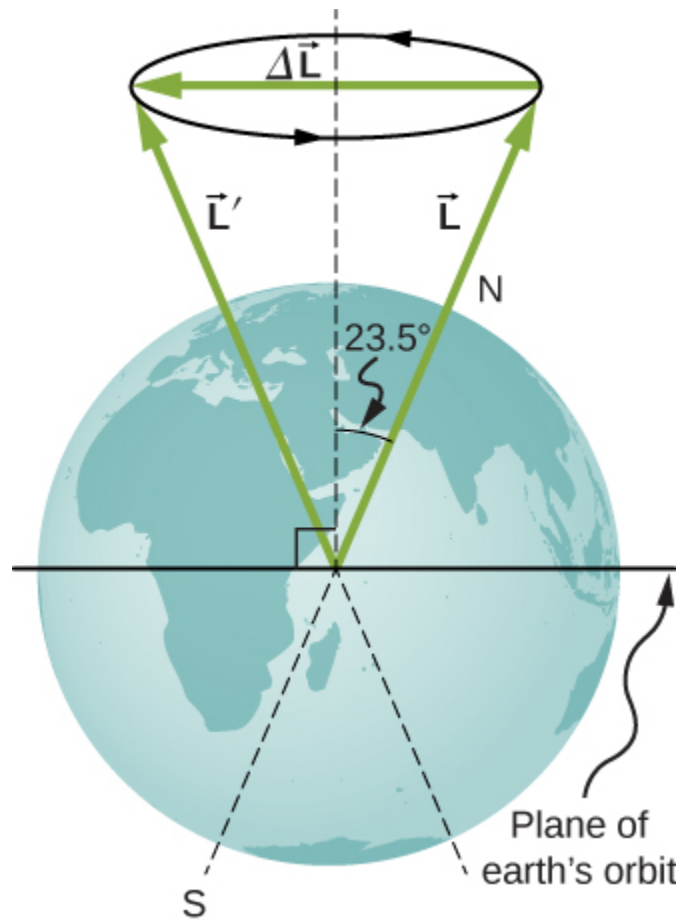
Exercise:**Problem:**

The axis of Earth makes a 23.5° angle with a direction perpendicular to the plane of Earth's orbit. As shown below, this axis precesses, making one complete rotation in 25,780 y.

(a) Calculate the change in angular momentum in half this time.

(b) What is the average torque producing this change in angular momentum?

(c) If this torque were created by a pair of forces acting at the most effective point on the equator, what would the magnitude of each force be?



Solution:

a. $L_{\text{Earth}} = 7.06 \times 10^{33} \text{ kg} \cdot \text{m}^2/\text{s},$

$\Delta L = 5.63 \times 10^{33} \text{ kg} \cdot \text{m}^2/\text{s};$

b. $\tau = 1.4 \times 10^{22} \text{ N} \cdot \text{m};$

c. The two forces at the equator would have the same magnitude but different directions, one in the north direction and the other in the south direction on the opposite side of Earth. The angle between the

forces and the lever arms to the center of Earth is 90° , so a given torque would have magnitude $\tau = FR_E \sin 90^\circ = FR_E$. Both would provide a torque in the same direction:

$$\tau = 2FR_E \Rightarrow F = 1.3 \times 10^{15} \text{ N}$$

Additional Problems

Exercise:

Problem:

A marble is rolling across the floor at a speed of 7.0 m/s when it starts up a plane inclined at 30° to the horizontal. (a) How far along the plane does the marble travel before coming to a rest? (b) How much time elapses while the marble moves up the plane?

Exercise:

Problem:

Repeat the preceding problem replacing the marble with a hollow sphere. Explain the new results.

Solution:

$$a_{\text{CM}} = -\frac{3}{10}g,$$

$$v^2 = v_0^2 + 2a_{\text{CM}}x \Rightarrow v^2 = (7.0 \text{ m/s})^2 - 2\left(\frac{3}{10}g\right)x,$$

$$v^2 = 0 \Rightarrow x = 8.34 \text{ m};$$

$$\text{b. } t = \frac{v-v_0}{a_{\text{CM}}}, \quad v = v_0 + a_{\text{CM}}t \Rightarrow t = 2.38 \text{ s};$$

The hollow sphere has a larger moment of inertia, and therefore is harder to bring to a rest than the marble, or solid sphere. The distance travelled is larger and the time elapsed is longer.

Exercise:

Problem:

The mass of a hoop of radius 1.0 m is 6.0 kg. It rolls across a horizontal surface with a speed of 10.0 m/s. (a) How much work is required to stop the hoop? (b) If the hoop starts up a surface at 30° to the horizontal with a speed of 10.0 m/s, how far along the incline will it travel before stopping and rolling back down?

Exercise:**Problem:**

Repeat the preceding problem for a hollow sphere of the same radius and mass and initial speed. Explain the differences in the results.

Solution:

- a. $W = -500.0 \text{ J}$;
- b. $K + U_{\text{grav}} = \text{constant}$,
 $500 \text{ J} + 0 = 0 + (6.0 \text{ kg})(9.8 \text{ m/s}^2)h$,
 $h = 8.5 \text{ m}$, $d = 17.0 \text{ m}$;

The moment of inertia is less for the hollow sphere, therefore less work is required to stop it. Likewise it rolls up the incline a shorter distance than the hoop.

Exercise:**Problem:**

A particle has mass 0.5 kg and is traveling along the line $x = 5.0 \text{ m}$ at 2.0 m/s in the positive y -direction. What is the particle's angular momentum about the origin?

Exercise:**Problem:**

A 4.0-kg particle moves in a circle of radius 2.0 m. The angular momentum of the particle varies in time according to $l = 5.0t^2$. (a) What is the torque on the particle about the center of the circle at $t = 3.4 \text{ s}$? (b) What is the angular velocity of the particle at $t = 3.4 \text{ s}$?

Solution:

- a. $\tau = 34.0 \text{ N} \cdot \text{m}$;
b. $l = mr^2\omega \Rightarrow \omega = 3.6 \text{ rad/s}$

Exercise:**Problem:**

A proton is accelerated in a cyclotron to $5.0 \times 10^6 \text{ m/s}$ in 0.01 s. The proton follows a circular path. If the radius of the cyclotron is 0.5 km, (a) What is the angular momentum of the proton about the center at its maximum speed? (b) What is the torque on the proton about the center as it accelerates to maximum speed?

Exercise:**Problem:**

(a) What is the angular momentum of the Moon in its orbit around Earth? (b) How does this angular momentum compare with the angular momentum of the Moon on its axis? Remember that the Moon keeps one side toward Earth at all times.

Solution:

- a. $d_M = 3.85 \times 10^8 \text{ m}$ average distance to the Moon; orbital period $27.32 \text{ d} = 2.36 \times 10^6 \text{ s}$; speed of the Moon $\frac{2\pi 3.85 \times 10^8 \text{ m}}{2.36 \times 10^6 \text{ s}} = 1.0 \times 10^3 \text{ m/s}$; mass of the Moon $7.35 \times 10^{22} \text{ kg}$,
 $L = 2.90 \times 10^{34} \text{ kgm}^2/\text{s}$;
b. radius of the Moon $1.74 \times 10^6 \text{ m}$; the orbital period is the same as (a): $\omega = 2.66 \times 10^{-6} \text{ rad/s}$,
 $L = 2.37 \times 10^{29} \text{ kg} \cdot \text{m}^2/\text{s}$;

The orbital angular momentum is 1.22×10^5 times larger than the rotational angular momentum for the Moon.

Exercise:

Problem:

A DVD is rotating at 500 rpm. What is the angular momentum of the DVD if has a radius of 6.0 cm and mass 20.0 g?

Exercise:**Problem:**

A potter's disk spins from rest up to 10 rev/s in 15 s. The disk has a mass 3.0 kg and radius 30.0 cm. What is the angular momentum of the disk at $t = 5$ s, $t = 10$ s?

Solution:

$$\begin{aligned} I &= 0.135 \text{ kg} \cdot \text{m}^2, \\ \alpha &= 4.19 \text{ rad/s}^2, \omega = \omega_0 + \alpha t, \\ \omega(5 \text{ s}) &= 21.0 \text{ rad/s}, L = 2.84 \text{ kg} \cdot \text{m}^2/\text{s}, \\ \omega(10 \text{ s}) &= 41.9 \text{ rad/s}, L = 5.66 \text{ kg} \cdot \text{m}^2/\text{s} \end{aligned}$$

Exercise:**Problem:**

Suppose you start an antique car by exerting a force of 300 N on its crank for 0.250 s. What is the angular momentum given to the engine if the handle of the crank is 0.300 m from the pivot and the force is exerted to create maximum torque the entire time?

Exercise:**Problem:**

A solid cylinder of mass 2.0 kg and radius 20 cm is rotating counterclockwise around a vertical axis through its center at 600 rev/min. A second solid cylinder of the same mass and radius is rotating clockwise around the same vertical axis at 900 rev/min. If the cylinders couple so that they rotate about the same vertical axis, what is the angular velocity of the combination?

Solution:

In the conservation of angular momentum equation, the rotation rate appears on both sides so we keep the (rev/min) notation as the angular velocity can be multiplied by a constant to get (rev/min):

$$L_i = -0.04 \text{ kg} \cdot \text{m}^2 (300.0 \text{ rev/min}),$$

$$L_f = 0.08 \text{ kg} \cdot \text{m}^2 f_f \Rightarrow f_f = -150.0 \text{ rev/min clockwise}$$

Exercise:**Problem:**

A boy stands at the center of a platform that is rotating without friction at 1.0 rev/s. The boy holds weights as far from his body as possible. At this position the total moment of inertia of the boy, platform, and weights is $5.0 \text{ kg} \cdot \text{m}^2$. The boy draws the weights in close to his body, thereby decreasing the total moment of inertia to $1.5 \text{ kg} \cdot \text{m}^2$. (a) What is the final angular velocity of the platform? (b) By how much does the rotational kinetic energy increase?

Exercise:**Problem:**

Eight children, each of mass 40 kg, climb on a small merry-go-round. They position themselves evenly on the outer edge and join hands. The merry-go-round has a radius of 4.0 m and a moment of inertia $1000.0 \text{ kg} \cdot \text{m}^2$. After the merry-go-round is given an angular velocity of 6.0 rev/min, the children walk inward and stop when they are 0.75 m from the axis of rotation. What is the new angular velocity of the merry-go-round? Assume there is negligible frictional torque on the structure.

Solution:

$$I_0 \omega_0 = I_f \omega_f,$$

$$I_0 = 6120.0 \text{ kg} \cdot \text{m}^2,$$

$$I_f = 1180.0 \text{ kg} \cdot \text{m}^2,$$

$$\omega_f = 31.1 \text{ rev/min}$$

Exercise:**Problem:**

A thin meter stick of mass 150 g rotates around an axis perpendicular to the stick's long axis at an angular velocity of 240 rev/min. What is the angular momentum of the stick if the rotation axis (a) passes through the center of the stick? (b) Passes through one end of the stick?

Exercise:**Problem:**

A satellite in the shape of a sphere of mass 20,000 kg and radius 5.0 m is spinning about an axis through its center of mass. It has a rotation rate of 8.0 rev/s. Two antennas deploy in the plane of rotation extending from the center of mass of the satellite. Each antenna can be approximated as a rod has mass 200.0 kg and length 7.0 m. What is the new rotation rate of the satellite?

Solution:

$$L_i = 1.00 \times 10^7 \text{ kg} \cdot \text{m}^2/\text{s},$$

$$I_f = 2.025 \times 10^5 \text{ kg} \cdot \text{m}^2,$$

$$\omega_f = 7.86 \text{ rev/s}$$

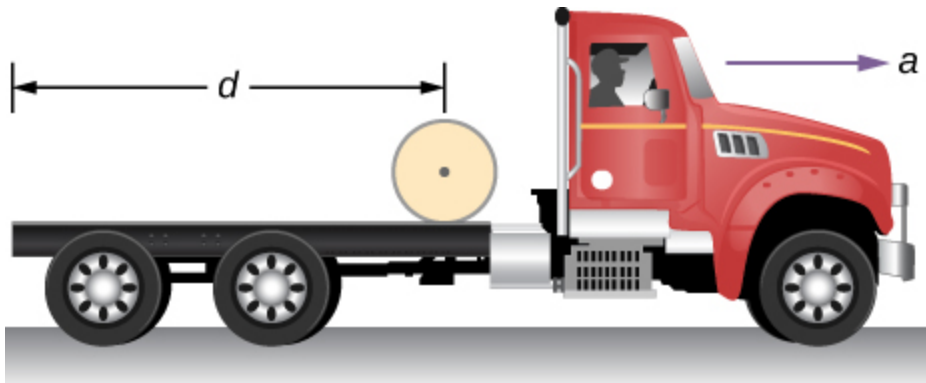
Exercise:**Problem:**

A top has moment of inertia $3.2 \times 10^{-4} \text{ kg} \cdot \text{m}^2$ and radius 4.0 cm from the center of mass to the pivot point. If it spins at 20.0 rev/s and is precessing, how many revolutions does it precess in 10.0 s?

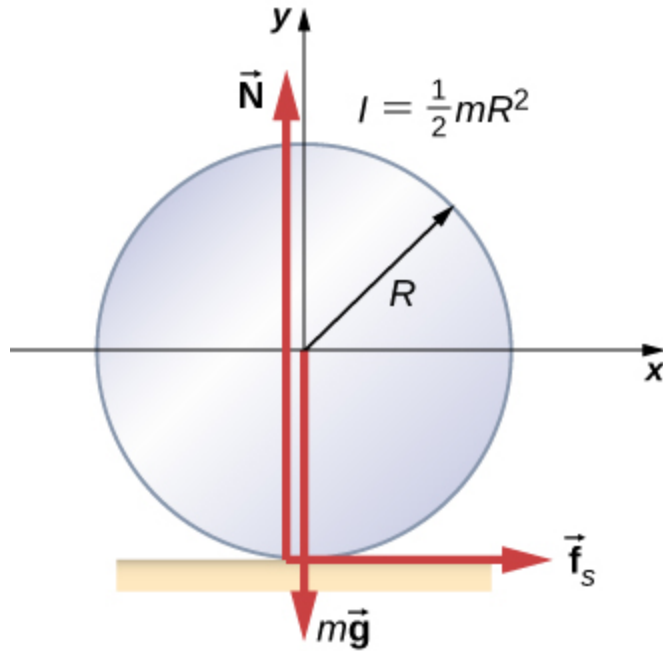
Challenge Problems**Exercise:**

Problem:

The truck shown below is initially at rest with solid cylindrical roll of paper sitting on its bed. If the truck moves forward with a uniform acceleration a , what distance s does it move before the paper rolls off its back end? (*Hint: If the roll accelerates forward with a' , then it accelerates backward relative to the truck with an acceleration $a - a'$. Also, $R\alpha = a - a'$.)*

**Solution:**

Assume the roll accelerates forward with respect to the ground with an acceleration a' . Then it accelerates backwards relative to the truck with an acceleration $(a - a')$.



Also, $R\alpha = a - a'$ $I = \frac{1}{2}mR^2$ $\sum F_x = f_s = ma'$,
 $\sum \tau = f_s R = I\alpha = I \frac{a-a'}{R}$ $f_s = \frac{I}{R^2}(a - a') = \frac{1}{2}m(a - a')$,
 Solving for a' : $f_s = \frac{1}{2}m(a - a')$; $a' = \frac{a}{3}$,
 $x - x_0 = v_0 t + \frac{1}{2}at^2$; $d = \frac{1}{3}at^2$; $t = \sqrt{\frac{3d}{a}}$;
 therefore, $s = 1.5d$

Exercise:

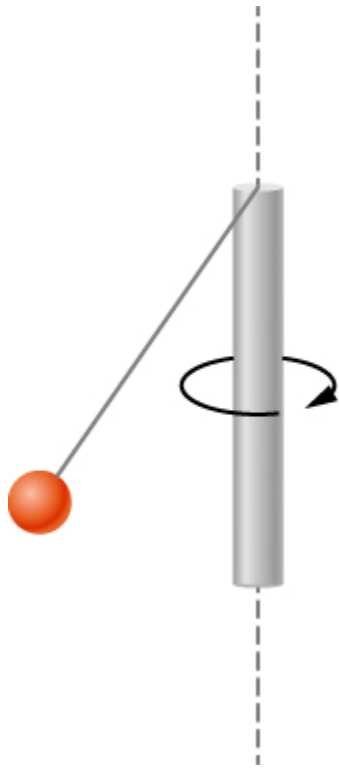
Problem:

A bowling ball of radius 8.5 cm is tossed onto a bowling lane with speed 9.0 m/s. The direction of the toss is to the left, as viewed by the observer, so the bowling ball starts to rotate counterclockwise when in contact with the floor. The coefficient of kinetic friction on the lane is 0.3. (a) What is the time required for the ball to come to the point where it is not slipping? What is the distance d to the point where the ball is rolling without slipping?

Exercise:

Problem:

A small ball of mass 0.50 kg is attached by a massless string to a vertical rod that is spinning as shown below. When the rod has an angular velocity of 6.0 rad/s, the string makes an angle of 30° with respect to the vertical. (a) If the angular velocity is increased to 10.0 rad/s, what is the new angle of the string? (b) Calculate the initial and final angular momenta of the ball. (c) Can the rod spin fast enough so that the ball is horizontal?



Solution:

a. The tension in the string provides the centripetal force such that $T \sin \theta = mr_{\perp} \omega^2$. The component of the tension that is vertical opposes the gravitational force such that $T \cos \theta = mg$. This gives $T = 5.7 \text{ N}$. We solve for $r_{\perp} = 0.16 \text{ m}$. This gives the length of the string as $r = 0.32 \text{ m}$.

At $\omega = 10.0 \text{ rad/s}$, there is a new angle, tension, and perpendicular radius to the rod. Dividing the two equations involving the tension to

eliminate it, we have $\frac{\sin \theta}{\cos \theta} = \frac{(0.32 \text{ m} \sin \theta)\omega^2}{g} \Rightarrow \frac{1}{\cos \theta} = \frac{0.32 m \omega^2}{g}$;
 $\cos \theta = 0.31 \Rightarrow \theta = 72.2^\circ$; b. $l_{\text{initial}} = 0.08 \text{ kg} \cdot \text{m}^2/\text{s}$,
 $l_{\text{final}} = 0.46 \text{ kg} \cdot \text{m}^2/\text{s}$; c. No, the cosine of the angle is inversely proportional to the square of the angular velocity, therefore in order for $\theta \rightarrow 90^\circ$, $\omega \rightarrow \infty$. The rod would have to spin infinitely fast.

Exercise:

Problem:

A bug flying horizontally at 1.0 m/s collides and sticks to the end of a uniform stick hanging vertically. After the impact, the stick swings out to a maximum angle of 5.0° from the vertical before rotating back. If the mass of the stick is 10 times that of the bug, calculate the length of the stick.

Glossary

precession

circular motion of the pole of the axis of a spinning object around another axis due to a torque

Introduction

class="introduction"

Two stilt
walkers in
standing
position. All
forces
acting on
each stilt
walker
balance out;
neither
changes its
translational
motion. In
addition, all
torques
acting on
each person
balance out,
and thus
neither of
them
changes its
rotational
motion. The
result is
static
equilibrium.
(credit:
modificatio
n of work
by Stuart
Redler)



In earlier chapters, you learned about forces and Newton's laws for translational motion. You then studied torques and the rotational motion of a body about a fixed axis of rotation. You also learned that static equilibrium means no motion at all and that dynamic equilibrium means motion without acceleration.

In this chapter, we combine the conditions for static translational equilibrium and static rotational equilibrium to describe situations typical for any kind of construction. What type of cable will support a suspension bridge? What type of foundation will support an office building? Will this prosthetic arm function correctly? These are examples of questions that contemporary engineers must be able to answer.

The elastic properties of materials are especially important in engineering applications, including bioengineering. For example, materials that can stretch or compress and then return to their original form or position make good shock absorbers. In this chapter, you will learn about some applications that combine equilibrium with elasticity to construct real structures that last.

Conditions for Static Equilibrium

By the end of this section, you will be able to:

- Identify the physical conditions of static equilibrium.
- Draw a free-body diagram for a rigid body acted on by forces.
- Explain how the conditions for equilibrium allow us to solve statics problems.

We say that a rigid body is in **equilibrium** when both its linear and angular acceleration are zero relative to an inertial frame of reference. This means that a body in equilibrium can be moving, but if so, its linear and angular velocities must be constant. We say that a rigid body is in **static equilibrium** when it is at rest *in our selected frame of reference*. Notice that the distinction between the state of rest and a state of uniform motion is artificial—that is, an object may be at rest in our selected frame of reference, yet to an observer moving at constant velocity relative to our frame, the same object appears to be in uniform motion with constant velocity. Because the motion is *relative*, what is in static equilibrium to us is in dynamic equilibrium to the moving observer, and vice versa. Since the laws of physics are identical for all inertial reference frames, in an inertial frame of reference, there is no distinction between static equilibrium and equilibrium.

According to Newton's second law of motion, the linear acceleration of a rigid body is caused by a net force acting on it, or

Equation:

$$\sum_k \vec{\mathbf{F}}_k = m\vec{\mathbf{a}}_{\text{CM}}.$$

Here, the sum is of all external forces acting on the body, where m is its mass and $\vec{\mathbf{a}}_{\text{CM}}$ is the linear acceleration of its center of mass (a concept we discussed in [Linear Momentum and Collisions](#) on linear momentum and collisions). In equilibrium, the linear acceleration is zero. If we set the acceleration to zero in [\[link\]](#), we obtain the following equation:

Note:

First Equilibrium Condition

The first equilibrium condition for the static equilibrium of a rigid body expresses *translational* equilibrium:

Equation:

$$\sum_k \vec{\mathbf{F}}_k = \vec{\mathbf{0}}.$$

The first equilibrium condition, [\[link\]](#), is the equilibrium condition for forces, which we encountered when studying applications of Newton's laws.

This vector equation is equivalent to the following three scalar equations for the components of the net force:

Equation:

$$\sum_k F_{kx} = 0, \quad \sum_k F_{ky} = 0, \quad \sum_k F_{kz} = 0.$$

Analogously to [\[link\]](#), we can state that the rotational acceleration $\vec{\alpha}$ of a rigid body about a fixed axis of rotation is caused by the net torque acting on the body, or

Equation:

$$\sum_k \vec{\tau}_k = I\vec{\alpha}.$$

Here I is the rotational inertia of the body in rotation about this axis and the summation is over *all* torques $\vec{\tau}_k$ of external forces in [\[link\]](#). In equilibrium, the rotational acceleration is zero. By setting to zero the right-hand side of [\[link\]](#), we obtain the second equilibrium condition:

Note:

Second Equilibrium Condition

The second equilibrium condition for the static equilibrium of a rigid body expresses *rotational* equilibrium:

Equation:

$$\sum_k \vec{\tau}_k = \vec{0}.$$

The second equilibrium condition, [\[link\]](#), is the equilibrium condition for torques that we encountered when we studied rotational dynamics. It is worth noting that this equation for equilibrium is generally valid for rotational equilibrium about any axis of rotation (fixed or otherwise). Again, this vector equation is equivalent to three scalar equations for the vector components of the net torque:

Equation:

$$\sum_k \tau_{kx} = 0, \quad \sum_k \tau_{ky} = 0, \quad \sum_k \tau_{kz} = 0.$$

The second equilibrium condition means that in equilibrium, there is no net external torque to cause rotation about any axis.

The first and second equilibrium conditions are stated in a particular reference frame. The first condition involves only forces and is therefore independent of the origin of the reference frame. However, the second condition involves torque, which is defined as a cross product, $\vec{\tau}_k = \vec{r}_k \times \vec{F}_k$, where the position vector \vec{r}_k with respect to the axis of rotation of the point where the force is applied enters the equation. Therefore, torque depends on the location of the axis in the reference frame. However, when rotational and translational equilibrium conditions hold simultaneously in one frame of reference, then they also hold in any other inertial frame of reference, so that the net torque about any axis of rotation is still zero. The explanation for this is fairly straightforward.

Suppose vector \vec{R} is the position of the origin of a new inertial frame of reference S' in the old inertial frame of reference S . From our study of relative motion, we know that in the new frame of reference S' , the position vector \vec{r}'_k of the point where the force \vec{F}_k is applied is related to \vec{r}_k via the equation

Equation:

$$\vec{r}'_k = \vec{r}_k - \vec{R}.$$

Now, we can sum all torques $\vec{\tau}'_k = \vec{r}'_k \times \vec{F}_k$ of all external forces in a new reference frame, S' :

Equation:

$$\sum_k \vec{\tau}'_k = \sum_k \vec{r}'_k \times \vec{F}_k = \sum_k (\vec{r}_k - \vec{R}) \times \vec{F}_k = \sum_k \vec{r}_k \times \vec{F}_k - \sum_k \vec{R} \times \vec{F}_k = \sum_k \vec{\tau}_k - \vec{R} \times \sum_k \vec{F}_k = \vec{0}$$

In the final step in this chain of reasoning, we used the fact that in equilibrium in the old frame of reference, S , the first term vanishes because of [\[link\]](#) and the second term vanishes because of [\[link\]](#). Hence, we see that the net torque in any inertial frame of reference S' is zero, provided that both conditions for equilibrium hold in an inertial frame of reference S .

The practical implication of this is that when applying equilibrium conditions for a rigid body, we are free to choose any point as the origin of the reference frame. Our choice of reference frame is dictated by the physical specifics of the problem we are solving. In one frame of reference, the mathematical form of the equilibrium conditions may be quite complicated, whereas in another frame, the same conditions may have a simpler mathematical form that is easy to solve. The origin of a selected frame of reference is called the pivot point.

In the most general case, equilibrium conditions are expressed by the six scalar equations ([\[link\]](#) and [\[link\]](#)). For planar equilibrium problems with rotation about a fixed axis, which we consider in this chapter, we can reduce the number of equations to three. The standard procedure is to adopt a frame of reference where the z -axis is the axis of rotation. With this choice of axis, the net torque has only a z -component, all forces that have non-zero torques lie in the xy -plane, and therefore contributions to the net torque come from only the x - and y -components of external forces. Thus, for planar problems with the axis of rotation perpendicular to the xy -plane, we have the following three equilibrium conditions for forces and torques:

Equation:

$$F_{1x} + F_{2x} + \cdots + F_{Nx} = 0$$

Equation:

$$F_{1y} + F_{2y} + \cdots + F_{Ny} = 0$$

Equation:

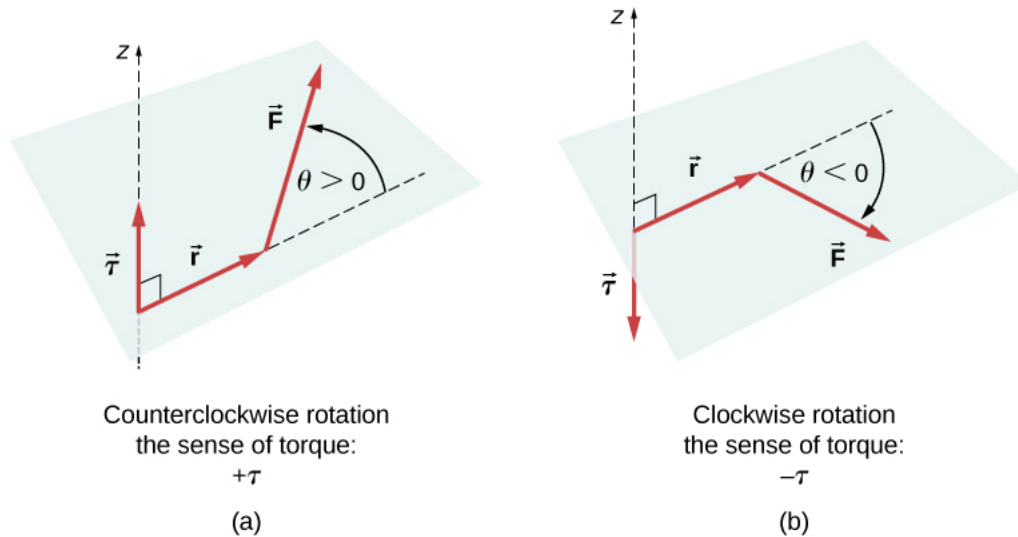
$$\tau_1 + \tau_2 + \cdots + \tau_N = 0$$

where the summation is over all N external forces acting on the body and over their torques. In [\[link\]](#), we simplified the notation by dropping the subscript z , but we understand here that the summation is over all contributions along the z -axis, which is the axis of rotation. In [\[link\]](#), the z -component of torque $\vec{\tau}_k$ from the force \vec{F}_k is

Equation:

$$\tau_k = r_k F_k \sin \theta$$

where r_k is the length of the lever arm of the force and F_k is the magnitude of the force (as you saw in [Fixed-Axis Rotation](#)). The angle θ is the angle between vectors \vec{r}_k and \vec{F}_k , measuring *from vector \vec{r}_k to vector \vec{F}_k* in the *counterclockwise* direction ([\[link\]](#)). When using [\[link\]](#), we often compute the magnitude of torque and assign its sense as either positive (+) or negative (−), depending on the direction of rotation caused by this torque alone. In [\[link\]](#), net torque is the sum of terms, with each term computed from [\[link\]](#), and each term must have the correct *sense*. Similarly, in [\[link\]](#), we assign the + sign to force components in the + x -direction and the − sign to components in the − x -direction. The same rule must be consistently followed in [\[link\]](#), when computing force components along the y -axis.



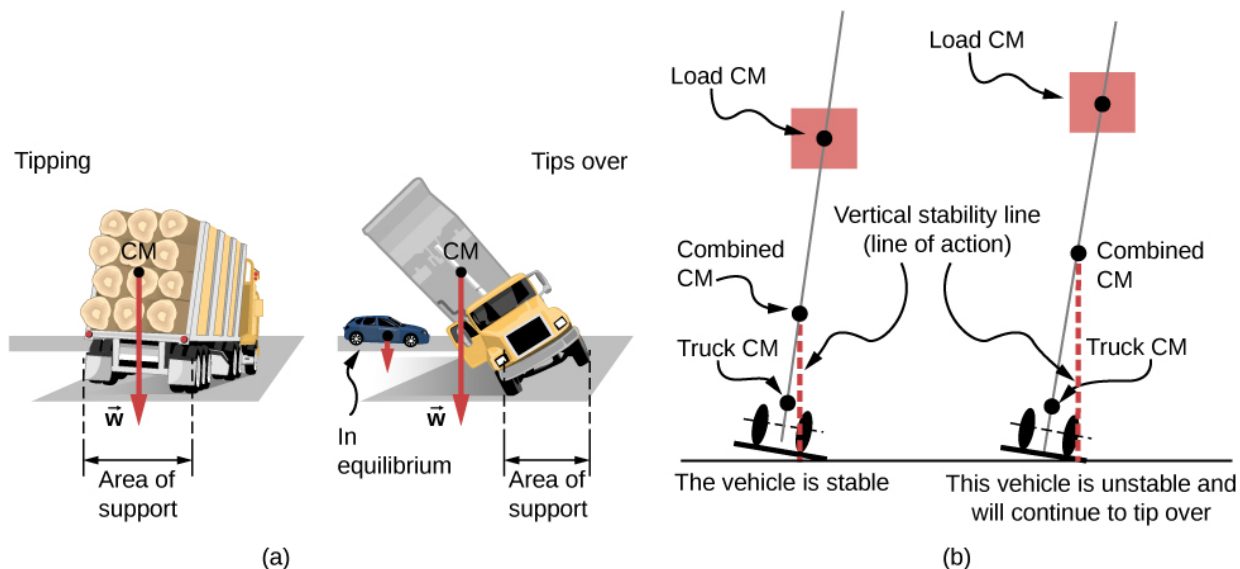
Torque of a force: (a) When the torque of a force causes counterclockwise rotation about the axis of rotation, we say that its *sense* is positive, which means the torque vector is parallel to the axis of rotation. (b) When torque of a force causes clockwise rotation about the axis, we say that its sense is negative, which means the torque vector is antiparallel to the axis of rotation.

Note:

View this [demonstration](#) to see two forces act on a rigid square in two dimensions. At all times, the static equilibrium conditions given by [\[link\]](#) through [\[link\]](#) are satisfied. You can vary magnitudes of the forces and their lever arms and observe the effect these changes have on the square.

In many equilibrium situations, one of the forces acting on the body is its weight. In free-body diagrams, the weight vector is attached to the **center of gravity** of the body. For all practical purposes, the center of gravity is identical to the center of mass, as you learned in [Linear Momentum and Collisions](#) on linear momentum and collisions. Only in situations where a body has a large spatial extension so that the gravitational field is nonuniform throughout its volume, are the center of gravity and the center of mass located at different points. In practical situations, however, even objects as large as buildings or cruise ships are located in a uniform gravitational field on Earth's surface, where the acceleration due to gravity has a constant magnitude of $g = 9.8 \text{ m/s}^2$. In these situations, the center of gravity is identical to the center of mass. Therefore, throughout this chapter, we use the center of mass (CM) as the point where the weight vector is attached. Recall that the CM has a special physical meaning: When an external force is applied to a body at exactly its CM, the body as a whole undergoes translational motion and such a force does not cause rotation.

When the CM is located off the axis of rotation, a net **gravitational torque** occurs on an object. Gravitational torque is the torque caused by weight. This gravitational torque may rotate the object if there is no support present to balance it. The magnitude of the gravitational torque depends on how far away from the pivot the CM is located. For example, in the case of a tipping truck ([\[link\]](#)), the pivot is located on the line where the tires make contact with the road's surface. If the CM is located high above the road's surface, the gravitational torque may be large enough to turn the truck over. Passenger cars with a low-lying CM, close to the pavement, are more resistant to tipping over than are trucks.



The distribution of mass affects the position of the center of mass (CM), where the weight vector \vec{w} is attached. If the center of gravity is within the area of support, the truck returns to its initial position after tipping [see the left panel in (b)]. But if the center of gravity lies outside the area of support, the truck turns over [see the right panel in (b)]. Both vehicles in (b) are out of equilibrium. Notice that the car in (a) is in equilibrium: The low location of its center of gravity makes it hard to tip over.

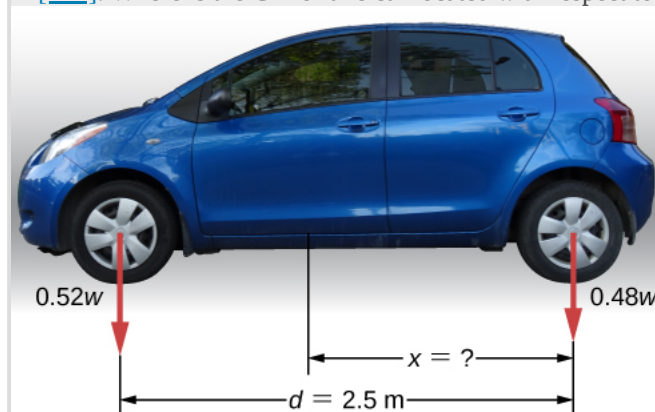
Note:

If you tilt a box so that one edge remains in contact with the table beneath it, then one edge of the base of support becomes a pivot. As long as the center of gravity of the box remains over the base of support, gravitational torque rotates the box back toward its original position of stable equilibrium. When the center of gravity moves outside of the base of support, gravitational torque rotates the box in the opposite direction, and the box rolls over. View this [demonstration](#) to experiment with stable and unstable positions of a box.

Example:

Center of Gravity of a Car

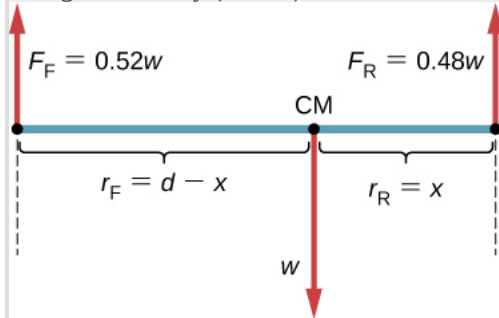
A passenger car with a 2.5-m wheelbase has 52% of its weight on the front wheels on level ground, as illustrated in [\[link\]](#). Where is the CM of this car located with respect to the rear axle?



The weight distribution between the axles of a car.
Where is the center of gravity located? (credit "car":
modification of work by Jane Whitney)

Strategy

We do not know the weight w of the car. All we know is that when the car rests on a level surface, $0.52w$ pushes down on the surface at contact points of the front wheels and $0.48w$ pushes down on the surface at contact points of the rear wheels. Also, the contact points are separated from each other by the distance $d = 2.5$ m. At these contact points, the car experiences normal reaction forces with magnitudes $F_F = 0.52w$ and $F_R = 0.48w$ on the front and rear axles, respectively. We also know that the car is an example of a rigid body in equilibrium whose entire weight w acts at its CM. The CM is located somewhere between the points where the normal reaction forces act, somewhere at a distance x from the point where F_R acts. Our task is to find x . Thus, we identify three forces acting on the body (the car), and we can draw a free-body diagram for the extended rigid body, as shown in [\[link\]](#).



The free-body diagram for the car clearly indicates force vectors acting on the car and distances to the center of mass (CM). When CM is selected as the pivot point, these distances are lever arms of normal reaction forces. Notice that vector magnitudes and lever arms do not need to be drawn to scale, but all quantities of relevance must be clearly labeled.

We are almost ready to write down equilibrium conditions [\[link\]](#) through [\[link\]](#) for the car, but first we must decide on the reference frame. Suppose we choose the x -axis along the length of the car, the y -axis vertical, and the z -axis perpendicular to this xy -plane. With this choice we only need to write [\[link\]](#) and [\[link\]](#) because all the y -components are identically zero. Now we need to decide on the location of the pivot point. We can choose any point as the location of the axis of rotation (z -axis). Suppose we place the axis of rotation at CM, as indicated in the free-body diagram for the car. At this point, we are ready to write the equilibrium conditions for the car.

Solution

Each equilibrium condition contains only three terms because there are $N = 3$ forces acting on the car. The first equilibrium condition, [\[link\]](#), reads

Equation:

$$+F_F - w + F_R = 0.$$

This condition is trivially satisfied because when we substitute the data, [\[link\]](#) becomes $+0.52w - w + 0.48w = 0$. The second equilibrium condition, [\[link\]](#), reads

Equation:

$$\tau_F + \tau_w + \tau_R = 0$$

where τ_F is the torque of force F_F , τ_w is the gravitational torque of force w , and τ_R is the torque of force F_R . When the pivot is located at CM, the gravitational torque is identically zero because the lever arm of the weight with respect to an axis that passes through CM is zero. The lines of action of both normal reaction forces are perpendicular to their lever arms, so in [\[link\]](#), we have $|\sin \theta| = 1$ for both forces. From the free-body diagram, we read that torque τ_F causes clockwise rotation about the pivot at CM, so its sense is negative; and torque τ_R causes counterclockwise rotation about the pivot at CM, so its sense is positive. With this information, we write the second equilibrium condition as

Equation:

$$-r_F F_F + r_R F_R = 0.$$

With the help of the free-body diagram, we identify the force magnitudes $F_R = 0.48w$ and $F_F = 0.52w$, and their corresponding lever arms $r_R = x$ and $r_F = d - x$. We can now write the second equilibrium condition, [\[link\]](#), explicitly in terms of the unknown distance x :

Equation:

$$-0.52(d - x)w + 0.48xw = 0.$$

Here the weight w cancels and we can solve the equation for the unknown position x of the CM. The answer is $x = 0.52d = 0.52(2.5 \text{ m}) = 1.3 \text{ m}$.

Solution

Choosing the pivot at the position of the front axle does not change the result. The free-body diagram for this pivot location is presented in [\[link\]](#). For this choice of pivot point, the second equilibrium condition is

Equation:

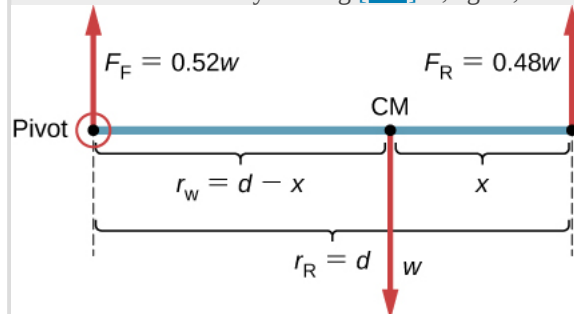
$$-r_w w + r_R F_R = 0.$$

When we substitute the quantities indicated in the diagram, we obtain

Equation:

$$-(d - x)w + 0.48dw = 0.$$

The answer obtained by solving [\[link\]](#) is, again, $x = 0.52d = 1.3 \text{ m}$.



The equivalent free-body diagram for the car; the pivot is clearly indicated.

Significance

This example shows that when solving static equilibrium problems, we are free to choose the pivot location. For different choices of the pivot point we have different sets of equilibrium conditions to solve. However, all choices lead to the same solution to the problem.

Note:

Exercise:

Problem: Check Your Understanding Solve [\[link\]](#) by choosing the pivot at the location of the rear axle.

Solution:

$$x = 1.3 \text{ m}$$

Note:

Exercise:

Problem:

Check Your Understanding Explain which one of the following situations satisfies both equilibrium conditions: (a) a tennis ball that does not spin as it travels in the air; (b) a pelican that is gliding in the air at a constant velocity at one altitude; or (c) a crankshaft in the engine of a parked car.

Solution:

(b), (c)

A special case of static equilibrium occurs when all external forces on an object act at or along the axis of rotation or when the spatial extension of the object can be disregarded. In such a case, the object can be effectively treated like a point mass. In this special case, we need not worry about the second equilibrium condition, [\[link\]](#), because all torques are identically zero and the first equilibrium condition (for forces) is the only condition to be satisfied. The free-body diagram and problem-solving strategy for this special case were outlined in [Newton's Laws of Motion](#) and [Applications of Newton's Laws](#). You will see a typical equilibrium situation involving only the first equilibrium condition in the next example.

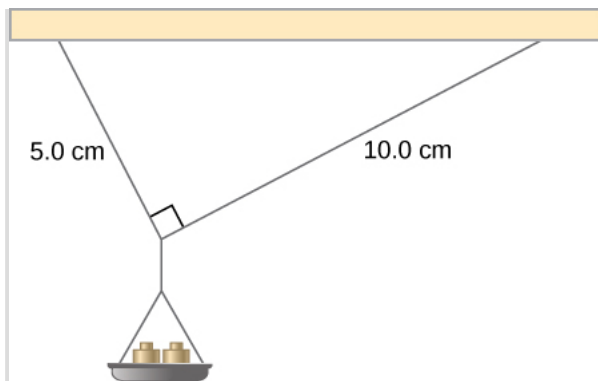
Note:

View this [demonstration](#) to see three weights that are connected by strings over pulleys and tied together in a knot. You can experiment with the weights to see how they affect the equilibrium position of the knot and, at the same time, see the vector-diagram representation of the first equilibrium condition at work.

Example:

A Breaking Tension

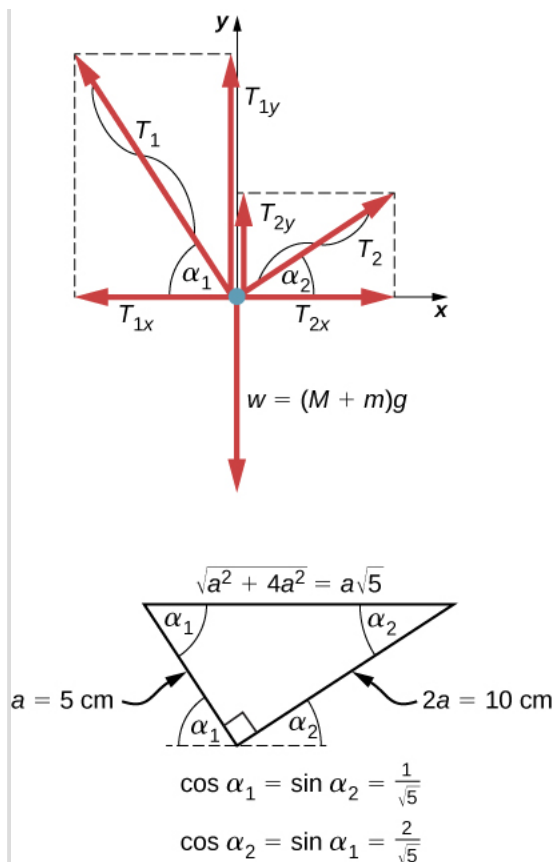
A small pan of mass 42.0 g is supported by two strings, as shown in [\[link\]](#). The maximum tension that the string can support is 2.80 N. Mass is added gradually to the pan until one of the strings snaps. Which string is it? How much mass must be added for this to occur?



Mass is added gradually to the pan until one of the strings snaps.

Strategy

This mechanical system consisting of strings, masses, and the pan is in static equilibrium. Specifically, the knot that ties the strings to the pan is in static equilibrium. The knot can be treated as a point; therefore, we need only the first equilibrium condition. The three forces pulling at the knot are the tension \vec{T}_1 in the 5.0-cm string, the tension \vec{T}_2 in the 10.0-cm string, and the weight \vec{w} of the pan holding the masses. We adopt a rectangular coordinate system with the y -axis pointing opposite to the direction of gravity and draw the free-body diagram for the knot (see [\[link\]](#)). To find the tension components, we must identify the direction angles α_1 and α_2 that the strings make with the horizontal direction that is the x -axis. As you can see in [\[link\]](#), the strings make two sides of a right triangle. We can use the Pythagorean theorem to solve this triangle, shown in [\[link\]](#), and find the sine and cosine of the angles α_1 and α_2 . Then we can resolve the tensions into their rectangular components, substitute in the first condition for equilibrium ([\[link\]](#) and [\[link\]](#)), and solve for the tensions in the strings. The string with a greater tension will break first.



Free-body diagram for the knot in [\[link\]](#).

Solution

The weight w pulling on the knot is due to the mass M of the pan and mass m added to the pan, or $w = (M + m)g$. With the help of the free-body diagram in [\[link\]](#), we can set up the equilibrium conditions for the knot:

Equation:

$$\begin{aligned} \text{in the } x\text{-direction,} \quad & -T_{1x} + T_{2x} = 0 \\ \text{in the } y\text{-direction,} \quad & +T_{1y} + T_{2y} - w = 0. \end{aligned}$$

From the free-body diagram, the magnitudes of components in these equations are

Equation:

$$\begin{aligned} T_{1x} &= T_1 \cos \alpha_1 = T_1 / \sqrt{5}, & T_{1y} &= T_1 \sin \alpha_1 = 2T_1 / \sqrt{5} \\ T_{2x} &= T_2 \cos \alpha_2 = 2T_2 / \sqrt{5}, & T_{2y} &= T_2 \sin \alpha_2 = T_2 / \sqrt{5}. \end{aligned}$$

We substitute these components into the equilibrium conditions and simplify. We then obtain two equilibrium equations for the tensions:

Equation:

$$\begin{aligned} \text{in } x\text{-direction,} \quad & T_1 = 2T_2 \\ \text{in } y\text{-direction,} \quad & \frac{2T_1}{\sqrt{5}} + \frac{T_2}{\sqrt{5}} = (M + m)g. \end{aligned}$$

The equilibrium equation for the x -direction tells us that the tension T_1 in the 5.0-cm string is twice the tension T_2 in the 10.0-cm string. Therefore, the shorter string will snap. When we use the first equation to eliminate T_2 from the second equation, we obtain the relation between the mass m on the pan and the tension T_1 in the shorter string:

Equation:

$$2.5T_1/\sqrt{5} = (M + m)g.$$

The string breaks when the tension reaches the critical value of $T_1 = 2.80$ N. The preceding equation can be solved for the critical mass m that breaks the string:

Equation:

$$m = \frac{2.5}{\sqrt{5}} \frac{T_1}{g} - M = \frac{2.5}{\sqrt{5}} \frac{2.80 \text{ N}}{9.8 \text{ m/s}^2} - 0.042 \text{ kg} = 0.277 \text{ kg} = 277.0 \text{ g}.$$

Significance

Suppose that the mechanical system considered in this example is attached to a ceiling inside an elevator going up. As long as the elevator moves up at a constant speed, the result stays the same because the weight w does not change. If the elevator moves up with acceleration, the critical mass is smaller because the weight of $M + m$ becomes larger by an apparent weight due to the acceleration of the elevator. Still, in all cases the shorter string breaks first.

Summary

- A body is in equilibrium when it remains either in uniform motion (both translational and rotational) or at rest. When a body in a selected inertial frame of reference neither rotates nor moves in translational motion, we say the body is in static equilibrium in this frame of reference.
- Conditions for equilibrium require that the sum of all external forces acting on the body is zero (first condition of equilibrium), and the sum of all external torques from external forces is zero (second condition of equilibrium). These two conditions must be simultaneously satisfied in equilibrium. If one of them is not satisfied, the body is not in equilibrium.
- The free-body diagram for a body is a useful tool that allows us to count correctly all contributions from all external forces and torques acting on the body. Free-body diagrams for the equilibrium of an extended rigid body must indicate a pivot point and lever arms of acting forces with respect to the pivot.

Conceptual Questions

Exercise:

Problem: What can you say about the velocity of a moving body that is in dynamic equilibrium?

Solution:

constant

Exercise:

Problem: Under what conditions can a rotating body be in equilibrium? Give an example.

Exercise:

Problem: What three factors affect the torque created by a force relative to a specific pivot point?

Solution:

magnitude and direction of the force, and its lever arm

Exercise:

Problem:

Mechanics sometimes put a length of pipe over the handle of a wrench when trying to remove a very tight bolt. How does this help?

For the next four problems, evaluate the statement as either true or false and explain your answer.

Exercise:

Problem: If there is only one external force (or torque) acting on an object, it cannot be in equilibrium.

Solution:

True, as the sum of forces cannot be zero in this case unless the force itself is zero.

Exercise:

Problem: If an object is in equilibrium there must be an even number of forces acting on it.

Exercise:

Problem: If an odd number of forces act on an object, the object cannot be in equilibrium.

Solution:

False, provided forces add to zero as vectors then equilibrium can be achieved.

Exercise:

Problem: A body moving in a circle with a constant speed is in rotational equilibrium.

Exercise:

Problem: What purpose is served by a long and flexible pole carried by wire-walkers?

Solution:

It helps a wire-walker to maintain equilibrium.

Problems

Exercise:

Problem:

When tightening a bolt, you push perpendicularly on a wrench with a force of 165 N at a distance of 0.140 m from the center of the bolt. How much torque are you exerting relative to the center of the bolt?

Exercise:

Problem:

When opening a door, you push on it perpendicularly with a force of 55.0 N at a distance of 0.850 m from the hinges. What torque are you exerting relative to the hinges?

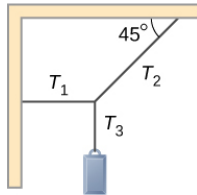
Solution:

$46.8 \text{ N} \cdot \text{m}$

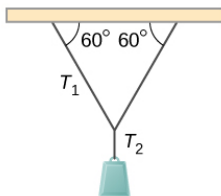
Exercise:

Problem:

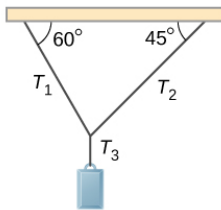
Find the magnitude of the tension in each supporting cable shown below. In each case, the weight of the suspended body is 100.0 N and the masses of the cables are negligible.



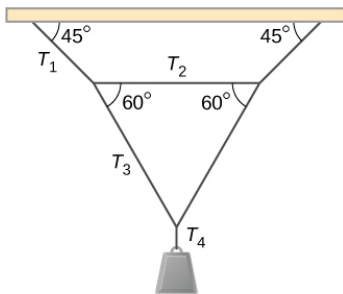
(a)



(b)



(c)

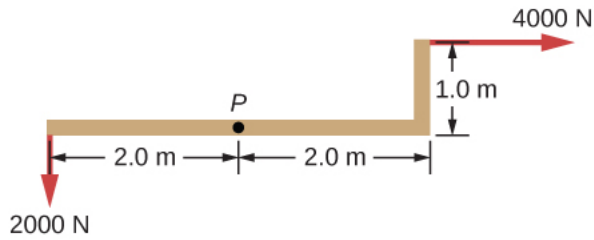


(d)

Exercise:

Problem:

What force must be applied at point P to keep the structure shown in equilibrium? The weight of the structure is negligible.



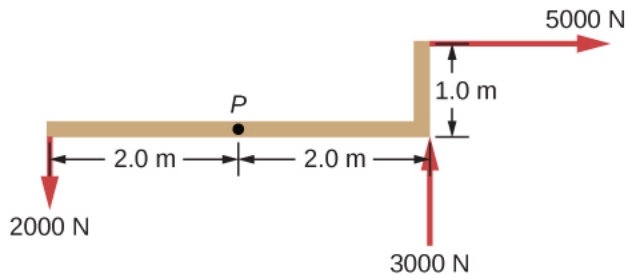
Solution:

4,472 N, 153.4°

Exercise:

Problem:

Is it possible to apply a force at P to keep in equilibrium the structure shown? The weight of the structure is negligible.



Exercise:

Problem:

Two children push on opposite sides of a door during play. Both push horizontally and perpendicular to the door. One child pushes with a force of 17.5 N at a distance of 0.600 m from the hinges, and the second child pushes at a distance of 0.450 m. What force must the second child exert to keep the door from moving? Assume friction is negligible.

Solution:

23.3 N

Exercise:

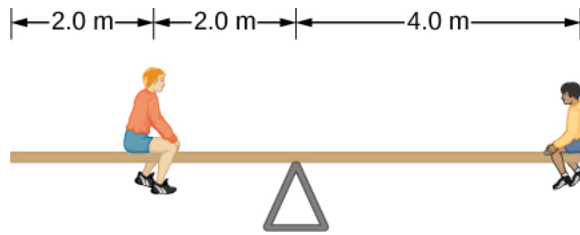
Problem:

A small 1000-kg SUV has a wheel base of 3.0 m. If 60% of its weight rests on the front wheels, how far behind the front wheels is the wagon's center of mass?

Exercise:

Problem:

The uniform seesaw is balanced at its center of mass, as seen below. The smaller boy on the right has a mass of 40.0 kg. What is the mass of his friend?



Solution:

80.0 kg

Glossary

center of gravity

point where the weight vector is attached

equilibrium

body is in equilibrium when its linear and angular accelerations are both zero relative to an inertial frame of reference

first equilibrium condition

expresses translational equilibrium; all external forces acting on the body balance out and their vector sum is zero

gravitational torque

torque on the body caused by its weight; it occurs when the center of gravity of the body is not located on the axis of rotation

second equilibrium condition

expresses rotational equilibrium; all torques due to external forces acting on the body balance out and their vector sum is zero

static equilibrium

body is in static equilibrium when it is at rest in our selected inertial frame of reference

Stress, Strain, and Elastic Modulus

By the end of this section, you will be able to:

- Explain the concepts of stress and strain in describing elastic deformations of materials
- Describe the types of elastic deformation of objects and materials

A model of a rigid body is an idealized example of an object that does not deform under the actions of external forces. It is very useful when analyzing mechanical systems—and many physical objects are indeed rigid to a great extent. The extent to which an object can be *perceived* as rigid depends on the physical properties of the material from which it is made. For example, a ping-pong ball made of plastic is brittle, and a tennis ball made of rubber is elastic when acted upon by squashing forces. However, under other circumstances, both a ping-pong ball and a tennis ball may bounce well as rigid bodies. Similarly, someone who designs prosthetic limbs may be able to approximate the mechanics of human limbs by modeling them as rigid bodies; however, the actual combination of bones and tissues is an elastic medium.

For the remainder of this chapter, we move from consideration of forces that affect the motion of an object to those that affect an object's shape. A change in shape due to the application of a force is known as a deformation. Even very small forces are known to cause some deformation. Deformation is experienced by objects or physical media under the action of external forces—for example, this may be squashing, squeezing, ripping, twisting, shearing, or pulling the objects apart. In the language of physics, two terms describe the forces on objects undergoing deformation: *stress* and *strain*.

Stress is a quantity that describes the magnitude of forces that cause deformation. Stress is generally defined as *force per unit area*. When forces pull on an object and cause its elongation, like the stretching of an elastic band, we call such stress a **tensile stress**. When forces cause a compression of an object, we call it a **compressive stress**. When an object is being squeezed from all sides, like a submarine in the depths of an ocean, we call this kind of stress a **bulk stress** (or **volume stress**). In other situations, the acting forces may be neither tensile nor compressive, and still produce a noticeable deformation. For example, suppose you hold a book tightly between the palms of your hands, then with one hand you press-and-pull on the front cover away from you, while with the other hand you press-and-pull on the back cover toward you. In such a case, when deforming forces act tangentially to the object's surface, we call them 'shear' forces and the stress they cause is called **shear stress**.

The SI unit of stress is the pascal (Pa). When one newton of force presses on a unit surface area of one meter squared, the resulting stress is one pascal:

Equation:

$$\text{one pascal} = 1.0 \text{ Pa} = \frac{1.0 \text{ N}}{1.0 \text{ m}^2}.$$

In the British system of units, the unit of stress is 'psi,' which stands for 'pound per square inch' (lb/in^2). Another unit that is often used for bulk stress is the atm (atmosphere). Conversion factors are

Equation:

$$\begin{aligned} 1 \text{ psi} &= 6895 \text{ Pa} \quad \text{and} \quad 1 \text{ Pa} = 1.450 \times 10^{-4} \text{ psi} \\ 1 \text{ atm} &= 1.013 \times 10^5 \text{ Pa} = 14.7 \text{ psi}. \end{aligned}$$

An object or medium under stress becomes deformed. The quantity that describes this deformation is called **strain**. Strain is given as a fractional change in either length (under tensile stress) or volume (under bulk stress) or geometry (under shear stress). Therefore, strain is a dimensionless number. Strain under a tensile stress is called **tensile strain**, strain under bulk stress is called **bulk strain** (or **volume strain**), and that caused by shear stress is called **shear strain**.

The greater the stress, the greater the strain; however, the relation between strain and stress does not need to be linear. Only when stress is sufficiently low is the deformation it causes in direct proportion to the stress value. The proportionality constant in this relation is called the **elastic modulus**. In the linear limit of low stress values, the general relation between stress and strain is

Note:
Equation:

$$\text{stress} = (\text{elastic modulus}) \times \text{strain}.$$

As we can see from dimensional analysis of this relation, the elastic modulus has the same physical unit as stress because strain is dimensionless.

We can also see from [\[link\]](#) that when an object is characterized by a large value of elastic modulus, the effect of stress is small. On the other hand, a small elastic modulus means that stress produces large strain and noticeable deformation. For example, a stress on a rubber band produces larger strain (deformation) than the same stress on a steel band of the same dimensions because the elastic modulus for rubber is two orders of magnitude smaller than the elastic modulus for steel.

The elastic modulus for tensile stress is called **Young’s modulus**; that for the bulk stress is called the **bulk modulus**; and that for shear stress is called the **shear modulus**. Note that the relation between stress and strain is an *observed* relation, measured in the laboratory. Elastic moduli for various materials are measured under various physical conditions, such as varying temperature, and collected in engineering data tables for reference ([\[link\]](#)). These tables are valuable references for industry and for anyone involved in engineering or construction. In the next section, we discuss strain-stress relations beyond the linear limit represented by [\[link\]](#), in the full range of stress values up to a fracture point. In the remainder of this section, we study the linear limit expressed by [\[link\]](#).

Material	Young’s modulus × 10 ¹⁰ Pa	Bulk modulus × 10 ¹⁰ Pa	Shear modulus × 10 ¹⁰ Pa
Aluminum	7.0	7.5	2.5
Bone (tension)	1.6	0.8	8.0
Bone (compression)	0.9		

Material	Young's modulus $\times 10^{10}\text{Pa}$	Bulk modulus $\times 10^{10}\text{Pa}$	Shear modulus $\times 10^{10}\text{Pa}$
Brass	9.0	6.0	3.5
Brick	1.5		
Concrete	2.0		
Copper	11.0	14.0	4.4
Crown glass	6.0	5.0	2.5
Granite	4.5	4.5	2.0
Hair (human)	1.0		
Hardwood	1.5		1.0
Iron	21.0	16.0	7.7
Lead	1.6	4.1	0.6
Marble	6.0	7.0	2.0
Nickel	21.0	17.0	7.8
Polystyrene	3.0		
Silk	6.0		
Spider thread	3.0		
Steel	20.0	16.0	7.5
Acetone		0.07	
Ethanol		0.09	
Glycerin		0.45	
Mercury		2.5	
Water		0.22	

Approximate Elastic Moduli for Selected Materials

Tensile or Compressive Stress, Strain, and Young's Modulus

Tension or compression occurs when two antiparallel forces of equal magnitude act on an object along only one of its dimensions, in such a way that the object does not move. One way to envision such a situation is illustrated in [\[link\]](#). A rod segment is either stretched or squeezed by a pair of forces acting along its length and perpendicular to its cross-section. The net effect of such forces is that the rod changes its length from the original length L_0 that it had before the forces appeared, to a new length L that it has under the action of the forces. This change in length $\Delta L = L - L_0$ may be either elongation (when L is larger than the original length L_0) or contraction (when L is smaller than the original length L_0). Tensile stress and strain occur when the forces are stretching an object, causing its elongation, and the length change ΔL is positive. Compressive stress and strain occur when the forces are contracting an object, causing its shortening, and the length change ΔL is negative.

In either of these situations, we define stress as the ratio of the deforming force F_{\perp} to the cross-sectional area A of the object being deformed. The symbol F_{\perp} that we reserve for the deforming force means that this force acts perpendicularly to the cross-section of the object. Forces that act parallel to the cross-section do not change the length of an object. The definition of the tensile stress is

Equation:

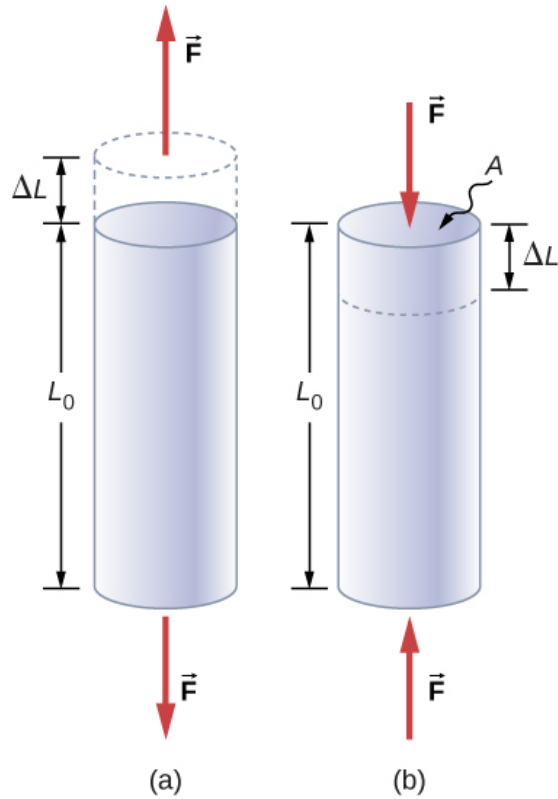
$$\text{tensile stress} = \frac{F_{\perp}}{A}.$$

Tensile strain is the measure of the deformation of an object under tensile stress and is defined as the fractional change of the object's length when the object experiences tensile stress

Equation:

$$\text{tensile strain} = \frac{\Delta L}{L_0}.$$

Compressive stress and strain are defined by the same formulas, [\[link\]](#) and [\[link\]](#), respectively. The only difference from the tensile situation is that for compressive stress and strain, we take absolute values of the right-hand sides in [\[link\]](#) and [\[link\]](#).



When an object is in either tension or compression, the net force on it is zero, but the object deforms by changing its original length L_0 . (a) Tension: The rod is elongated by ΔL . (b) Compression: The rod is contracted by ΔL . In both cases, the deforming force acts along the length of the rod and perpendicular to its cross-section. In the linear range of low stress, the cross-sectional area of the rod does not change.

Young's modulus Y is the elastic modulus when deformation is caused by either tensile or compressive stress, and is defined by [\[link\]](#). Dividing this equation by tensile strain, we obtain the expression for Young's modulus:

Note:

Equation:

$$Y = \frac{\text{tensile stress}}{\text{tensile strain}} = \frac{F_{\perp}/A}{\Delta L/L_0} = \frac{F_{\perp}}{A} \frac{L_0}{\Delta L}.$$

Example:**Compressive Stress in a Pillar**

A sculpture weighing 10,000 N rests on a horizontal surface at the top of a 6.0-m-tall vertical pillar [\[link\]](#). The pillar's cross-sectional area is 0.20 m^2 and it is made of granite with a mass density of 2700 kg/m^3 . Find the compressive stress at the cross-section located 3.0 m below the top of the pillar and the value of the compressive strain of the top 3.0-m segment of the pillar.



Nelson's Column in Trafalgar Square, London, England. (credit: modification of work by Cristian Bortes)

Strategy

First we find the weight of the 3.0-m-long top section of the pillar. The normal force that acts on the cross-section located 3.0 m down from the top is the sum of the pillar's weight and the sculpture's weight. Once we have the normal force, we use [\[link\]](#) to find the stress. To find the compressive strain, we find the value of Young's modulus for granite in [\[link\]](#) and invert [\[link\]](#).

Solution

The volume of the pillar segment with height $h = 3.0 \text{ m}$ and cross-sectional area $A = 0.20 \text{ m}^2$ is

Equation:

$$V = Ah = (0.20 \text{ m}^2)(3.0 \text{ m}) = 0.60 \text{ m}^3.$$

With the density of granite $\rho = 2.7 \times 10^3 \text{ kg/m}^3$, the mass of the pillar segment is

Equation:

$$m = \rho V = (2.7 \times 10^3 \text{ kg/m}^3)(0.60 \text{ m}^3) = 1.60 \times 10^3 \text{ kg}.$$

The weight of the pillar segment is

Equation:

$$w_p = mg = (1.60 \times 10^3 \text{ kg})(9.80 \text{ m/s}^2) = 1.568 \times 10^4 \text{ N}.$$

The weight of the sculpture is $w_s = 1.0 \times 10^4 \text{ N}$, so the normal force on the cross-sectional surface located 3.0 m below the sculpture is

Equation:

$$F_{\perp} = w_p + w_s = (1.568 + 1.0) \times 10^4 \text{ N} = 2.568 \times 10^4 \text{ N}.$$

Therefore, the stress is

Equation:

$$\text{stress} = \frac{F_{\perp}}{A} = \frac{2.568 \times 10^4 \text{ N}}{0.20 \text{ m}^2} = 1.284 \times 10^5 \text{ Pa} = 128.4 \text{ kPa}.$$

Young's modulus for granite is $Y = 4.5 \times 10^{10} \text{ Pa} = 4.5 \times 10^7 \text{ kPa}$. Therefore, the compressive strain at this position is

Equation:

$$\text{strain} = \frac{\text{stress}}{Y} = \frac{128.4 \text{ kPa}}{4.5 \times 10^7 \text{ kPa}} = 2.85 \times 10^{-6}.$$

Significance

Notice that the normal force acting on the cross-sectional area of the pillar is not constant along its length, but varies from its smallest value at the top to its largest value at the bottom of the pillar. Thus, if the pillar has a uniform cross-sectional area along its length, the stress is largest at its base.

Note:**Exercise:****Problem:**

Check Your Understanding Find the compressive stress and strain at the base of Nelson's column.

Solution:

206.8 kPa; 4.6×10^{-5}

Example:**Stretching a Rod**

A 2.0-m-long steel rod has a cross-sectional area of 0.30 cm^2 . The rod is a part of a vertical support that holds a heavy 550-kg platform that hangs attached to the rod's lower end. Ignoring the weight of the rod, what is the tensile stress in the rod and the elongation of the rod under the stress?

Strategy

First we compute the tensile stress in the rod under the weight of the platform in accordance with [\[link\]](#). Then we invert [\[link\]](#) to find the rod's elongation, using $L_0 = 2.0 \text{ m}$. From [\[link\]](#), Young's modulus for steel is $Y = 2.0 \times 10^{11} \text{ Pa}$.

Solution

Substituting numerical values into the equations gives us

Equation:

$$\frac{F_{\perp}}{A} = \frac{(550 \text{ kg})(9.8 \text{ m/s}^2)}{3.0 \times 10^{-5} \text{ m}^2} = 1.8 \times 10^8 \text{ Pa}$$

$$\Delta L = \frac{F_{\perp}}{A} \frac{L_0}{Y} = (1.8 \times 10^8 \text{ Pa}) \frac{2.0 \text{ m}}{2.0 \times 10^{11} \text{ Pa}} = 1.8 \times 10^{-3} \text{ m} = 1.8 \text{ mm}.$$

Significance

Similarly as in the example with the column, the tensile stress in this example is not uniform along the length of the rod. Unlike in the previous example, however, if the weight of the rod is taken into consideration, the stress in the rod is largest at the top and smallest at the bottom of the rod where the equipment is attached.

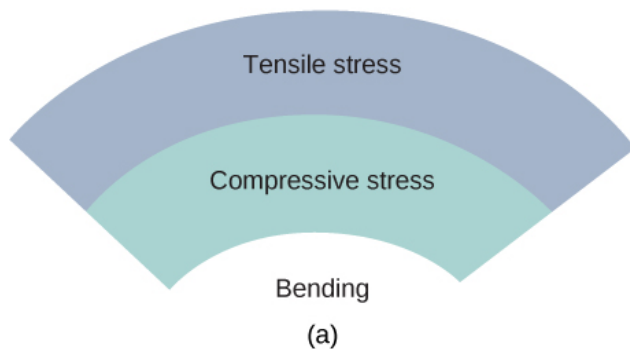
Note:**Exercise:****Problem:**

Check Your Understanding A 2.0-m-long wire stretches 1.0 mm when subjected to a load. What is the tensile strain in the wire?

Solution:

$$5.0 \times 10^{-4}$$

Objects can often experience both compressive stress and tensile stress simultaneously [\[link\]](#). One example is a long shelf loaded with heavy books that sags between the end supports under the weight of the books. The top surface of the shelf is in compressive stress and the bottom surface of the shelf is in tensile stress. Similarly, long and heavy beams sag under their own weight. In modern building construction, such bending strains can be almost eliminated with the use of I-beams [\[link\]](#).



(a) An object bending downward experiences tensile stress (stretching) in the upper section and compressive stress (compressing) in the lower section. (b) Elite weightlifters often bend iron bars

temporarily during lifting, as in the 2012 Olympics competition. (credit b: modification of work by Oleksandr Kocherzhenko)



Steel I-beams are used in construction to reduce bending strains. (credit: modification of work by “US Army Corps of Engineers Europe District”/Flickr)

Note:

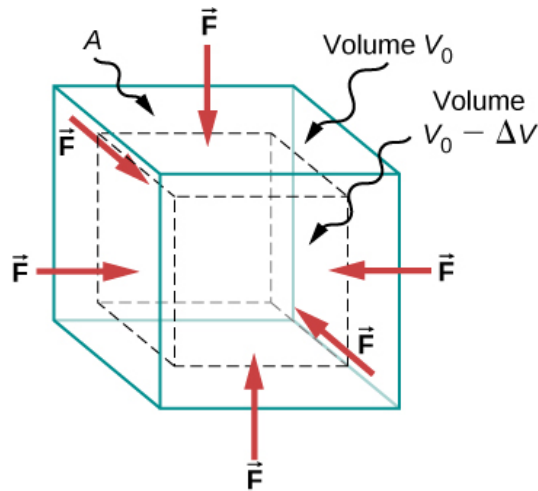
A heavy box rests on a table supported by three columns. View this [demonstration](#) to move the box to see how the compression (or tension) in the columns is affected when the box changes its position.

Bulk Stress, Strain, and Modulus

When you dive into water, you feel a force pressing on every part of your body from all directions. What you are experiencing then is bulk stress, or in other words, **pressure**. Bulk stress always tends to decrease the volume enclosed by the surface of a submerged object. The forces of this “squeezing” are always perpendicular to the submerged surface [\[link\]](#). The effect of these forces is to decrease the volume of the submerged object by an amount ΔV compared with the volume V_0 of the object in the absence of bulk stress. This kind of deformation is called bulk strain and is described by a change in volume relative to the original volume:

Equation:

$$\text{bulk strain} = \frac{\Delta V}{V_0}.$$



An object under increasing bulk stress always undergoes a decrease in its volume. Equal forces perpendicular to the surface act from all directions. The effect of these forces is to decrease the volume by the amount ΔV compared to the original volume, V_0 .

The bulk strain results from the bulk stress, which is a force F_{\perp} normal to a surface that presses on the unit surface area A of a submerged object. This kind of physical quantity, or pressure p , is defined as
Equation:

$$\text{pressure} = p \equiv \frac{F_{\perp}}{A}.$$

We will study pressure in fluids in greater detail in [Fluid Mechanics](#). An important characteristic of pressure is that it is a scalar quantity and does not have any particular direction; that is, pressure acts equally in all possible directions. When you submerge your hand in water, you sense the same amount of pressure acting on the top surface of your hand as on the bottom surface, or on the side surface, or on the surface of the skin between your fingers. What you are perceiving in this case is an increase in pressure Δp over what you are used to feeling when your hand is not submerged in water. What you feel when your hand is not submerged in the water is the **normal pressure** p_0 of one atmosphere, which serves as a reference point. The bulk stress is this increase in pressure, or Δp , over the normal level, p_0 .

When the bulk stress increases, the bulk strain increases in response, in accordance with [\[link\]](#). The proportionality constant in this relation is called the bulk modulus, B , or

Note:

Equation:

$$B = \frac{\text{bulk stress}}{\text{bulk strain}} = -\frac{\Delta p}{\Delta V/V_0} = -\Delta p \frac{V_0}{\Delta V}.$$

The minus sign that appears in [\[link\]](#) is for consistency, to ensure that B is a positive quantity. Note that the minus sign ($-$) is necessary because an increase Δp in pressure (a positive quantity) always causes a decrease ΔV in volume, and decrease in volume is a negative quantity. The reciprocal of the bulk modulus is called **compressibility** k , or

Equation:

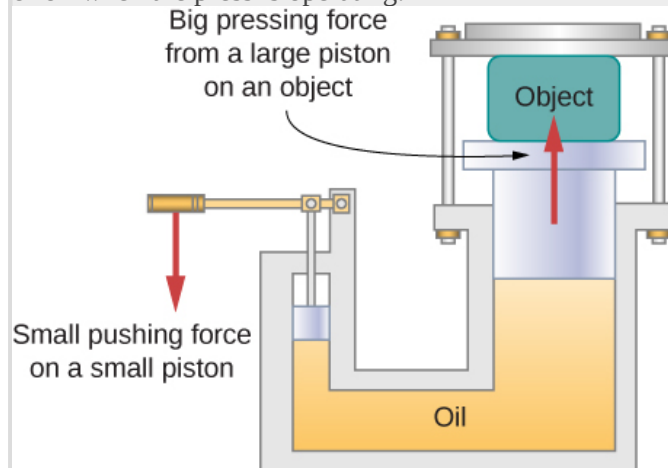
$$k = \frac{1}{B} = -\frac{\Delta V/V_0}{\Delta p}.$$

The term ‘compressibility’ is used in relation to fluids (gases and liquids). Compressibility describes the change in the volume of a fluid per unit increase in pressure. Fluids characterized by a large compressibility are relatively easy to compress. For example, the compressibility of water is $4.64 \times 10^{-5}/\text{atm}$ and the compressibility of acetone is $1.45 \times 10^{-4}/\text{atm}$. This means that under a 1.0-atm increase in pressure, the relative decrease in volume is approximately three times as large for acetone as it is for water.

Example:

Hydraulic Press

In a hydraulic press [\[link\]](#), a 250-liter volume of oil is subjected to a 2300-psi pressure increase. If the compressibility of oil is $2.0 \times 10^{-5}/\text{atm}$, find the bulk strain and the absolute decrease in the volume of oil when the press is operating.



In a hydraulic press, when a small piston is displaced downward, the pressure in the oil is transmitted throughout the oil to the large piston, causing the large piston to move upward. A small force applied to a small piston causes a large pressing force, which the large piston exerts on an

object that is either lifted or squeezed. The device acts as a mechanical lever.

Strategy

We must invert [\[link\]](#) to find the bulk strain. First, we convert the pressure increase from psi to atm, $\Delta p = 2300 \text{ psi} = 2300 / 14.7 \text{ atm} \approx 160 \text{ atm}$, and identify $V_0 = 250 \text{ L}$.

Solution

Substituting values into the equation, we have

Equation:

$$\text{bulk strain} = \frac{\Delta V}{V_0} = \frac{\Delta p}{B} = k\Delta p = (2.0 \times 10^{-5}/\text{atm})(160 \text{ atm}) = 0.0032$$
$$\text{answer: } \Delta V = 0.0032 V_0 = 0.0032(250 \text{ L}) = 0.78 \text{ L}.$$

Significance

Notice that since the compressibility of water is 2.32 times larger than that of oil, if the working substance in the hydraulic press of this problem were changed to water, the bulk strain as well as the volume change would be 2.32 times larger.

Note:**Exercise:****Problem:**

Check Your Understanding If the normal force acting on each face of a cubical 1.0-m^3 piece of steel is changed by $1.0 \times 10^7 \text{ N}$, find the resulting change in the volume of the piece of steel.

Solution:

63 mL

Shear Stress, Strain, and Modulus

The concepts of shear stress and strain concern only solid objects or materials. Buildings and tectonic plates are examples of objects that may be subjected to shear stresses. In general, these concepts do not apply to fluids.

Shear deformation occurs when two antiparallel forces of equal magnitude are applied tangentially to opposite surfaces of a solid object, causing no deformation in the transverse direction to the line of force, as in the typical example of shear stress illustrated in [\[link\]](#). Shear deformation is characterized by a gradual shift Δx of layers in the direction tangent to the acting forces. This gradation in Δx occurs in the transverse direction along some distance L_0 . Shear strain is defined by the ratio of the largest displacement Δx to the transverse distance L_0

Equation:

$$\text{shear strain} = \frac{\Delta x}{L_0}.$$

Shear strain is caused by shear stress. Shear stress is due to forces that act *parallel* to the surface. We use the symbol F_{\parallel} for such forces. The magnitude F_{\parallel} per surface area A where shearing force is applied is the measure of shear stress

Equation:

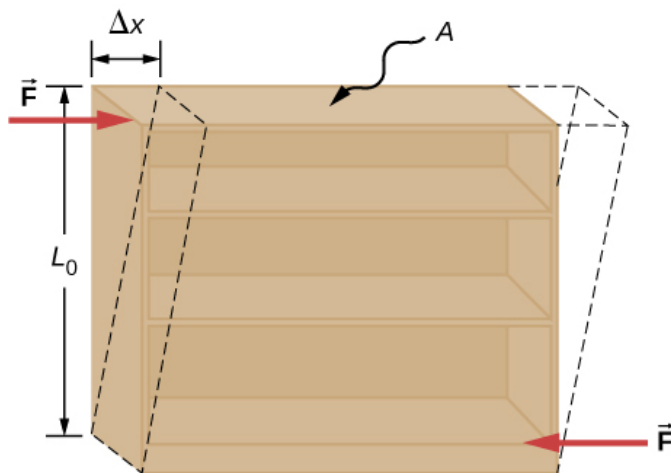
$$\text{shear stress} = \frac{F_{\parallel}}{A}.$$

The shear modulus is the proportionality constant in [\[link\]](#) and is defined by the ratio of stress to strain. Shear modulus is commonly denoted by S :

Note:

Equation:

$$S = \frac{\text{shear stress}}{\text{shear strain}} = \frac{F_{\parallel}/A}{\Delta x/L_0} = \frac{F_{\parallel}}{A} \frac{L_0}{\Delta x}.$$



An object under shear stress: Two antiparallel forces of equal magnitude are applied tangentially to opposite parallel surfaces of the object. The dashed-line contour depicts the resulting deformation. There is no change in the direction transverse to the acting forces and the transverse length L_0 is unaffected. Shear deformation is characterized by a gradual shift Δx of layers in the direction tangent to the forces.

Example:**An Old Bookshelf**

A cleaning person tries to move a heavy, old bookcase on a carpeted floor by pushing tangentially on the surface of the very top shelf. However, the only noticeable effect of this effort is similar to that seen in [\[link\]](#), and it disappears when the person stops pushing. The bookcase is 180.0 cm tall and 90.0 cm wide with four 30.0-cm-deep shelves, all partially loaded with books. The total weight of the bookcase and books is 600.0 N. If the person gives the top shelf a 50.0-N push that displaces the top shelf horizontally by 15.0 cm relative to the motionless bottom shelf, find the shear modulus of the bookcase.

Strategy

The only pieces of relevant information are the physical dimensions of the bookcase, the value of the tangential force, and the displacement this force causes. We identify $F_{\parallel} = 50.0 \text{ N}$, $\Delta x = 15.0 \text{ cm}$, $L_0 = 180.0 \text{ cm}$, and $A = (30.0 \text{ cm})(90.0 \text{ cm}) = 2700.0 \text{ cm}^2$, and we use [\[link\]](#) to compute the shear modulus.

Solution

Substituting numbers into the equations, we obtain for the shear modulus

Equation:

$$S = \frac{F_{\parallel}}{A} \frac{L_0}{\Delta x} = \frac{50.0 \text{ N}}{2700.0 \text{ cm}^2} \frac{180.0 \text{ cm}}{15.0 \text{ cm}} = \frac{2}{9} \frac{\text{N}}{\text{cm}^2} = \frac{2}{9} \times 10^4 \frac{\text{N}}{\text{m}^2} = \frac{20}{9} \times 10^3 \text{ Pa} = 2.222 \text{ kPa}.$$

We can also find shear stress and strain, respectively:

Equation:

$$\frac{F_{\parallel}}{A} = \frac{50.0 \text{ N}}{2700.0 \text{ cm}^2} = \frac{5}{27} \text{ kPa} = 185.2 \text{ Pa}$$

$$\frac{\Delta x}{L_0} = \frac{15.0 \text{ cm}}{180.0 \text{ cm}} = \frac{1}{12} = 0.083.$$

Significance

If the person in this example gave the shelf a healthy push, it might happen that the induced shear would collapse it to a pile of rubbish. Much the same shear mechanism is responsible for failures of earth-filled dams and levees; and, in general, for landslides.

Note:**Exercise:****Problem:**

Check Your Understanding Explain why the concepts of Young's modulus and shear modulus do not apply to fluids.

Solution:

Fluids have different mechanical properties than those of solids; fluids flow.

Summary

- External forces on an object (or medium) cause its deformation, which is a change in its size and shape. The strength of the forces that cause deformation is expressed by stress, which in SI units is

measured in the unit of pressure (pascal). The extent of deformation under stress is expressed by strain, which is dimensionless.

- For a small stress, the relation between stress and strain is linear. The elastic modulus is the proportionality constant in this linear relation.
- Tensile (or compressive) strain is the response of an object or medium to tensile (or compressive) stress. Here, the elastic modulus is called Young's modulus. Tensile (or compressive) stress causes elongation (or shortening) of the object or medium and is due to an external forces acting along only one direction perpendicular to the cross-section.
- Bulk strain is the response of an object or medium to bulk stress. Here, the elastic modulus is called the bulk modulus. Bulk stress causes a change in the volume of the object or medium and is caused by forces acting on the body from all directions, perpendicular to its surface. Compressibility of an object or medium is the reciprocal of its bulk modulus.
- Shear strain is the deformation of an object or medium under shear stress. The shear modulus is the elastic modulus in this case. Shear stress is caused by forces acting along the object's two parallel surfaces.

Conceptual Questions

Note: Unless stated otherwise, the weights of the wires, rods, and other elements are assumed to be negligible. Elastic moduli of selected materials are given in [\[link\]](#).

Exercise:

Problem:

Why can a squirrel jump from a tree branch to the ground and run away undamaged, while a human could break a bone in such a fall?

Solution:

In contact with the ground, stress in squirrel's limbs is smaller than stress in human's limbs.

Exercise:

Problem:

When a glass bottle full of vinegar warms up, both the vinegar and the glass expand, but the vinegar expands significantly more with temperature than does the glass. The bottle will break if it is filled up to its very tight cap. Explain why and how a pocket of air above the vinegar prevents the bottle from breaking.

Exercise:

Problem:

A thin wire strung between two nails in the wall is used to support a large picture. Is the wire likely to snap if it is strung tightly or if it is strung so that it sags considerably?

Solution:

tightly

Exercise:

Problem:

Review the relationship between stress and strain. Can you find any similarities between the two quantities?

Exercise:**Problem:**

What type of stress are you applying when you press on the ends of a wooden rod? When you pull on its ends?

Solution:

compressive; tensile

Exercise:

Problem: Can compressive stress be applied to a rubber band?

Exercise:

Problem: Can Young's modulus have a negative value? What about the bulk modulus?

Solution:

no

Exercise:**Problem:**

If a hypothetical material has a negative bulk modulus, what happens when you squeeze a piece of it?

Exercise:

Problem: Discuss how you might measure the bulk modulus of a liquid.

Problems**Exercise:****Problem:**

The "lead" in pencils is a graphite composition with a Young's modulus of approximately $1.0 \times 10^9 \text{ N/m}^2$. Calculate the change in length of the lead in an automatic pencil if you tap it straight into the pencil with a force of 4.0 N. The lead is 0.50 mm in diameter and 60 mm long.

Solution:

1.2 mm

Exercise:

Problem:

TV broadcast antennas are the tallest artificial structures on Earth. In 1987, a 72.0-kg physicist placed himself and 400 kg of equipment at the top of a 610-m-high antenna to perform gravity experiments. By how much was the antenna compressed, if we consider it to be equivalent to a steel cylinder 0.150 m in radius?

Exercise:**Problem:**

By how much does a 65.0-kg mountain climber stretch her 0.800-cm diameter nylon rope when she hangs 35.0 m below a rock outcropping? (For nylon, $Y = 1.35 \times 10^9 \text{Pa}$.)

Solution:

9.0 cm

Exercise:**Problem:**

When water freezes, its volume increases by 9.05%. What force per unit area is water capable of exerting on a container when it freezes?

Exercise:**Problem:**

A farmer making grape juice fills a glass bottle to the brim and caps it tightly. The juice expands more than the glass when it warms up, in such a way that the volume increases by 0.2%. Calculate the force exerted by the juice per square centimeter if its bulk modulus is $1.8 \times 10^9 \text{N/m}^2$, assuming the bottle does not break.

Solution:

$4.0 \times 10^2 \text{N/cm}^2$

Exercise:**Problem:**

A disk between vertebrae in the spine is subjected to a shearing force of 600.0 N. Find its shear deformation, using the shear modulus of $1.0 \times 10^9 \text{N/m}^2$. The disk is equivalent to a solid cylinder 0.700 cm high and 4.00 cm in diameter.

Exercise:**Problem:**

A vertebra is subjected to a shearing force of 500.0 N. Find the shear deformation, taking the vertebra to be a cylinder 3.00 cm high and 4.00 cm in diameter. How does your result compare with the result obtained in the preceding problem? Are spinal problems more common in disks than in vertebrae?

Solution:

0.149 μm

Exercise:

Problem:

Calculate the force a piano tuner applies to stretch a steel piano wire by 8.00 mm, if the wire is originally 1.35 m long and its diameter is 0.850 mm.

Exercise:

Problem:

A 20.0-m-tall hollow aluminum flagpole is equivalent in strength to a solid cylinder 4.00 cm in diameter. A strong wind bends the pole as much as a horizontal 900.0-N force on the top would do. How far to the side does the top of the pole flex?

Solution:

0.57 mm

Exercise:

Problem:

A copper wire of diameter 1.0 cm stretches 1.0% when it is used to lift a load upward with an acceleration of 2.0 m/s^2 . What is the weight of the load?

Exercise:

Problem:

As an oil well is drilled, each new section of drill pipe supports its own weight and the weight of the pipe and the drill bit beneath it. Calculate the stretch in a new 6.00-m-long steel pipe that supports a 100-kg drill bit and a 3.00-km length of pipe with a linear mass density of 20.0 kg/m. Treat the pipe as a solid cylinder with a 5.00-cm diameter.

Solution:

8.59 mm

Exercise:

Problem:

A large uniform cylindrical steel rod of density $\rho = 7.8 \text{ g/cm}^3$ is 2.0 m long and has a diameter of 5.0 cm. The rod is fastened to a concrete floor with its long axis vertical. What is the normal stress in the rod at the cross-section located at (a) 1.0 m from its lower end? (b) 1.5 m from the lower end?

Exercise:

Problem:

A 90-kg mountain climber hangs from a nylon rope and stretches it by 25.0 cm. If the rope was originally 30.0 m long and its diameter is 1.0 cm, what is Young's modulus for the nylon?

Solution:

$1.35 \times 10^9 \text{ Pa}$

Exercise:**Problem:**

A suspender rod of a suspension bridge is 25.0 m long. If the rod is made of steel, what must its diameter be so that it does not stretch more than 1.0 cm when a 2.5×10^4 -kg truck passes by it? Assume that the rod supports all of the weight of the truck.

Exercise:**Problem:**

A copper wire is 1.0 m long and its diameter is 1.0 mm. If the wire hangs vertically, how much weight must be added to its free end in order to stretch it 3.0 mm?

Solution:

259.0 N

Exercise:**Problem:**

A 100-N weight is attached to a free end of a metallic wire that hangs from the ceiling. When a second 100-N weight is added to the wire, it stretches 3.0 mm. The diameter and the length of the wire are 1.0 mm and 2.0 m, respectively. What is Young's modulus of the metal used to manufacture the wire?

Exercise:**Problem:**

The bulk modulus of a material is 1.0×10^{11} N/m². What fractional change in volume does a piece of this material undergo when it is subjected to a bulk stress increase of 10^7 N/m²? Assume that the force is applied uniformly over the surface.

Solution:

0.01%

Exercise:**Problem:**

Normal forces of magnitude 1.0×10^6 N are applied uniformly to a spherical surface enclosing a volume of a liquid. This causes the radius of the surface to decrease from 50.000 cm to 49.995 cm. What is the bulk modulus of the liquid?

Exercise:**Problem:**

During a walk on a rope, a tightrope walker creates a tension of 3.94×10^3 N in a wire that is stretched between two supporting poles that are 15.0 m apart. The wire has a diameter of 0.50 cm when it is not stretched. When the walker is on the wire in the middle between the poles the wire makes an angle of 5.0° below the horizontal. How much does this tension stretch the steel wire when the walker is in this position?

Solution:

1.44 cm

Exercise:**Problem:**

When using a pencil eraser, you exert a vertical force of 6.00 N at a distance of 2.00 cm from the hardwood-eraser joint. The pencil is 6.00 mm in diameter and is held at an angle of 20.0° to the horizontal. (a) By how much does the wood flex perpendicular to its length? (b) How much is it compressed lengthwise?

Exercise:**Problem:**

Normal forces are applied uniformly over the surface of a spherical volume of water whose radius is 20.0 cm. If the pressure on the surface is increased by 200 MPa, by how much does the radius of the sphere decrease?

Solution:

0.63 cm

Glossary

bulk modulus

elastic modulus for the bulk stress

bulk strain

(or **volume strain**) strain under the bulk stress, given as fractional change in volume

bulk stress

(or **volume stress**) stress caused by compressive forces, in all directions

compressibility

reciprocal of the bulk modulus

compressive strain

strain that occurs when forces are contracting an object, causing its shortening

compressive stress

stress caused by compressive forces, only in one direction

elastic modulus

proportionality constant in linear relation between stress and strain, in SI pascals

normal pressure

pressure of one atmosphere, serves as a reference level for pressure

pascal (Pa)

SI unit of stress, SI unit of pressure

pressure

force pressing in normal direction on a surface per the surface area, the bulk stress in fluids

shear modulus

elastic modulus for shear stress

shear strain

strain caused by shear stress

shear stress

stress caused by shearing forces

strain

dimensionless quantity that gives the amount of deformation of an object or medium under stress

stress

quantity that contains information about the magnitude of force causing deformation, defined as force per unit area

tensile strain

strain under tensile stress, given as fractional change in length, which occurs when forces are stretching an object, causing its elongation

tensile stress

stress caused by tensile forces, only in one direction, which occurs when forces are stretching an object, causing its elongation

Young's modulus

elastic modulus for tensile or compressive stress

Elasticity and Plasticity

By the end of this section, you will be able to:

- Explain the limit where a deformation of material is elastic
- Describe the range where materials show plastic behavior
- Analyze elasticity and plasticity on a stress-strain diagram

We referred to the proportionality constant between stress and strain as the elastic modulus. But why do we call it that? What does it mean for an object to be elastic and how do we describe its behavior?

Elasticity is the tendency of solid objects and materials to return to their original shape after the external forces (load) causing a deformation are removed. An object is **elastic** when it comes back to its original size and shape when the load is no longer present. Physical reasons for elastic behavior vary among materials and depend on the microscopic structure of the material. For example, the elasticity of polymers and rubbers is caused by stretching polymer chains under an applied force. In contrast, the elasticity of metals is caused by resizing and reshaping the crystalline cells of the lattices (which are the material structures of metals) under the action of externally applied forces.

The two parameters that determine the elasticity of a material are its *elastic modulus* and its *elastic limit*. A high elastic modulus is typical for materials that are hard to deform; in other words, materials that require a high load to achieve a significant strain. An example is a steel band. A low elastic modulus is typical for materials that are easily deformed under a load; for example, a rubber band. If the stress under a load becomes too high, then when the load is removed, the material no longer comes back to its original shape and size, but relaxes to a different shape and size: The material becomes permanently deformed. The **elastic limit** is the stress value beyond which the material no longer behaves elastically but becomes permanently deformed.

Our perception of an elastic material depends on both its elastic limit and its elastic modulus. For example, all rubbers are characterized by a low elastic modulus and a high elastic limit; hence, it is easy to stretch them and the

stretch is noticeably large. Among materials with identical elastic limits, the most elastic is the one with the lowest elastic modulus.

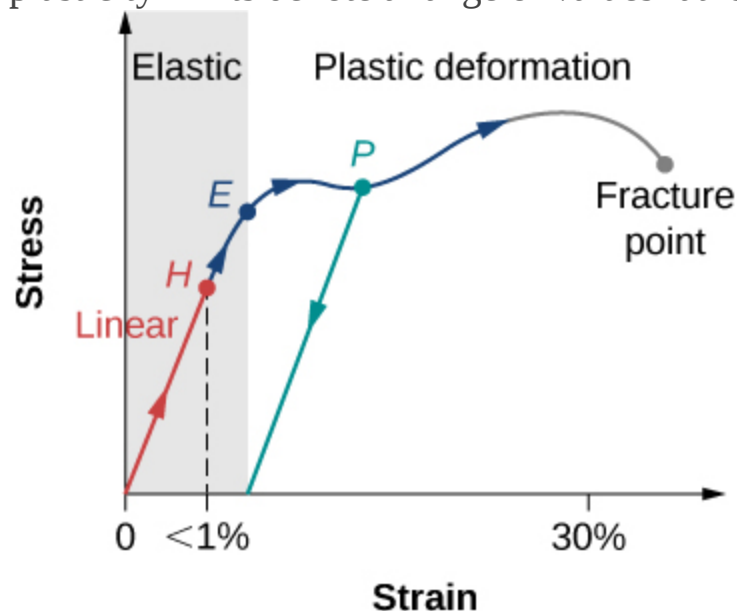
When the load increases from zero, the resulting stress is in direct proportion to strain in the way given by [\[link\]](#), but only when stress does not exceed some limiting value. For stress values within this linear limit, we can describe elastic behavior in analogy with Hooke's law for a spring. According to Hooke's law, the stretch value of a spring under an applied force is directly proportional to the magnitude of the force. Conversely, the response force from the spring to an applied stretch is directly proportional to the stretch. In the same way, the deformation of a material under a load is directly proportional to the load, and, conversely, the resulting stress is directly proportional to strain. The linearity limit (or the **proportionality limit**) is the largest stress value beyond which stress is no longer proportional to strain. Beyond the linearity limit, the relation between stress and strain is no longer linear. When stress becomes larger than the linearity limit but still within the elasticity limit, behavior is still elastic, but the relation between stress and strain becomes nonlinear.

For stresses beyond the elastic limit, a material exhibits **plastic behavior**. This means the material deforms irreversibly and does not return to its original shape and size, even when the load is removed. When stress is gradually increased beyond the elastic limit, the material undergoes plastic deformation. Rubber-like materials show an increase in stress with the increasing strain, which means they become more difficult to stretch and, eventually, they reach a fracture point where they break. Ductile materials such as metals show a gradual decrease in stress with the increasing strain, which means they become easier to deform as stress-strain values approach the breaking point. Microscopic mechanisms responsible for plasticity of materials are different for different materials.

We can graph the relationship between stress and strain on a **stress-strain diagram**. Each material has its own characteristic strain-stress curve. A typical stress-strain diagram for a ductile metal under a load is shown in [\[link\]](#). In this figure, strain is a fractional elongation (not drawn to scale). When the load is gradually increased, the linear behavior (red line) that starts at the no-load point (the origin) ends at the linearity limit at point *H*.

For further load increases beyond point H , the stress-strain relation is nonlinear but still elastic. In the figure, this nonlinear region is seen between points H and E . Ever larger loads take the stress to the elasticity limit E , where elastic behavior ends and plastic deformation begins. Beyond the elasticity limit, when the load is removed, for example at P , the material relaxes to a new shape and size along the green line. This is to say that the material becomes permanently deformed and does not come back to its initial shape and size when stress becomes zero.

The material undergoes plastic deformation for loads large enough to cause stress to go beyond the elasticity limit at E . The material continues to be plastically deformed until the stress reaches the fracture point (breaking point). Beyond the fracture point, we no longer have one sample of material, so the diagram ends at the fracture point. For the completeness of this qualitative description, it should be said that the linear, elastic, and plasticity limits denote a range of values rather than one sharp point.



Typical stress-strain plot for a metal under a load: The graph ends at the fracture point. The arrows show the direction of changes under an ever-increasing load. Points H and E are the linearity and elasticity limits, respectively. Between points H and E ,

the behavior is nonlinear. The green line originating at P illustrates the metal's response when the load is removed. The permanent deformation has a strain value at the point where the green line intercepts the horizontal axis.

The value of stress at the fracture point is called breaking stress (or **ultimate stress**). Materials with similar elastic properties, such as two metals, may have very different breaking stresses. For example, ultimate stress for aluminum is $2.2 \times 10^8 \text{ Pa}$ and for steel it may be as high as $20.0 \times 10^8 \text{ Pa}$, depending on the kind of steel. We can make a quick estimate, based on [\[link\]](#), that for rods with a 1-in^2 cross-sectional area, the breaking load for an aluminum rod is $3.2 \times 10^4 \text{ lb}$, and the breaking load for a steel rod is about nine times larger.

Summary

- An object or material is elastic if it comes back to its original shape and size when the stress vanishes. In elastic deformations with stress values lower than the proportionality limit, stress is proportional to strain. When stress goes beyond the proportionality limit, the deformation is still elastic but nonlinear up to the elasticity limit.
- An object or material has plastic behavior when stress is larger than the elastic limit. In the plastic region, the object or material does not come back to its original size or shape when stress vanishes but acquires a permanent deformation. Plastic behavior ends at the breaking point.

Key Equations

First Equilibrium Condition	$\sum_k \vec{\mathbf{F}}_k = \vec{\mathbf{0}}$
Second Equilibrium Condition	$\sum_k \vec{\tau}_k = \vec{\mathbf{0}}$
Linear relation between stress and strain	stress = (elastic modulus) \times strain
Young's modulus	$Y = \frac{\text{tensile stress}}{\text{tensile strain}} = \frac{F_{\perp}}{A} \frac{L_0}{\Delta L}$
Bulk modulus	$B = \frac{\text{bulk stress}}{\text{bulk strain}} = -\Delta p \frac{V_0}{\Delta V}$
Shear modulus	$S = \frac{\text{shear stress}}{\text{shear strain}} = \frac{F_{\parallel}}{A} \frac{L_0}{\Delta x}$

Conceptual Questions

Note: Unless stated otherwise, the weights of the wires, rods, and other elements are assumed to be negligible. Elastic moduli of selected materials are given in [\[link\]](#).

Exercise:

Problem:

What is meant when a fishing line is designated as “a 10-lb test?”

Exercise:

Problem:

Steel rods are commonly placed in concrete before it sets. What is the purpose of these rods?

Solution:

It acts as “reinforcement,” increasing a range of strain values before the structure reaches its breaking point.

Problems

Exercise:

Problem:

A uniform rope of cross-sectional area 0.50 cm^2 breaks when the tensile stress in it reaches $6.00 \times 10^6 \text{ N/m}^2$. (a) What is the maximum load that can be lifted slowly at a constant speed by the rope? (b) What is the maximum load that can be lifted by the rope with an acceleration of 4.00 m/s^2 ?

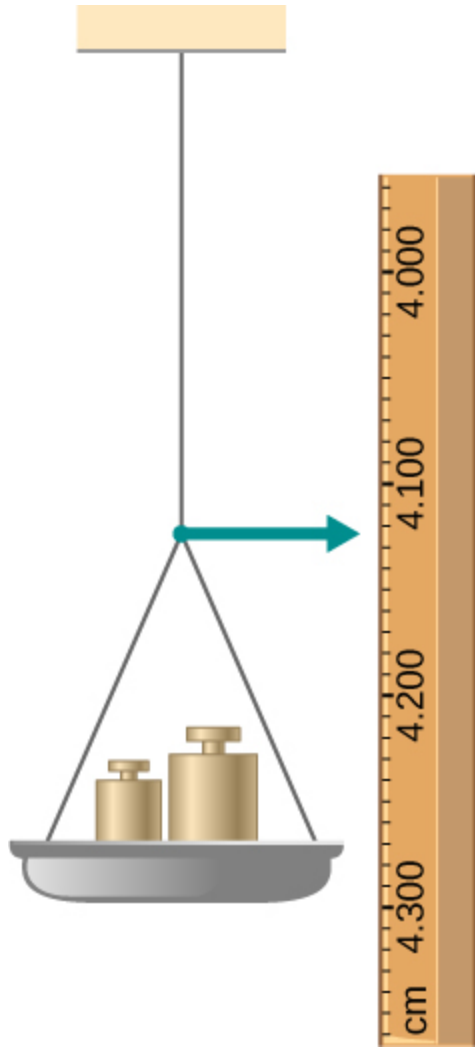
Exercise:

Problem:

One end of a vertical metallic wire of length 2.0 m and diameter 1.0 mm is attached to a ceiling, and the other end is attached to a 5.0-N weight pan, as shown below. The position of the pointer before the pan is 4.000 cm . Different weights are then added to the pan area, and the position of the pointer is recorded in the table shown. Plot stress versus strain for this wire, then use the resulting curve to determine Young's modulus and the proportionality limit of the metal. What metal is this most likely to be?

Added load (including pan) (N)	Scale reading (cm)
0	4.000
15	4.036

Added load (including pan) (N)	Scale reading (cm)
25	4.073
35	4.109
45	4.146
55	4.181
65	4.221
75	4.266
85	4.316



Exercise:

Problem:

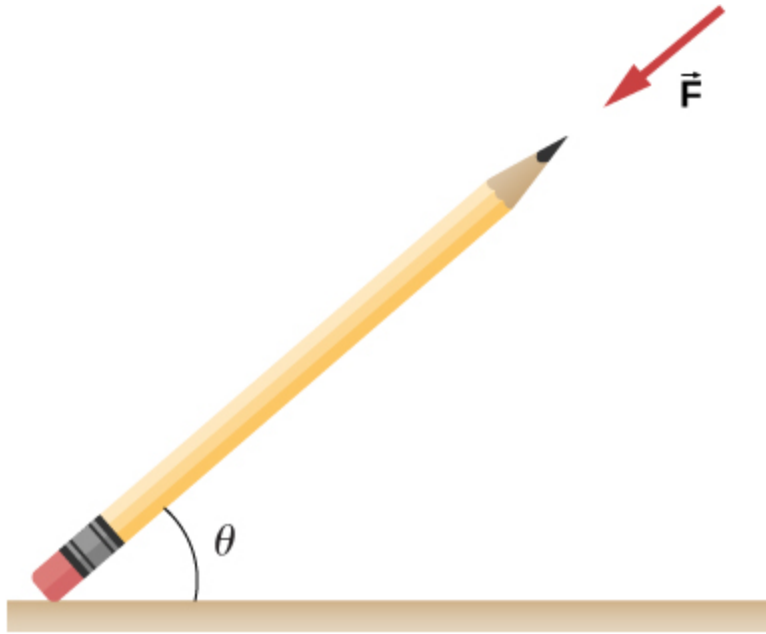
An aluminum ($\rho = 2.7 \text{ g/cm}^3$) wire is suspended from the ceiling and hangs vertically. How long must the wire be before the stress at its upper end reaches the proportionality limit, which is $8.0 \times 10^7 \text{ N/m}^2$?

Additional Problems

Exercise:

Problem:

The coefficient of static friction between the rubber eraser of the pencil and the tabletop is $\mu_s = 0.80$. If the force \vec{F} is applied along the axis of the pencil, as shown below, what is the minimum angle at which the pencil can stand without slipping? Ignore the weight of the pencil.

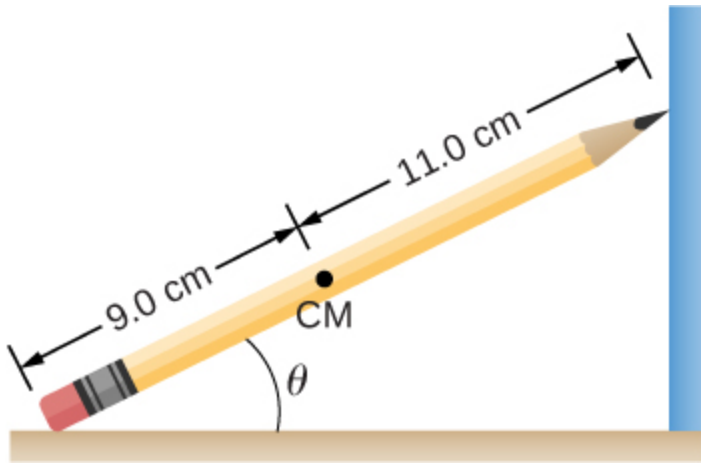


Solution:

$$\tan^{-1}(1/\mu_s) = 51.3^\circ$$

Exercise:**Problem:**

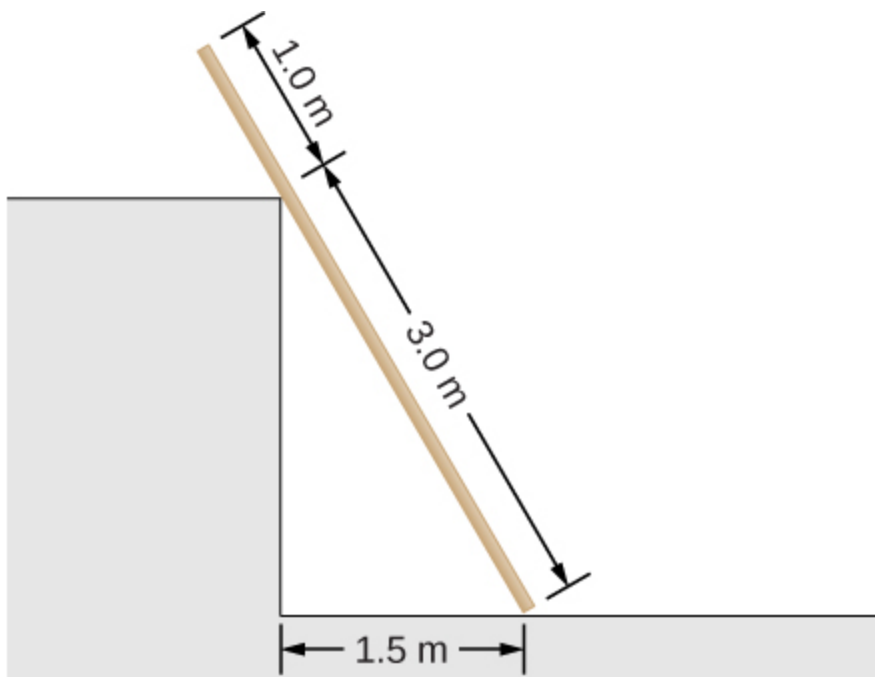
A pencil rests against a corner, as shown below. The sharpened end of the pencil touches a smooth vertical surface and the eraser end touches a rough horizontal floor. The coefficient of static friction between the eraser and the floor is $\mu_s = 0.80$. The center of mass of the pencil is located 9.0 cm from the tip of the eraser and 11.0 cm from the tip of the pencil lead. Find the minimum angle θ for which the pencil does not slip.



Exercise:

Problem:

A uniform 4.0-m plank weighing 200.0 N rests against the corner of a wall, as shown below. There is no friction at the point where the plank meets the corner. (a) Find the forces that the corner and the floor exert on the plank. (b) What is the minimum coefficient of static friction between the floor and the plank to prevent the plank from slipping?



Solution:

a. at corner 66.7 N at 30° with the horizontal; at floor 177 N at 109° with the horizontal; b. $\mu_s = 0.346$

Exercise:**Problem:**

A 40-kg boy jumps from a height of 3.0 m, lands on one foot and comes to rest in 0.10 s after he hits the ground. Assume that he comes to rest with a constant acceleration opposite to the motion. If the total cross-sectional area of the bones in his legs just above his ankles is 3.0 cm^2 , what is the compression stress in these bones? Leg bones can be fractured when they are subjected to stress greater than $1.7 \times 10^8 \text{ Pa}$. Is the boy in danger of breaking his leg?

Exercise:**Problem:**

Two thin rods, one made of steel and the other of aluminum, are joined end to end. Each rod is 2.0 m long and has cross-sectional area 9.1 mm^2 . If a 10,000-N tensile force is applied at each end of the combination, find: (a) stress in each rod; (b) strain in each rod; and, (c) elongation of each rod.

Solution:

a. $1.10 \times 10^9 \text{ N/m}^2$; b. 5.5×10^{-3} ; c. 11.0 mm, 31.4 mm

Exercise:**Problem:**

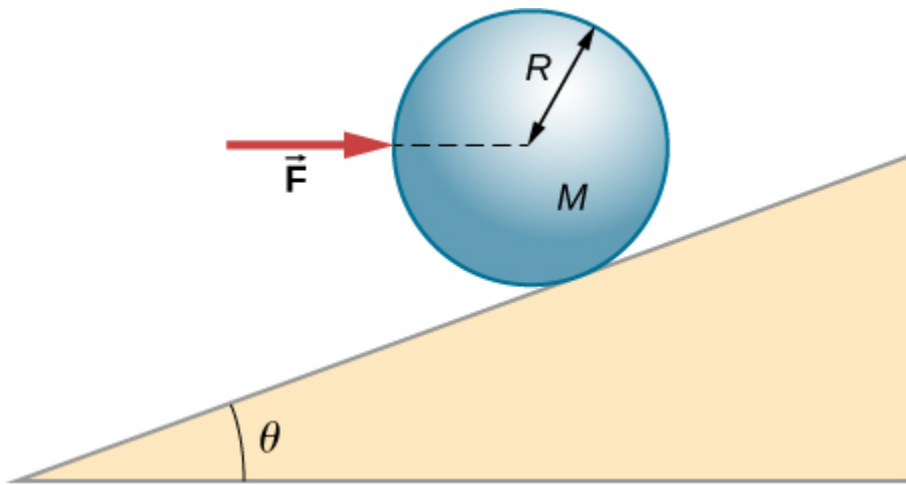
Two rods, one made of copper and the other of steel, have the same dimensions. If the copper rod stretches by 0.15 mm under some stress, how much does the steel rod stretch under the same stress?

Challenge Problems

Exercise:

Problem:

A horizontal force \vec{F} is applied to a uniform sphere in direction exact toward the center of the sphere, as shown below. Find the magnitude of this force so that the sphere remains in static equilibrium. What is the frictional force of the incline on the sphere?



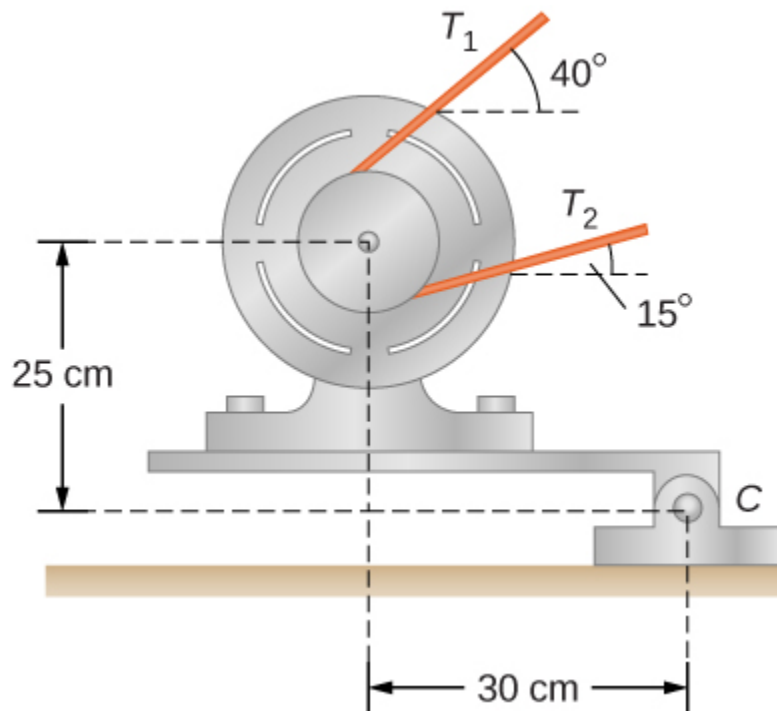
Solution:

$$F = Mg \tan \theta; f = 0$$

Exercise:

Problem:

When a motor is set on a pivoted mount seen below, its weight can be used to maintain tension in the drive belt. When the motor is not running the tensions T_1 and T_2 are equal. The total mass of the platform and the motor is 100.0 kg, and the diameter of the drive belt pulley is 16.0 cm. when the motor is off, find: (a) the tension in the belt, and (b) the force at the hinged platform support at point C . Assume that the center of mass of the motor plus platform is at the center of the motor.

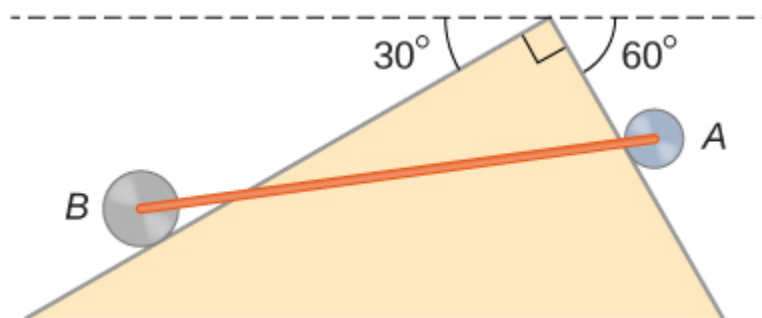


Exercise:

Problem:

Two wheels A and B with weights w and $2w$, respectively, are connected by a uniform rod with weight $w/2$, as shown below. The wheels are free to roll on the sloped surfaces. Determine the angle that the rod forms with the horizontal when the system is in equilibrium.

Hint: There are five forces acting on the rod, which is two weights of the wheels, two normal reaction forces at points where the wheels make contacts with the wedge, and the weight of the rod.



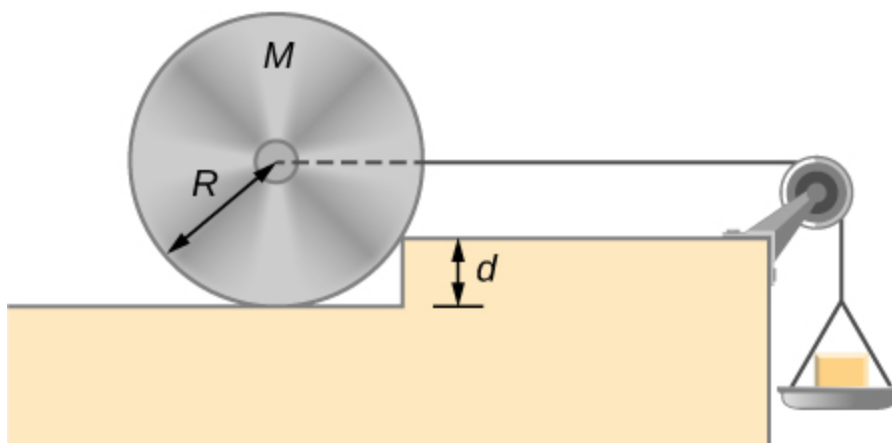
Solution:

with the horizontal, $\theta = 42.2^\circ$; $\alpha = 17.8^\circ$ with the steeper side of the wedge

Exercise:

Problem:

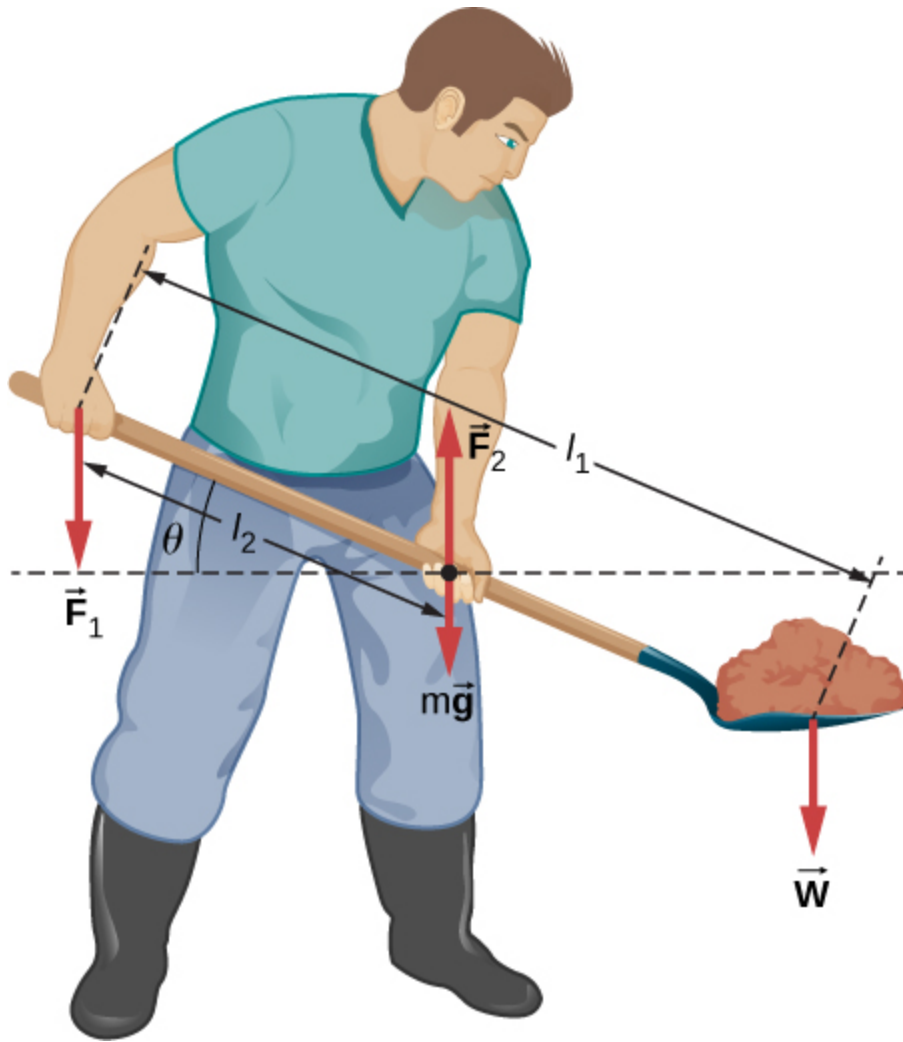
Weights are gradually added to a pan until a wheel of mass M and radius R is pulled over an obstacle of height d , as shown below. What is the minimum mass of the weights plus the pan needed to accomplish this?



Exercise:

Problem:

In order to lift a shovelful of dirt, a gardener pushes downward on the end of the shovel and pulls upward at distance l_2 from the end, as shown below. The weight of the shovel is $m\vec{g}$ and acts at the point of application of \vec{F}_2 . Calculate the magnitudes of the forces \vec{F}_1 and \vec{F}_2 as functions of l_1 , l_2 , mg , and the weight W of the load. Why do your answers not depend on the angle θ that the shovel makes with the horizontal?



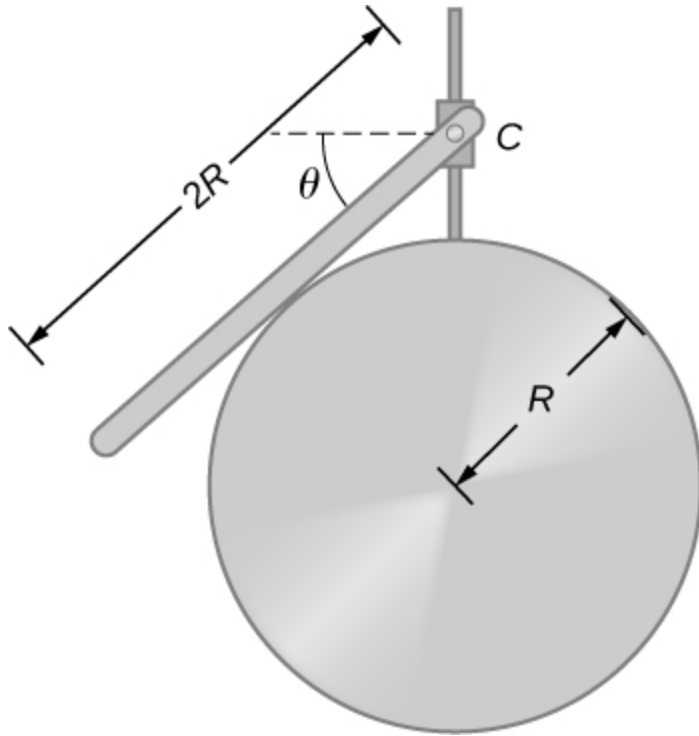
Solution:

$$W(l_1/l_2 - 1); Wl_1/l_2 + mg$$

Exercise:

Problem:

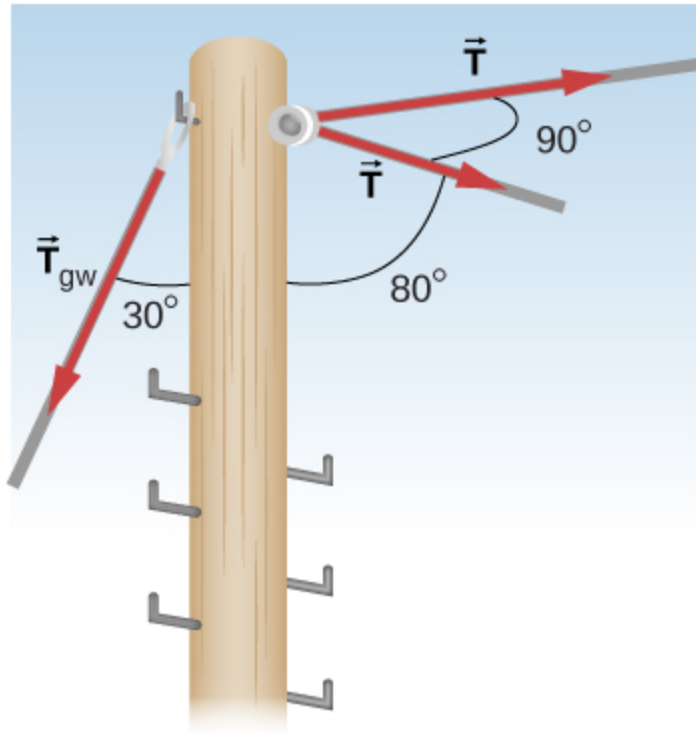
A uniform rod of length $2R$ and mass M is attached to a small collar C and rests on a cylindrical surface of radius R , as shown below. If the collar can slide without friction along the vertical guide, find the angle θ for which the rod is in static equilibrium.



Exercise:

Problem:

The pole shown below is at a 90.0° bend in a power line and is therefore subjected to more shear force than poles in straight parts of the line. The tension in each line is $4.00 \times 10^4 \text{ N}$, at the angles shown. The pole is 15.0 m tall, has an 18.0 cm diameter, and can be considered to have half the strength of hardwood. (a) Calculate the compression of the pole. (b) Find how much it bends and in what direction. (c) Find the tension in a guy wire used to keep the pole straight if it is attached to the top of the pole at an angle of 30.0° with the vertical. The guy wire is in the opposite direction of the bend.



Solution:

a. 1.1 mm; b. 6.6 mm to the right; c. $1.11 \times 10^5 \text{ N}$

Glossary

breaking stress (ultimate stress)
value of stress at the fracture point

elastic
object that comes back to its original size and shape when the load is no longer present

elastic limit
stress value beyond which material no longer behaves elastically and becomes permanently deformed

linearity limit (proportionality limit)

largest stress value beyond which stress is no longer proportional to strain

plastic behavior

material deforms irreversibly, does not go back to its original shape and size when load is removed and stress vanishes

stress-strain diagram

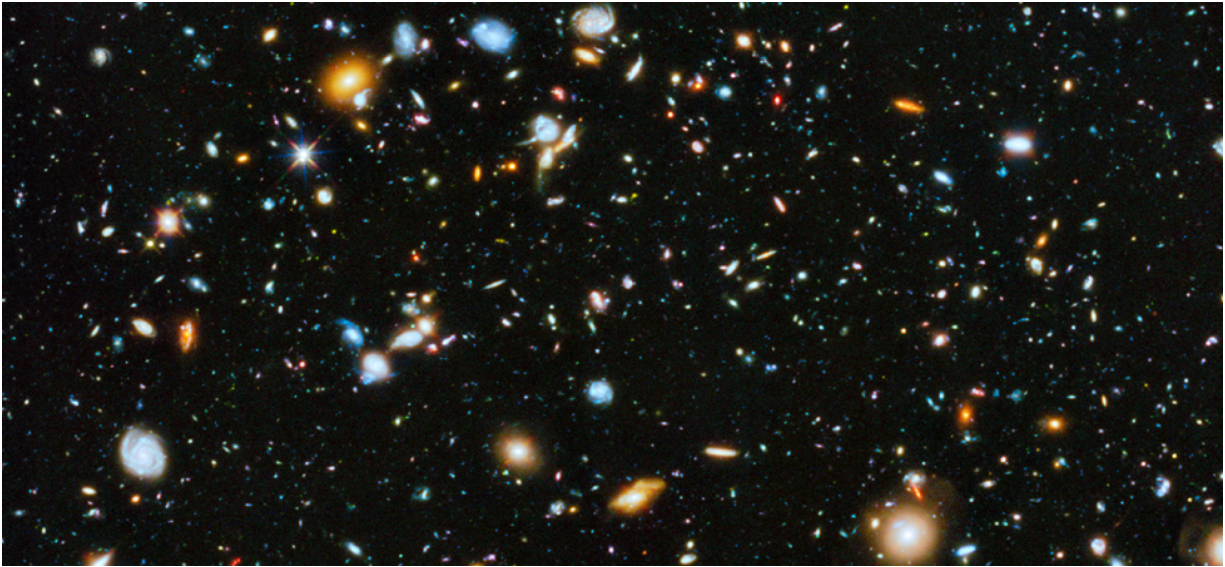
graph showing the relationship between stress and strain, characteristic of a material

Introduction

class="introduction"

Our visible
Universe
contains
billions of
galaxies,
whose very
existence is
due to the
force of
gravity.
Gravity is
ultimately
responsible
for the energy
output of all
stars—
initiating
thermonuclear
reactions in
stars,
allowing the
Sun to heat
Earth, and
making
galaxies
visible from
unfathomable
distances.
Most of the
dots you see
in this image
are not stars,
but galaxies.
(credit:

modification
of work by
NASA/ESA)



In this chapter, we study the nature of the gravitational force for objects as small as ourselves and for systems as massive as entire galaxies. We show how the gravitational force affects objects on Earth and the motion of the Universe itself. Gravity is the first force to be postulated as an action-at-a-distance force, that is, objects exert a gravitational force on one another without physical contact and that force falls to zero only at an infinite distance. Earth exerts a gravitational force on you, but so do our Sun, the Milky Way galaxy, and the billions of galaxies, like those shown above, which are so distant that we cannot see them with the naked eye.

Newton's Law of Universal Gravitation

By the end of this section, you will be able to:

- List the significant milestones in the history of gravitation
- Calculate the gravitational force between two point masses
- Estimate the gravitational force between collections of mass

We first review the history of the study of gravitation, with emphasis on those phenomena that for thousands of years have inspired philosophers and scientists to search for an explanation. Then we examine the simplest form of Newton's law of universal gravitation and how to apply it.

The History of Gravitation

The earliest philosophers wondered why objects naturally tend to fall toward the ground. Aristotle (384–322 BCE) believed that it was the nature of rocks to seek Earth and the nature of fire to seek the Heavens. Brahmagupta (598~665 CE) postulated that Earth was a sphere and that objects possessed a natural affinity for it, falling toward the center from wherever they were located.

The motions of the Sun, our Moon, and the planets have been studied for thousands of years as well. These motions were described with amazing accuracy by Ptolemy (90–168 CE), whose method of epicycles described the paths of the planets as circles within circles. However, there is little evidence that anyone connected the motion of astronomical bodies with the motion of objects falling to Earth—until the seventeenth century.

Nicolaus Copernicus (1473–1543) is generally credited as being the first to challenge Ptolemy's geocentric (Earth-centered) system and suggest a heliocentric system, in which the Sun is at the center of the solar system. This idea was supported by the incredibly precise naked-eye measurements of planetary motions by Tycho Brahe and their analysis by Johannes Kepler and Galileo Galilei. Kepler showed that the motion of each planet is an ellipse (the first of his three laws, discussed in [Kepler's Laws of Planetary Motion](#)), and Robert Hooke (the same Hooke who formulated Hooke's law for springs) intuitively suggested that these motions are due to the planets being attracted to the Sun. However, it was Isaac Newton who connected the acceleration of objects near Earth's surface with the centripetal acceleration of the Moon in its orbit about Earth.

Finally, in [Einstein's Theory of Gravity](#), we look at the theory of general relativity proposed by Albert Einstein in 1916. His theory comes from a vastly different perspective, in which gravity is a manifestation of mass warping space and time. The consequences of his theory gave rise to many remarkable predictions, essentially all of which have been confirmed over the many decades following the publication of the theory (including the 2015 measurement of gravitational waves from the merger of two black holes).

Newton's Law of Universal Gravitation

Newton noted that objects at Earth's surface (hence at a distance of R_E from the center of Earth) have an acceleration of g , but the Moon, at a distance of about $60 R_E$, has a centripetal acceleration about $(60)^2$ times smaller than g . He could explain this by postulating that a force exists between any two objects, whose magnitude is given by the product of the two masses divided by the square of the distance between them. We now know that this inverse square law is ubiquitous in nature, a function of geometry for point sources. The strength of any source at a distance r is spread over the surface of a

sphere centered about the mass. The surface area of that sphere is proportional to r^2 . In later chapters, we see this same form in the electromagnetic force.

Note:

Newton's Law of Gravitation

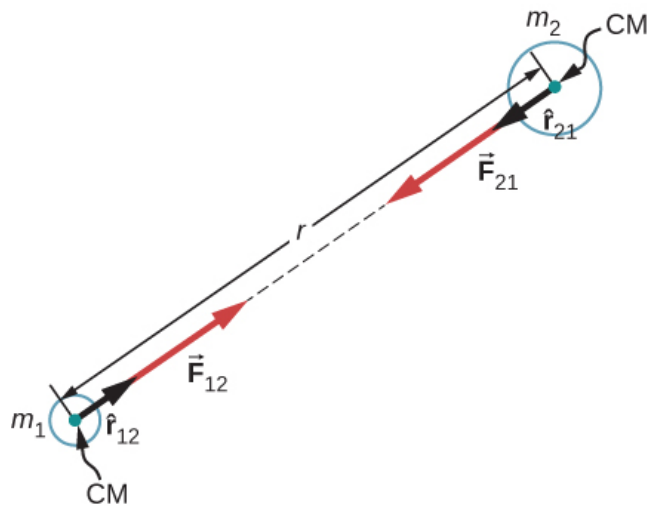
Newton's law of gravitation can be expressed as

Equation:

$$\vec{\mathbf{F}}_{12} = G \frac{m_1 m_2}{r^2} \hat{\mathbf{r}}_{12}$$

where $\vec{\mathbf{F}}_{12}$ is the force on object 1 exerted by object 2 and $\hat{\mathbf{r}}_{12}$ is a unit vector that points from object 1 toward object 2.

As shown in [\[link\]](#), the $\vec{\mathbf{F}}_{12}$ vector points from object 1 toward object 2, and hence represents an attractive force between the objects. The equal but opposite force $\vec{\mathbf{F}}_{21}$ is the force on object 2 exerted by object 1.



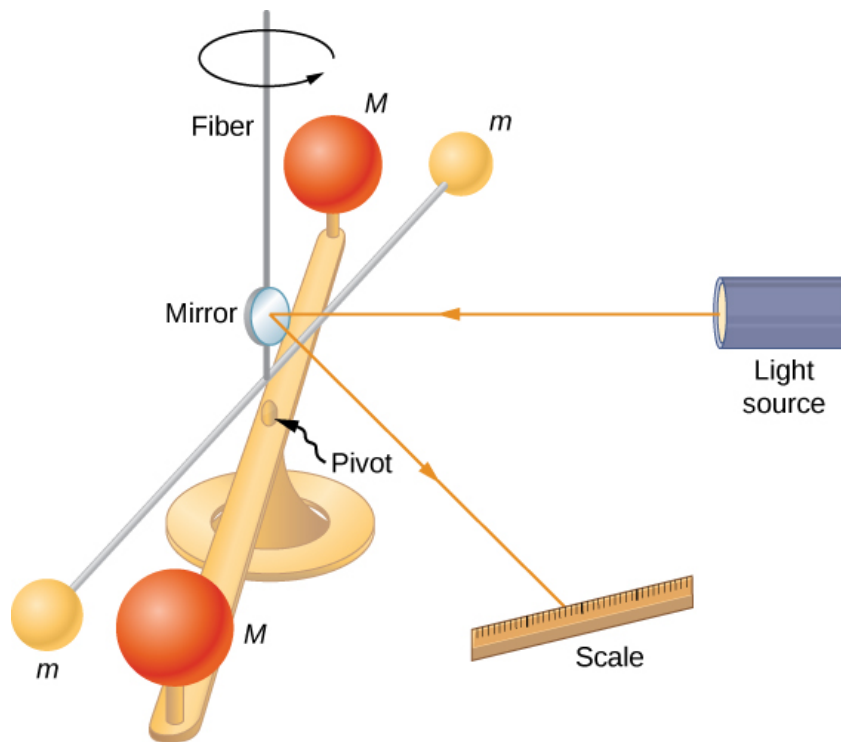
Gravitational force acts along a line joining the centers of mass of two objects.

These equal but opposite forces reflect Newton's third law, which we discussed earlier. Note that strictly speaking, [\[link\]](#) applies to point masses—all the mass is located at one point. But it applies equally to any spherically symmetric objects, where r is the distance between the centers of mass of those objects. In many cases, it works reasonably well for nonsymmetrical objects, if their separation is large compared to their size, and we take r to be the distance between the center of mass of each body.

The Cavendish Experiment

A century after Newton published his law of universal gravitation, Henry Cavendish determined the proportionality constant G by performing a painstaking experiment. He constructed a device similar to that shown in [\[link\]](#), in which small masses are suspended from a wire. Once in equilibrium, two fixed, larger masses are placed symmetrically near the smaller ones. The gravitational attraction creates a torsion (twisting) in the supporting wire that can be measured.

The constant G is called the **universal gravitational constant** and Cavendish determined it to be $G = 6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$. The word ‘universal’ indicates that scientists think that this constant applies to masses of any composition and that it is the same throughout the Universe. The value of G is an incredibly small number, showing that the force of gravity is very weak. The attraction between masses as small as our bodies, or even objects the size of skyscrapers, is incredibly small. For example, two 1.0-kg masses located 1.0 meter apart exert a force of $6.7 \times 10^{-11} \text{ N}$ on each other. This is the weight of a typical grain of pollen.



Cavendish used an apparatus similar to this to measure the gravitational attraction between two spheres (m) suspended from a wire and two stationary spheres (M). This is a common experiment performed in undergraduate laboratories, but it is quite challenging. Passing trucks outside the laboratory can create vibrations that overwhelm the gravitational forces.

Although gravity is the weakest of the four fundamental forces of nature, its attractive nature is what holds us to Earth, causes the planets to orbit the Sun and the Sun to orbit our galaxy, and binds galaxies into clusters, ranging from a few to millions. Gravity is the force that forms the Universe.

Note:

Newton's Law of Gravitation

To determine the motion caused by the gravitational force, follow these steps:

1. Identify the two masses, one or both, for which you wish to find the gravitational force.
2. Draw a free-body diagram, sketching the force acting on each mass and indicating the distance between their centers of mass.
3. Apply Newton's second law of motion to each mass to determine how it will move.

Example:

A Collision in Orbit

Consider two nearly spherical *Soyuz* payload vehicles, in orbit about Earth, each with mass 9000 kg and diameter 4.0 m. They are initially at rest relative to each other, 10.0 m from center to center. (As we will see in [Kepler's Laws of Planetary Motion](#), both orbit Earth at the same speed and interact nearly the same as if they were isolated in deep space.) Determine the gravitational force between them and their initial acceleration. Estimate how long it takes for them to drift together, and how fast they are moving upon impact.

Strategy

We use Newton's law of gravitation to determine the force between them and then use Newton's second law to find the acceleration of each. For the *estimate*, we assume this acceleration is constant, and we use the constant-acceleration equations from [Motion along a Straight Line](#) to find the time and speed of the collision.

Solution

The magnitude of the force is

Equation:

$$|\vec{F}_{12}| = F_{12} = G \frac{m_1 m_2}{r^2} = 6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2 / \text{kg}^2 \frac{(9000 \text{ kg})(9000 \text{ kg})}{(10 \text{ m})^2} = 5.4 \times 10^{-5} \text{ N}.$$

The initial acceleration of each payload is

Equation:

$$a = \frac{F}{m} = \frac{5.4 \times 10^{-5} \text{ N}}{9000 \text{ kg}} = 6.0 \times 10^{-9} \text{ m/s}^2.$$

The vehicles are 4.0 m in diameter, so the vehicles move from 10.0 m to 4.0 m apart, or a distance of 3.0 m each. A similar calculation to that above, for when the vehicles are 4.0 m apart, yields an acceleration of $3.8 \times 10^{-8} \text{ m/s}^2$, and the average of these two values is $2.2 \times 10^{-8} \text{ m/s}^2$. If we assume a constant acceleration of this value and they start from rest, then the vehicles collide with speed given by

Equation:

$$v^2 = v_0^2 + 2a(x - x_0), \text{ where } v_0 = 0,$$

so

Equation:

$$v = \sqrt{2(2.2 \times 10^{-9} \text{ N})(3.0 \text{ m})} = 3.6 \times 10^{-4} \text{ m/s}.$$

We use $v = v_0 + at$ to find $t = v/a = 1.7 \times 10^4 \text{ s}$ or about 4.6 hours.

Significance

These calculations—including the initial force—are only estimates, as the vehicles are probably not spherically symmetrical. But you can see that the force is incredibly small. Astronauts must tether themselves when doing work outside even the massive International Space Station (ISS), as in [\[link\]](#), because the gravitational attraction cannot save them from even the smallest push away from the station.



This photo shows Ed White tethered to the Space Shuttle during a spacewalk. (credit: NASA)

Note:

Exercise:

Problem:

Check Your Understanding What happens to force and acceleration as the vehicles fall together? What will our estimate of the velocity at a collision higher or lower than the speed actually be? And finally, what would happen if the masses were not identical? Would the force on each be the same or different? How about their accelerations?

Solution:

The force of gravity on each object increases with the square of the inverse distance as they fall together, and hence so does the acceleration. For example, if the distance is halved, the force and acceleration are quadrupled. Our average is accurate only for a linearly increasing acceleration, whereas the acceleration actually increases at a greater rate. So our calculated speed is too small. From Newton's third law (action-reaction forces), the force of gravity between any two objects must be the same. But the accelerations will not be if they have different masses.

The effect of gravity between two objects with masses on the order of these space vehicles is indeed small. Yet, the effect of gravity on you from Earth is significant enough that a fall into Earth of only a few feet can be dangerous. We examine the force of gravity near Earth's surface in the next section.

Example:**Attraction between Galaxies**

Find the acceleration of our galaxy, the Milky Way, due to the nearest comparably sized galaxy, the Andromeda galaxy ([link](#)). The approximate mass of each galaxy is 800 billion solar masses (a solar mass is the mass of our Sun), and they are separated by 2.5 million light-years. (Note that the mass of Andromeda is not so well known but is believed to be slightly larger than our galaxy.) Each galaxy has a diameter of roughly 100,000 light-years ($1 \text{ light-year} = 9.5 \times 10^{15} \text{ m}$).



Galaxies interact gravitationally over immense distances. The Andromeda galaxy is the nearest spiral galaxy to the Milky Way, and they will eventually collide. (credit: Boris Štromar)

Strategy

As in the preceding example, we use Newton's law of gravitation to determine the force between them and then use Newton's second law to find the acceleration of the Milky Way. We can consider the galaxies to be point masses, since their sizes are about 25 times smaller than their separation. The mass of the Sun (see [Appendix D](#)) is $2.0 \times 10^{30} \text{ kg}$ and a light-year is the distance light travels in one year, $9.5 \times 10^{15} \text{ m}$.

Solution

The magnitude of the force is

Equation:

$$F_{12} = G \frac{m_1 m_2}{r^2} = (6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2 / \text{kg}^2) \frac{[(800 \times 10^9)(2.0 \times 10^{30} \text{ kg})]^2}{[(2.5 \times 10^6)(9.5 \times 10^{15} \text{ m})]^2} = 3.0 \times 10^{29} \text{ N}.$$

The acceleration of the Milky Way is

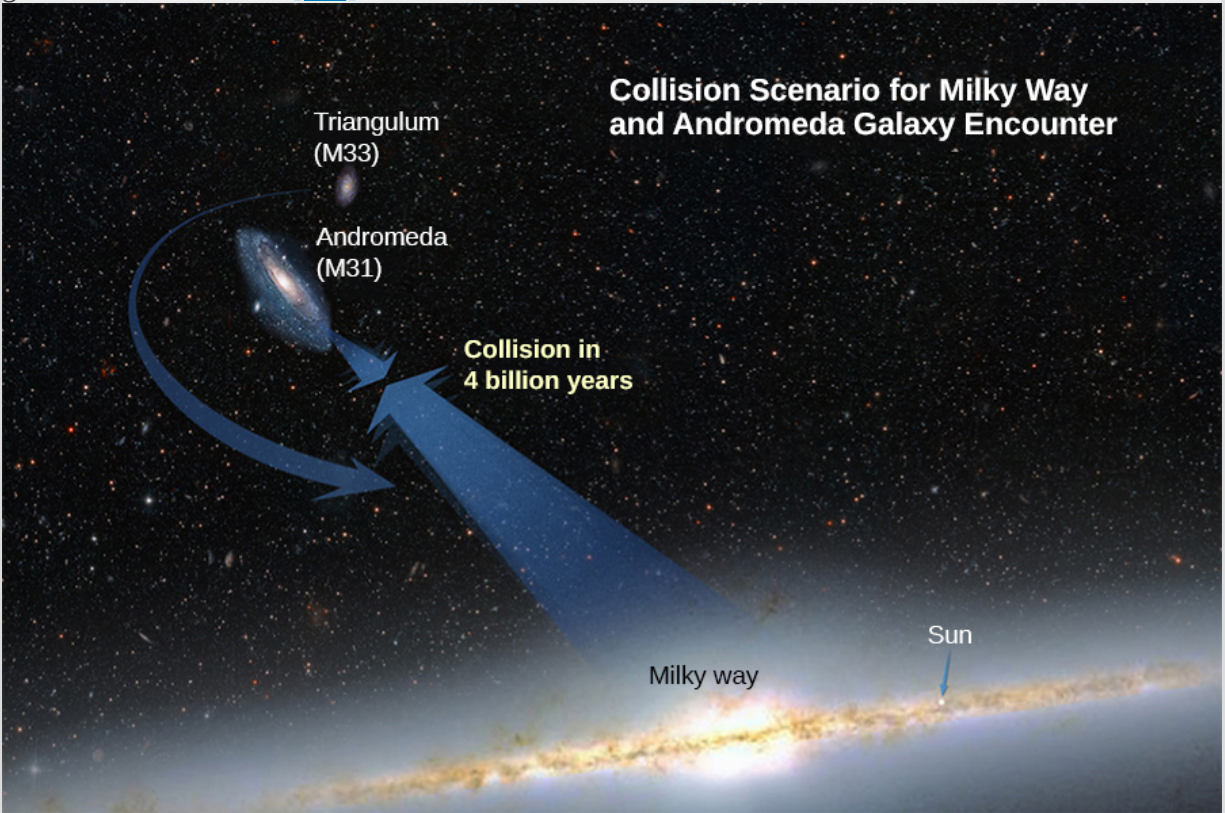
Equation:

$$a = \frac{F}{m} = \frac{3.0 \times 10^{29} \text{ N}}{(800 \times 10^9)(2.0 \times 10^{30} \text{ kg})} = 1.9 \times 10^{-13} \text{ m/s}^2.$$

Significance

Does this value of acceleration seem astoundingly small? If they start from rest, then they would accelerate directly toward each other, "colliding" at their center of mass. Let's estimate the time for

this to happen. The initial acceleration is $\sim 10^{-13} \text{ m/s}^2$, so using $v = at$, we see that it would take $\sim 10^{13} \text{ s}$ for each galaxy to reach a speed of 1.0 m/s , and they would be only $\sim 0.5 \times 10^{13} \text{ m}$ closer. That is nine orders of magnitude smaller than the initial distance between them. In reality, such motions are rarely simple. These two galaxies, along with about 50 other smaller galaxies, are all gravitationally bound into our local cluster. Our local cluster is gravitationally bound to other clusters in what is called a supercluster. All of this is part of the great cosmic dance that results from gravitation, as shown in [\[link\]](#).



Based on the results of this example, plus what astronomers have observed elsewhere in the Universe, our galaxy will collide with the Andromeda Galaxy in about 4 billion years. (credit: modification of work by NASA; ESA; A. Feild and R. van der Marel, STScI)

Summary

- All masses attract one another with a gravitational force proportional to their masses and inversely proportional to the square of the distance between them.
- Spherically symmetrical masses can be treated as if all their mass were located at the center.
- Nonsymmetrical objects can be treated as if their mass were concentrated at their center of mass, provided their distance from other masses is large compared to their size.

Conceptual Questions

Exercise:**Problem:**

Action at a distance, such as is the case for gravity, was once thought to be illogical and therefore untrue. What is the ultimate determinant of the truth in science, and why was this action at a distance ultimately accepted?

Solution:

The ultimate truth is experimental verification. Field theory was developed to help explain how force is exerted without objects being in contact for both gravity and electromagnetic forces that act at the speed of light. It has only been since the twentieth century that we have been able to measure that the force is not conveyed immediately.

Exercise:**Problem:**

In the law of universal gravitation, Newton assumed that the force was proportional to the product of the two masses ($\sim m_1 m_2$). While all scientific conjectures must be experimentally verified, can you provide arguments as to why this must be? (You may wish to consider simple examples in which any other form would lead to contradictory results.)

Problems**Exercise:****Problem:**

Evaluate the magnitude of gravitational force between two 5-kg spherical steel balls separated by a center-to-center distance of 15 cm.

Solution:

$$7.4 \times 10^{-8} \text{ N}$$

Exercise:**Problem:**

Estimate the gravitational force between two sumo wrestlers, with masses 220 kg and 240 kg, when they are embraced and their centers are 1.2 m apart.

Exercise:

Problem:

Astrology makes much of the position of the planets at the moment of one's birth. The only known force a planet exerts on Earth is gravitational. (a) Calculate the gravitational force exerted on a 4.20-kg baby by a 100-kg father 0.200 m away at birth (he is assisting, so he is close to the child). (b) Calculate the force on the baby due to Jupiter if it is at its closest distance to Earth, some 6.29×10^{11} m away. How does the force of Jupiter on the baby compare to the force of the father on the baby? Other objects in the room and the hospital building also exert similar gravitational forces. (Of course, there could be an unknown force acting, but scientists first need to be convinced that there is even an effect, much less that an unknown force causes it.)

Solution:

a. 7.01×10^{-7} N; b. The mass of Jupiter is

$$m_J = 1.90 \times 10^{27} \text{ kg}$$

$$F_J = 1.35 \times 10^{-6} \text{ N}$$

$$\frac{F_f}{F_J} = 0.521$$

Exercise:**Problem:**

A mountain 10.0 km from a person exerts a gravitational force on him equal to 2.00% of his weight. (a) Calculate the mass of the mountain. (b) Compare the mountain's mass with that of Earth. (c) What is unreasonable about these results? (d) Which premises are unreasonable or inconsistent? (Note that accurate gravitational measurements can easily detect the effect of nearby mountains and variations in local geology.)

Exercise:**Problem:**

The International Space Station has a mass of approximately 370,000 kg. (a) What is the force on a 150-kg suited astronaut if she is 20 m from the center of mass of the station? (b) How accurate do you think your answer would be?



(credit: ©ESA–David Ducros)

Solution:

a. 9.25×10^{-6} N; b. Not very, as the ISS is not even symmetrical, much less spherically symmetrical.

Exercise:

Problem:

Asteroid Toutatis passed near Earth in 2006 at four times the distance to our Moon. This was the closest approach we will have until 2060. If it has mass of 5.0×10^{13} kg, what force did it exert on Earth at its closest approach?

Exercise:

Problem:

(a) What was the acceleration of Earth caused by asteroid Toutatis (see previous problem) at its closest approach? (b) What was the acceleration of Toutatis at this point?

Solution:

a. 1.41×10^{-15} m/s²; b. 1.69×10^{-4} m/s²

Glossary

Newton's law of gravitation

every mass attracts every other mass with a force proportional to the product of their masses, inversely proportional to the square of the distance between them, and with direction along the line connecting the center of mass of each

universal gravitational constant

constant representing the strength of the gravitational force, that is believed to be the same throughout the universe

Gravitation Near Earth's Surface

By the end of this section, you will be able to:

- Explain the connection between the constants G and g
- Determine the mass of an astronomical body from free-fall acceleration at its surface
- Describe how the value of g varies due to location and Earth's rotation

In this section, we observe how Newton's law of gravitation applies at the surface of a planet and how it connects with what we learned earlier about free fall. We also examine the gravitational effects within spherical bodies.

Weight

Recall that the acceleration of a free-falling object near Earth's surface is approximately $g = 9.80 \text{ m/s}^2$. The force causing this acceleration is called the weight of the object, and from Newton's second law, it has the value mg . This weight is present regardless of whether the object is in free fall. We now know that this force is the gravitational force between the object and Earth. If we substitute mg for the magnitude of \vec{F}_{12} in Newton's law of universal gravitation, m for m_1 , and M_E for m_2 , we obtain the scalar equation

Equation:

$$mg = G \frac{mM_E}{r^2}$$

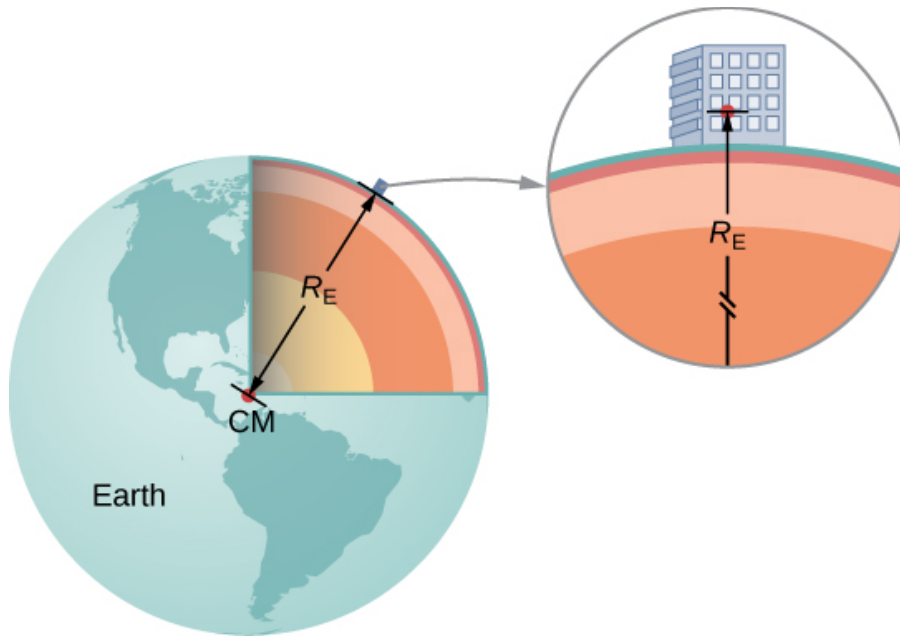
where r is the distance between the centers of mass of the object and Earth. The average radius of Earth is about 6370 km. Hence, for objects within a few kilometers of Earth's surface, we can take $r = R_E$ ([link](#)). The mass m of the object cancels, leaving

Note:

Equation:

$$g = G \frac{M_E}{r^2}.$$

This explains why all masses free fall with the same acceleration. We have ignored the fact that Earth also accelerates toward the falling object, but that is acceptable as long as the mass of Earth is much larger than that of the object.



We can take the distance between the centers of mass of Earth and an object on its surface to be the radius of Earth, provided that its size is much less than the radius of Earth.

Example:

Masses of Earth and Moon

Have you ever wondered how we know the mass of Earth? We certainly can't place it on a scale. The values of g and the radius of Earth were measured with reasonable accuracy centuries ago.

- Use the standard values of g , R_E , and [\[link\]](#) to find the mass of Earth.
- Estimate the value of g on the Moon. Use the fact that the Moon has a radius of about 1700 km (a value of this accuracy was determined many centuries ago) and assume it has the same average density as Earth, 5500 kg/m^3 .

Strategy

With the known values of g and R_E , we can use [\[link\]](#) to find M_E . For the Moon, we use the assumption of equal average density to determine the mass from a ratio of the volumes of Earth and the Moon.

Solution

- Rearranging [\[link\]](#), we have

Equation:

$$M_E = \frac{gR_E^2}{G} = \frac{9.80 \text{ m/s}^2 (6.37 \times 10^6 \text{ m})^2}{6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2} = 5.95 \times 10^{24} \text{ kg}.$$

b. The volume of a sphere is proportional to the radius cubed, so a simple ratio gives us
Equation:

$$\frac{M_M}{M_E} = \frac{R_M^3}{R_E^3} \rightarrow M_M = \left(\frac{(1.7 \times 10^6 \text{ m})^3}{(6.37 \times 10^6 \text{ m})^3} \right) (5.95 \times 10^{24} \text{ kg}) = 1.1 \times 10^{23} \text{ kg}.$$

We now use [\[link\]](#).

Equation:

$$g_M = G \frac{M_M}{r_M^2} = (6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2) \frac{(1.1 \times 10^{23} \text{ kg})}{(1.7 \times 10^6 \text{ m})^2} = 2.5 \text{ m/s}^2$$

Significance

As soon as Cavendish determined the value of G in 1798, the mass of Earth could be calculated. (In fact, that was the ultimate purpose of Cavendish's experiment in the first place.) The value we calculated for g of the Moon is incorrect. The average density of the Moon is actually only 3340 kg/m^3 and $g = 1.6 \text{ m/s}^2$ at the surface. Newton attempted to measure the mass of the Moon by comparing the effect of the Sun on Earth's ocean tides compared to that of the Moon. His value was a factor of two too small. The most accurate values for g and the mass of the Moon come from tracking the motion of spacecraft that have orbited the Moon. But the mass of the Moon can actually be determined accurately without going to the Moon. Earth and the Moon orbit about a common center of mass, and careful astronomical measurements can determine that location. The ratio of the Moon's mass to Earth's is the ratio of [the distance from the common center of mass to the Moon's center] to [the distance from the common center of mass to Earth's center].

Later in this chapter, we will see that the mass of other astronomical bodies also can be determined by the period of small satellites orbiting them. But until Cavendish determined the value of G , the masses of all these bodies were unknown.

Example:

Gravity above Earth's Surface

What is the value of g 400 km above Earth's surface, where the International Space Station is in orbit?

Strategy

Using the value of M_E and noting the radius is $r = R_E + 400 \text{ km}$, we use [\[link\]](#) to find g . From [\[link\]](#) we have

Equation:

$$g = G \frac{M_E}{r^2} = 6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2 \frac{5.96 \times 10^{24} \text{ kg}}{(6.37 \times 10^6 + 400 \times 10^3 \text{ m})^2} = 8.67 \text{ m/s}^2.$$

Significance

We often see video of astronauts in space stations, apparently weightless. But clearly, the force of gravity is acting on them. Comparing the value of g we just calculated to that on Earth (9.80 m/s^2), we see that the astronauts in the International Space Station still have 88% of their weight. They only appear to be weightless because they are in free fall. We will come back to this in [Satellite Orbits and Energy](#).

Note:

Exercise:

Problem:

Check Your Understanding How does your weight at the top of a tall building compare with that on the first floor? Do you think engineers need to take into account the change in the value of g when designing structural support for a very tall building?

Solution:

The tallest buildings in the world are all less than 1 km. Since g is proportional to the distance squared from Earth's center, a simple ratio shows that the change in g at 1 km above Earth's surface is less than 0.0001%. There would be no need to consider this in structural design.

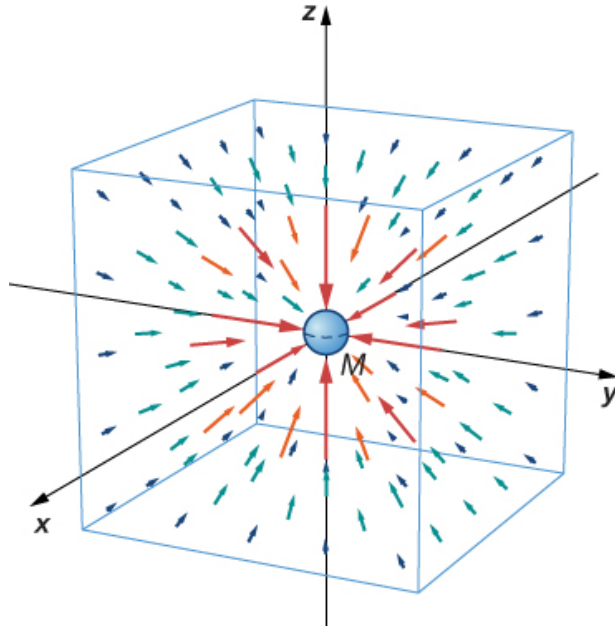
The Gravitational Field

[\[link\]](#) is a scalar equation, giving the magnitude of the gravitational acceleration as a function of the distance from the center of the mass that causes the acceleration. But we could have retained the vector form for the force of gravity in [\[link\]](#), and written the acceleration in vector form as

Equation:

$$\vec{g} = G \frac{M}{r^2} \hat{r}.$$

We identify the vector field represented by \vec{g} as the **gravitational field** caused by mass M . We can picture the field as shown [\[link\]](#). The lines are directed radially inward and are symmetrically distributed about the mass.



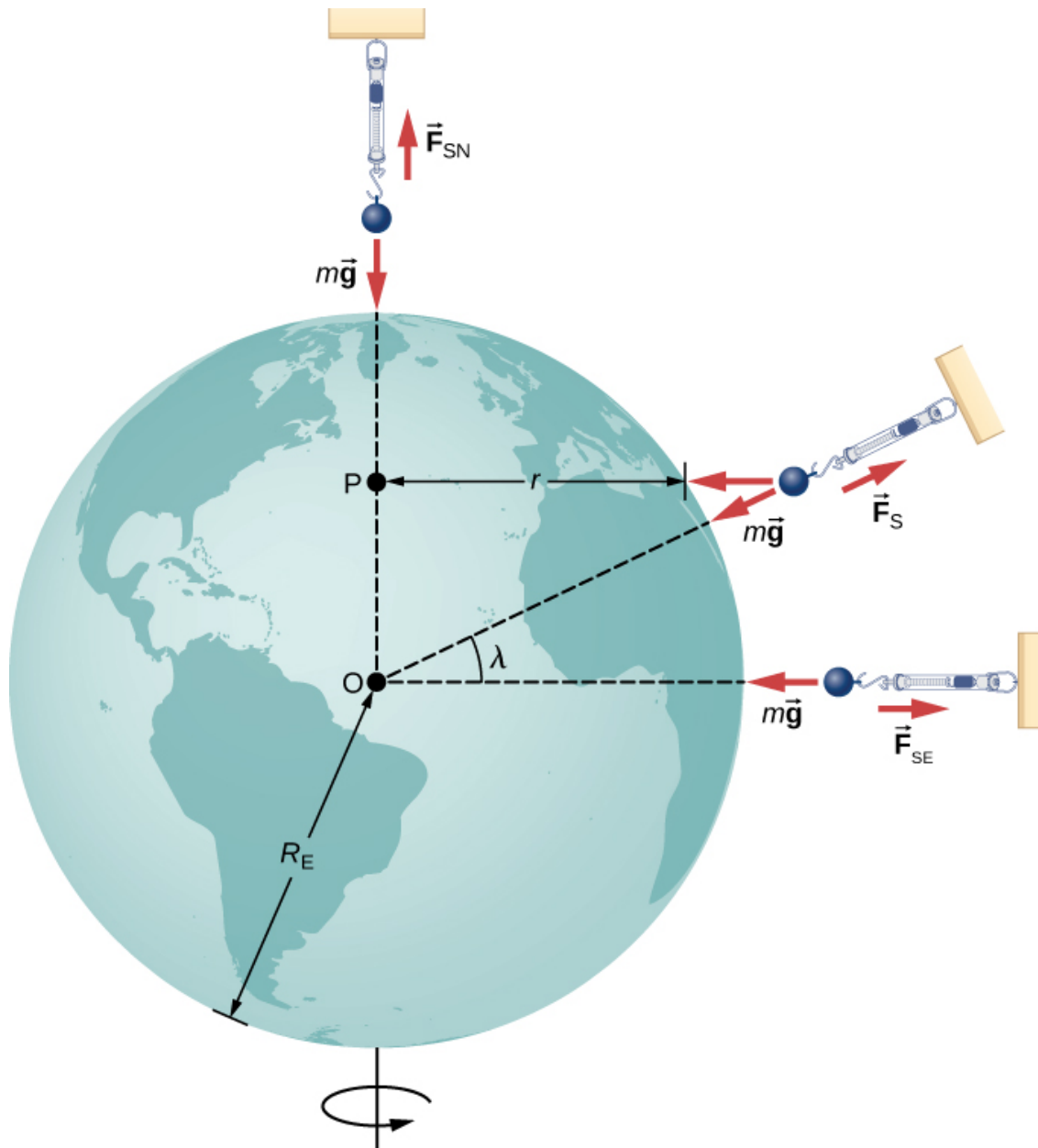
A three-dimensional representation of the gravitational field created by mass M . Note that the lines are uniformly distributed in all directions. (The box has been added only to aid in visualization.)

As is true for any vector field, the direction of \vec{g} is parallel to the field lines at any point. The strength of \vec{g} at any point is inversely proportional to the line spacing. Another way to state this is that the magnitude of the field in any region is proportional to the number of lines that pass through a unit surface area, effectively a density of lines. Since the lines are equally spaced in all directions, the number of lines per unit surface area at a distance r from the mass is the total number of lines divided by the surface area of a sphere of radius r , which is proportional to r^2 . Hence, this picture perfectly represents the inverse square law, in addition to indicating the direction of the field. In the field picture, we say that a mass m interacts with the gravitational field of mass M . We will use the concept of fields to great advantage in the later chapters on electromagnetism.

Apparent Weight: Accounting for Earth's Rotation

As we saw in [Applications of Newton's Laws](#), objects moving at constant speed in a circle have a centripetal acceleration directed toward the center of the circle, which means that there must be a net force directed toward the center of that circle. Since all objects on the surface of Earth move through a circle every 24 hours, there must be a net centripetal force on each object directed toward the center of that circle.

Let's first consider an object of mass m located at the equator, suspended from a scale ([link](#)). The scale exerts an upward force \vec{F}_s away from Earth's center. This is the reading on the scale, and hence it is the **apparent weight** of the object. The weight ($m\vec{g}$) points toward Earth's center. If Earth were not rotating, the acceleration would be zero and, consequently, the net force would be zero, resulting in $F_s = mg$. This would be the true reading of the weight.



For a person standing at the equator, the centripetal acceleration (a_c) is in the same direction as the force of gravity. At latitude λ , the angle between a_c

and the force of gravity is λ and the magnitude of a_c decreases with $\cos\lambda$.

With rotation, the sum of these forces must provide the centripetal acceleration, a_c . Using Newton's second law, we have

Note:

Equation:

$$\sum F = F_s - mg = ma_c \quad \text{where} \quad a_c = -\frac{v^2}{r}.$$

Note that a_c points in the same direction as the weight; hence, it is negative. The tangential speed v is the speed at the equator and r is R_E . We can calculate the speed simply by noting that objects on the equator travel the circumference of Earth in 24 hours. Instead, let's use the alternative expression for a_c from [Motion in Two and Three Dimensions](#). Recall that the tangential speed is related to the angular speed (ω) by $v = r\omega$. Hence, we have $a_c = -r\omega^2$. By rearranging [\[link\]](#) and substituting $r = R_E$, the apparent weight at the equator is

Equation:

$$F_s = m(g - R_E\omega^2).$$

The angular speed of Earth everywhere is

Equation:

$$\omega = \frac{2\pi \text{ rad}}{24 \text{ hr} \times 3600 \text{ s/hr}} = 7.27 \times 10^{-5} \text{ rad/s}.$$

Substituting for the values of R_E and ω , we have $R_E\omega^2 = 0.0337 \text{ m/s}^2$. This is only 0.34% of the value of gravity, so it is clearly a small correction.

Example:

Zero Apparent Weight

How fast would Earth need to spin for those at the equator to have zero apparent weight?

How long would the length of the day be?

Strategy

Using [\[link\]](#), we can set the apparent weight (F_s) to zero and determine the centripetal acceleration required. From that, we can find the speed at the equator. The length of day is the time required for one complete rotation.

Solution

From [\[link\]](#), we have $\sum F = F_s - mg = ma_c$, so setting $F_s = 0$, we get $g = a_c$. Using the expression for a_c , substituting for Earth's radius and the standard value of gravity, we get

Equation:

$$a_c = \frac{v^2}{r} = g$$

$$v = \sqrt{gr} = \sqrt{(9.80 \text{ m/s}^2)(6.37 \times 10^6 \text{ m})} = 7.91 \times 10^3 \text{ m/s}.$$

The period T is the time for one complete rotation. Therefore, the tangential speed is the circumference divided by T , so we have

Equation:

$$v = \frac{2\pi r}{T}$$

$$T = \frac{2\pi r}{v} = \frac{2\pi(6.37 \times 10^6 \text{ m})}{7.91 \times 10^3 \text{ m/s}} = 5.06 \times 10^3 \text{ s}.$$

This is about 84 minutes.

Significance

We will see later in this chapter that this speed and length of day would also be the orbital speed and period of a satellite in orbit at Earth's surface. While such an orbit would not be possible near Earth's surface due to air resistance, it certainly is possible only a few hundred miles above Earth.

Results Away from the Equator

At the poles, $a_c \rightarrow 0$ and $F_s = mg$, just as is the case without rotation. At any other latitude λ , the situation is more complicated. The centripetal acceleration is directed toward point P in the figure, and the radius becomes $r = R_E \cos \lambda$. The vector sum of the weight and \vec{F}_s must point toward point P , hence \vec{F}_s no longer points away from the center of Earth. (The difference is small and exaggerated in the figure.) A plumb bob will always point along this deviated direction. All buildings are built aligned along this deviated direction, not along a radius through the center of Earth. For the tallest buildings, this represents a deviation of a few feet at the top.

It is also worth noting that Earth is not a perfect sphere. The interior is partially liquid, and this enhances Earth bulging at the equator due to its rotation. The radius of Earth is about 30

km greater at the equator compared to the poles. It is left as an exercise to compare the strength of gravity at the poles to that at the equator using [\[link\]](#). The difference is comparable to the difference due to rotation and is in the same direction. Apparently, you really can lose “weight” by moving to the tropics.

Gravity Away from the Surface

Earlier we stated without proof that the law of gravitation applies to spherically symmetrical objects, where the mass of each body acts as if it were at the center of the body. Since [\[link\]](#) is derived from [\[link\]](#), it is also valid for symmetrical mass distributions, but both equations are valid only for values of $r \geq R_E$. As we saw in [\[link\]](#), at 400 km above Earth’s surface, where the International Space Station orbits, the value of g is 8.67 m/s^2 . (We will see later that this is also the centripetal acceleration of the ISS.)

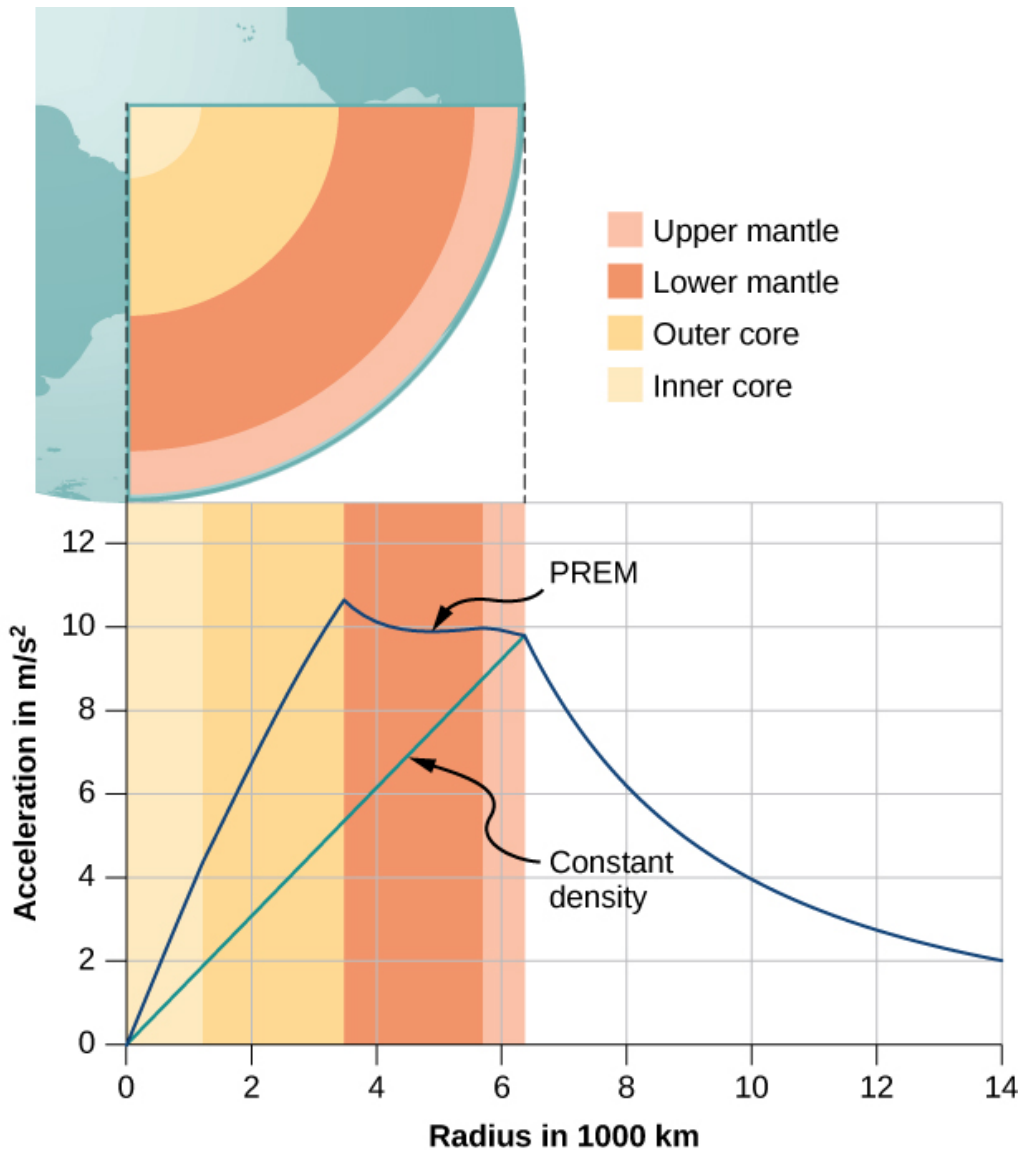
For $r < R_E$, [\[link\]](#) and [\[link\]](#) are not valid. However, we can determine g for these cases using a principle that comes from Gauss’s law, which is a powerful mathematical tool that we study in more detail later in the course. A consequence of Gauss’s law, applied to gravitation, is that only the mass *within* r contributes to the gravitational force. Also, that mass, just as before, can be considered to be located at the center. The gravitational effect of the mass *outside* r has zero net effect.

Two very interesting special cases occur. For a spherical planet with constant density, the mass within r is the density times the volume within r . This mass can be considered located at the center. Replacing M_E with only the mass within r , $M = \rho \times (\text{volume of a sphere})$, and R_E with r , [\[link\]](#) becomes

Equation:

$$g = G \frac{M_E}{R_E^2} = G \frac{\rho (4/3\pi r^3)}{r^2} = \frac{4}{3} G \rho \pi r.$$

The value of g , and hence your weight, decreases linearly as you descend down a hole to the center of the spherical planet. At the center, you are weightless, as the mass of the planet pulls equally in all directions. Actually, Earth’s density is not constant, nor is Earth solid throughout. [\[link\]](#) shows the profile of g if Earth had constant density and the more likely profile based upon estimates of density derived from seismic data.



For $r < R_E$, the value of g for the case of constant density is the straight green line. The blue line from the PREM (Preliminary Reference Earth Model) is probably closer to the actual profile for g .

The second interesting case concerns living on a spherical shell planet. This scenario has been proposed in many science fiction stories. Ignoring significant engineering issues, the shell could be constructed with a desired radius and total mass, such that g at the surface is the same as Earth's. Can you guess what happens once you descend in an elevator to the inside of the shell, where there is no mass between you and the center? What benefits would this provide for traveling great distances from one point on the sphere to another? And finally, what effect would there be if the planet was spinning?

Summary

- The weight of an object is the gravitational attraction between Earth and the object.
- The gravitational field is represented as lines that indicate the direction of the gravitational force; the line spacing indicates the strength of the field.
- Apparent weight differs from actual weight due to the acceleration of the object.

Conceptual Questions

Exercise:

Problem:

Must engineers take Earth's rotation into account when constructing very tall buildings at any location other than the equator or very near the poles?

Solution:

The centripetal acceleration is not directed along the gravitational force and therefore the correct line of the building (i.e., the plumb bob line) is not directed towards the center of Earth. But engineers use either a plumb bob or a transit, both of which respond to both the direction of gravity and acceleration. No special consideration for their location on Earth need be made.

Problems

Exercise:

Problem:

(a) Calculate Earth's mass given the acceleration due to gravity at the North Pole is measured to be 9.832 m/s^2 and the radius of the Earth at the pole is 6356 km. (b) Compare this with the NASA's Earth Fact Sheet value of $5.9726 \times 10^{24} \text{ kg}$.

Exercise:

Problem:

(a) What is the acceleration due to gravity on the surface of the Moon? (b) On the surface of Mars? The mass of Mars is $6.418 \times 10^{23} \text{ kg}$ and its radius is $3.38 \times 10^6 \text{ m}$.

Solution:

a. 1.62 m/s^2 ; b. 3.75 m/s^2

Exercise:

Problem:

(a) Calculate the acceleration due to gravity on the surface of the Sun. (b) By what factor would your weight increase if you could stand on the Sun? (Never mind that you cannot.)

Exercise:**Problem:**

The mass of a particle is 15 kg. (a) What is its weight on Earth? (b) What is its weight on the Moon? (c) What is its mass on the Moon? (d) What is its weight in outer space far from any celestial body? (e) What is its mass at this point?

Solution:

a. 147 N; b. 25.5 N; c. 15 kg; d. 0; e. 15 kg

Exercise:**Problem:**

On a planet whose radius is 1.2×10^7 m, the acceleration due to gravity is 18 m/s^2 . What is the mass of the planet?

Exercise:**Problem:**

The mean diameter of the planet Saturn is 1.2×10^8 m, and its mean mass density is 0.69 g/cm^3 . Find the acceleration due to gravity at Saturn's surface.

Solution:

12 m/s^2

Exercise:**Problem:**

The mean diameter of the planet Mercury is 4.88×10^6 m, and the acceleration due to gravity at its surface is 3.78 m/s^2 . Estimate the mass of this planet.

Exercise:**Problem:**

The acceleration due to gravity on the surface of a planet is three times as large as it is on the surface of Earth. The mass density of the planet is known to be twice that of Earth. What is the radius of this planet in terms of Earth's radius?

Solution:

$$(3/2)R_E$$

Exercise:**Problem:**

A body on the surface of a planet with the same radius as Earth's weighs 10 times more than it does on Earth. What is the mass of this planet in terms of Earth's mass?

Glossary

apparent weight

reading of the weight of an object on a scale that does not account for acceleration

gravitational field

vector field that surrounds the mass creating the field; the field is represented by field lines, in which the direction of the field is tangent to the lines, and the magnitude (or field strength) is inversely proportional to the spacing of the lines; other masses respond to this field

Gravitational Potential Energy and Total Energy

By the end of this section, you will be able to:

- Determine changes in gravitational potential energy over great distances
- Apply conservation of energy to determine escape velocity
- Determine whether astronomical bodies are gravitationally bound

We studied gravitational potential energy in [Potential Energy and Conservation of Energy](#), where the value of g remained constant. We now develop an expression that works over distances such that g is not constant. This is necessary to correctly calculate the energy needed to place satellites in orbit or to send them on missions in space.

Gravitational Potential Energy beyond Earth

We defined work and potential energy in [Work and Kinetic Energy](#) and [Potential Energy and Conservation of Energy](#). The usefulness of those definitions is the ease with which we can solve many problems using conservation of energy. Potential energy is particularly useful for forces that change with position, as the gravitational force does over large distances. In [Potential Energy and Conservation of Energy](#), we showed that the change in gravitational potential energy near Earth's surface is $\Delta U = mg(y_2 - y_1)$. This works very well if g does not change significantly between y_1 and y_2 . We return to the definition of work and potential energy to derive an expression that is correct over larger distances.

Recall that work (W) is the integral of the dot product between force and distance. Essentially, it is the product of the component of a force along a displacement times that displacement. We define ΔU as the *negative* of the work done by the force we associate with the potential energy. For clarity, we derive an expression for moving a mass m from distance r_1 from the center of Earth to distance r_2 . However, the result can easily be generalized to any two objects changing their separation from one value to another.

Consider [\[link\]](#), in which we take m from a distance r_1 from Earth's center to a distance that is r_2 from the center. Gravity is a conservative force (its magnitude and direction are functions of location only), so we can take any path we wish, and the result for the calculation of work is the same. We take the path shown, as it greatly simplifies the integration. We first move *radially* outward from distance r_1 to distance r_2 , and then move along the arc of a circle until we reach the final position. During the radial portion, \vec{F} is opposite to the direction we travel along $d\vec{r}$, so $E = K_1 + U_1 = K_2 + U_2$. Along the arc, \vec{F} is perpendicular to $d\vec{r}$, so $\vec{F} \cdot d\vec{r} = 0$. No work is done as we move along the arc. Using the expression for the gravitational force and noting the values for $\vec{F} \cdot d\vec{r}$ along the two segments of our path, we have

Equation:

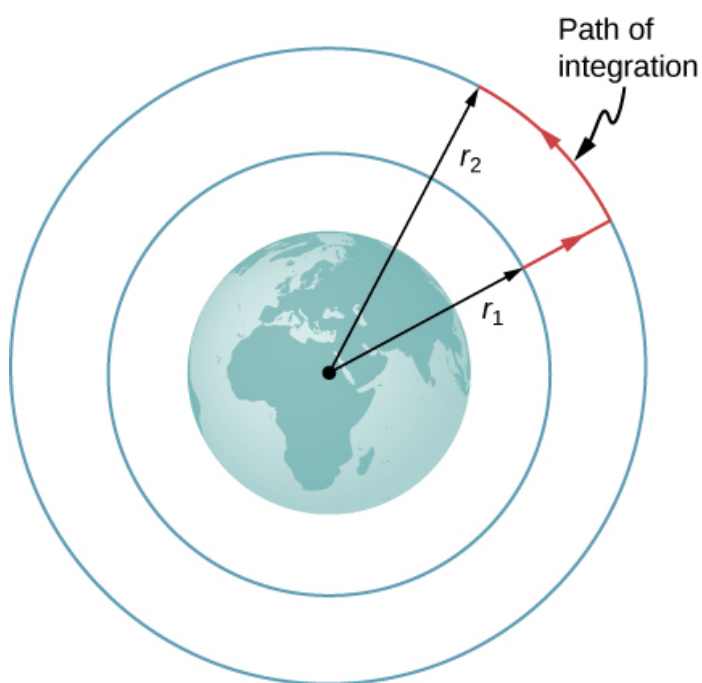
$$\Delta U = - \int_{r_1}^{r_2} \vec{F} \cdot d\vec{r} = GM_E m \int_{r_1}^{r_2} \frac{dr}{r^2} = GM_E m \left(\frac{1}{r_1} - \frac{1}{r_2} \right).$$

Since $\Delta U = U_2 - U_1$, we can adopt a simple expression for U :

Note:

Equation:

$$U = -\frac{GM_{\text{E}}m}{r}.$$



The work integral, which determines the change in potential energy, can be evaluated along the path shown in red.

Note two important items with this definition. First, $U \rightarrow 0$ as $r \rightarrow \infty$. The potential energy is zero when the two masses are infinitely far apart. Only the difference in U is important, so the choice of $U = 0$ for $r = \infty$ is merely one of convenience. (Recall that in earlier gravity problems, you were free to take $U = 0$ at the top or bottom of a building, or anywhere.) Second, note that U becomes increasingly more negative as the masses get closer. That is consistent with what you learned about potential energy in [Potential Energy and Conservation of Energy](#). As the two masses are separated, positive work must be done against the force of

gravity, and hence, U increases (becomes less negative). All masses naturally fall together under the influence of gravity, falling from a higher to a lower potential energy.

Example:**Lifting a Payload**

How much energy is required to lift the 9000-kg *Soyuz* vehicle from Earth's surface to the height of the ISS, 400 km above the surface?

Strategy

Use [\[link\]](#) to find the change in potential energy of the payload. That amount of work or energy must be supplied to lift the payload.

Solution

Paying attention to the fact that we start at Earth's surface and end at 400 km above the surface, the change in U is

Equation:

$$\Delta U = U_{\text{orbit}} - U_{\text{Earth}} = -\frac{GM_{\text{E}}m}{R_{\text{E}} + 400 \text{ km}} - \left(-\frac{GM_{\text{E}}m}{R_{\text{E}}} \right).$$

We insert the values

Equation:

$$m = 9000 \text{ kg}, \quad M_{\text{E}} = 5.96 \times 10^{24} \text{ kg}, \quad R_{\text{E}} = 6.37 \times 10^6 \text{ m}$$

and convert 400 km into $4.00 \times 10^5 \text{ m}$. We find $\Delta U = 3.32 \times 10^{10} \text{ J}$. It is positive, indicating an increase in potential energy, as we would expect.

Significance

For perspective, consider that the average US household energy use in 2013 was 909 kWh per month. That is energy of

Equation:

$$909 \text{ kWh} \times 1000 \text{ W/kW} \times 3600 \text{ s/h} = 3.27 \times 10^9 \text{ J per month.}$$

So our result is an energy expenditure equivalent to 10 months. But this is just the energy needed to raise the payload 400 km. If we want the *Soyuz* to be in orbit so it can rendezvous with the ISS and not just fall back to Earth, it needs a lot of kinetic energy. As we see in the next section, that kinetic energy is about five times that of ΔU . In addition, far more energy is expended lifting the propulsion system itself. Space travel is not cheap.

Note:**Exercise:**

Problem:

Check Your Understanding Why not use the simpler expression $\Delta U = mg(y_2 - y_1)$? How significant would the error be? (Recall the previous result, in [\[link\]](#), that the value g at 400 km above the Earth is 8.67 m/s^2 .)

Solution:

The value of g drops by about 10% over this change in height. So $\Delta U = mg(y_2 - y_1)$ will give too large a value. If we use $g = 9.80 \text{ m/s}^2$, then we get

$$\Delta U = mg(y_2 - y_1) = 3.53 \times 10^{10} \text{ J}$$

which is about 6% greater than that found with the correct method.

Conservation of Energy

In [Potential Energy and Conservation of Energy](#), we described how to apply conservation of energy for systems with conservative forces. We were able to solve many problems, particularly those involving gravity, more simply using conservation of energy. Those principles and problem-solving strategies apply equally well here. The only change is to place the new expression for potential energy into the conservation of energy equation, $E = K_1 + U_1 = K_2 + U_2$.

Note:**Equation:**

$$\frac{1}{2}mv_1^2 - \frac{GMm}{r_1} = \frac{1}{2}mv_2^2 - \frac{GMm}{r_2}$$

Note that we use M , rather than M_E , as a reminder that we are not restricted to problems involving Earth. However, we still assume that $m \ll M$. (For problems in which this is not true, we need to include the kinetic energy of both masses and use conservation of momentum to relate the velocities to each other. But the principle remains the same.)

Escape velocity

Escape velocity is often defined to be the *minimum* initial velocity of an object that is required to escape the surface of a planet (or any large body like a moon) and never return. As usual, we assume no energy lost to an atmosphere, should there be any.

Consider the case where an object is launched from the surface of a planet with an initial velocity directed away from the planet. With the *minimum* velocity needed to escape, the object would *just* come to rest infinitely far away, that is, the object gives up the last of its kinetic energy just as it reaches infinity, where the force of gravity becomes zero. Since $U \rightarrow 0$ as $r \rightarrow \infty$, this means the total energy is zero. Thus, we find the escape velocity from the surface of an astronomical body of mass M and radius R by setting the total energy equal to zero. At the surface of the body, the object is located at $r_1 = R$ and it has escape velocity $v_1 = v_{\text{esc}}$. It reaches $r_2 = \infty$ with velocity $v_2 = 0$. Substituting into [\[link\]](#), we have

Equation:

$$\frac{1}{2}mv_{\text{esc}}^2 - \frac{GMm}{R} = \frac{1}{2}m0^2 - \frac{GMm}{\infty} = 0.$$

Solving for the escape velocity,

Note:

Equation:

$$v_{\text{esc}} = \sqrt{\frac{2GM}{R}}.$$

Notice that m has canceled out of the equation. The escape velocity is the same for all objects, regardless of mass. Also, we are not restricted to the surface of the planet; R can be any starting point beyond the surface of the planet.

Example:

Escape from Earth

What is the escape speed from the surface of Earth? Assume there is no energy loss from air resistance. Compare this to the escape speed from the Sun, starting from Earth's orbit.

Strategy

We use [\[link\]](#), clearly defining the values of R and M . To escape Earth, we need the mass and radius of Earth. For escaping the Sun, we need the mass of the Sun, and the orbital distance between Earth and the Sun.

Solution

Substituting the values for Earth's mass and radius directly into [\[link\]](#), we obtain

Equation:

$$v_{\text{esc}} = \sqrt{\frac{2GM}{R}} = \sqrt{\frac{2(6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2)(5.96 \times 10^{24} \text{ kg})}{6.37 \times 10^6 \text{ m}}} = 1.12 \times 10^4 \text{ m/s}.$$

That is about 11 km/s or 25,000 mph. To escape the Sun, starting from Earth's orbit, we use $R = R_{\text{ES}} = 1.50 \times 10^{11} \text{ m}$ and $M_{\text{Sun}} = 1.99 \times 10^{30} \text{ kg}$. The result is $v_{\text{esc}} = 4.21 \times 10^4 \text{ m/s}$ or about 42 km/s.

Significance

The speed needed to escape the Sun (leave the solar system) is nearly four times the escape speed from Earth's surface. But there is help in both cases. Earth is rotating, at a speed of nearly 1.7 km/s at the equator, and we can use that velocity to help escape, or to achieve orbit. For this reason, many commercial space companies maintain launch facilities near the equator. To escape the Sun, there is even more help. Earth revolves about the Sun at a speed of approximately 30 km/s. By launching in the direction that Earth is moving, we need only an additional 12 km/s. The use of gravitational assist from other planets, essentially a gravity slingshot technique, allows space probes to reach even greater speeds. In this slingshot technique, the vehicle approaches the planet and is accelerated by the planet's gravitational attraction. It has its greatest speed at the closest point of approach, although it accelerates opposite to the motion in equal measure as it moves away. But relative to the planet, the vehicle's speed far before the approach, and long after, are the same. If the directions are chosen correctly, that can result in a significant increase (or decrease if needed) in the vehicle's speed relative to the rest of the solar system.

Note:

Visit this [website](#) to learn more about escape velocity.

Note:**Exercise:****Problem:**

Check Your Understanding If we send a probe out of the solar system starting from Earth's surface, do we only have to escape the Sun?

Solution:

The probe must overcome both the gravitational pull of Earth and the Sun. In the second calculation of our example, we found the speed necessary to escape the Sun from a distance of Earth's orbit, not from Earth itself. The proper way to find this value is to

start with the energy equation, [\[link\]](#), in which you would include a potential energy term for both Earth and the Sun.

Energy and gravitationally bound objects

As stated previously, escape velocity can be defined as the initial velocity of an object that can escape the surface of a moon or planet. More generally, it is the speed at *any* position such that the *total* energy is zero. If the total energy is zero or greater, the object escapes. If the total energy is negative, the object cannot escape. Let's see why that is the case.

As noted earlier, we see that $U \rightarrow 0$ as $r \rightarrow \infty$. If the total energy is zero, then as m reaches a value of r that approaches infinity, U becomes zero and so must the kinetic energy. Hence, m comes to rest infinitely far away from M . It has “just escaped” M . If the total energy is positive, then kinetic energy remains at $r = \infty$ and certainly m does not return. When the total energy is zero or greater, then we say that m is not gravitationally bound to M .

On the other hand, if the total energy is negative, then the kinetic energy must reach zero at some finite value of r , where U is negative and equal to the total energy. The object can never exceed this finite distance from M , since to do so would require the kinetic energy to become negative, which is not possible. We say m is **gravitationally bound** to M .

We have simplified this discussion by assuming that the object was headed directly away from the planet. What is remarkable is that the result applies for any velocity. Energy is a scalar quantity and hence [\[link\]](#) is a scalar equation—the direction of the velocity plays no role in conservation of energy. It is possible to have a gravitationally bound system where the masses do not “fall together,” but maintain an orbital motion about each other.

We have one important final observation. Earlier we stated that if the total energy is zero or greater, the object escapes. Strictly speaking, [\[link\]](#) and [\[link\]](#) apply for point objects. They apply to finite-sized, spherically symmetric objects as well, provided that the value for r in [\[link\]](#) is always greater than the sum of the radii of the two objects. If r becomes less than this sum, then the objects collide. (Even for greater values of r , but near the sum of the radii, gravitational tidal forces could create significant effects if both objects are planet sized. We examine tidal effects in [Tidal Forces](#).) Neither positive nor negative total energy precludes finite-sized masses from colliding. For real objects, direction is important.

Example:

How Far Can an Object Escape?

Let's consider the preceding example again, where we calculated the escape speed from Earth and the Sun, starting from Earth's orbit. We noted that Earth already has an orbital speed of 30 km/s. As we see in the next section, that is the tangential speed needed to stay in circular orbit. If an object had this speed at the distance of Earth's orbit, but was headed directly away

from the Sun, how far would it travel before coming to rest? Ignore the gravitational effects of any other bodies.

Strategy

The object has initial kinetic and potential energies that we can calculate. When its speed reaches zero, it is at its maximum distance from the Sun. We use [\[link\]](#), conservation of energy, to find the distance at which kinetic energy is zero.

Solution

The initial position of the object is Earth's radius of orbit and the initial speed is given as 30 km/s. The final velocity is zero, so we can solve for the distance at that point from the conservation of energy equation. Using $R_{\text{ES}} = 1.50 \times 10^{11} \text{ m}$ and $M_{\text{Sun}} = 1.99 \times 10^{30} \text{ kg}$, we have

Equation:

$$\begin{aligned}\frac{1}{2}mv_1^2 - \frac{GMm}{r_1} &= \frac{1}{2}mv_2^2 - \frac{GMm}{r_2} \\ \frac{1}{2} \cancel{m} (3.0 \times 10^3 \text{ m/s})^2 - \frac{(6.67 \times 10^{-11} \text{ N}\cdot\text{m/kg}^2)(1.99 \times 10^{30} \text{ kg}) \cancel{m}}{1.50 \times 10^{11} \text{ m}} \\ &= \frac{1}{2} \cancel{m} 0^2 - \frac{(6.67 \times 10^{-11} \text{ N}\cdot\text{m/kg}^2)(1.99 \times 10^{30} \text{ kg}) \cancel{m}}{r_2}\end{aligned}$$

where the mass m cancels. Solving for r_2 we get $r_2 = 3.0 \times 10^{11} \text{ m}$. Note that this is twice the initial distance from the Sun and takes us past Mars's orbit, but not quite to the asteroid belt.

Significance

The object in this case reached a distance *exactly* twice the initial orbital distance. We will see the reason for this in the next section when we calculate the speed for circular orbits.

Note:

Exercise:

Problem:

Check Your Understanding Assume you are in a spacecraft in orbit about the Sun at Earth's orbit, but far away from Earth (so that it can be ignored). How could you redirect your tangential velocity to the radial direction such that you could then pass by Mars's orbit? What would be required to change just the direction of the velocity?

Solution:

You change the direction of your velocity with a force that is perpendicular to the velocity at all points. In effect, you must constantly adjust the thrusters, creating a centripetal force until your momentum changes from tangential to radial. A simple momentum vector diagram shows that the net *change* in momentum is $\sqrt{2}$ times the magnitude of momentum itself. This turns out to be a very inefficient way to reach Mars. We discuss the most efficient way in [Kepler's Laws of Planetary Motion](#).

Summary

- The acceleration due to gravity changes as we move away from Earth, and the expression for gravitational potential energy must reflect this change.
- The total energy of a system is the sum of kinetic and gravitational potential energy, and this total energy is conserved in orbital motion.
- Objects must have a minimum velocity, the escape velocity, to leave a planet and not return.
- Objects with total energy less than zero are bound; those with zero or greater are unbound.

Conceptual Questions

Exercise:

Problem:

It was stated that a satellite with negative total energy is in a bound orbit, whereas one with zero or positive total energy is in an unbounded orbit. Why is this true? What choice for gravitational potential energy was made such that this is true?

Exercise:

Problem:

It was shown that the energy required to lift a satellite into a *low* Earth orbit (the change in potential energy) is only a small fraction of the kinetic energy needed to keep it in orbit. Is this true for larger orbits? Is there a trend to the ratio of kinetic energy to change in potential energy as the size of the orbit increases?

Solution:

As we move to larger orbits, the change in potential energy increases, whereas the orbital velocity decreases. Hence, the ratio is highest near Earth's surface (technically infinite if we orbit at Earth's surface with no elevation change), moving to zero as we reach infinitely far away.

Problems

Exercise:

Problem: Find the escape speed of a projectile from the surface of Mars.

Solution:

5000 m/s

Exercise:

Problem: Find the escape speed of a projectile from the surface of Jupiter.

Exercise:

Problem:

What is the escape speed of a satellite located at the Moon's orbit about Earth? Assume the Moon is not nearby.

Solution:

1440 m/s

Exercise:

Problem:

(a) Evaluate the gravitational potential energy between two 5.00-kg spherical steel balls separated by a center-to-center distance of 15.0 cm. (b) Assuming that they are both initially at rest relative to each other in deep space, use conservation of energy to find how fast will they be traveling upon impact. Each sphere has a radius of 5.10 cm.

Exercise:

Problem:

An average-sized asteroid located 5.0×10^7 km from Earth with mass 2.0×10^{13} kg is detected headed directly toward Earth with speed of 2.0 km/s. What will its speed be just before it hits our atmosphere? (You may ignore the size of the asteroid.)

Solution:

11 km/s

Exercise:

Problem:

(a) What will be the kinetic energy of the asteroid in the previous problem just before it hits Earth? b) Compare this energy to the output of the largest fission bomb, 2100 TJ. What impact would this have on Earth?

Exercise:

Problem:

(a) What is the change in energy of a 1000-kg payload taken from rest at the surface of Earth and placed at rest on the surface of the Moon? (b) What would be the answer if the payload were taken from the Moon's surface to Earth? Is this a reasonable calculation of the energy needed to move a payload back and forth?

Solution:

a. 5.85×10^{10} J; b. -5.85×10^{10} J; No. It assumes the kinetic energy is recoverable. This would not even be reasonable if we had an elevator between Earth and the Moon.

Glossary

escape velocity

initial velocity an object needs to escape the gravitational pull of another; it is more accurately defined as the velocity of an object with zero total mechanical energy

gravitationally bound

two objects are gravitationally bound if their orbits are closed; gravitationally bound systems have a negative total mechanical energy

Satellite Orbits and Energy

By the end of this section, you will be able to:

- Describe the mechanism for circular orbits
- Find the orbital periods and speeds of satellites
- Determine whether objects are gravitationally bound

The Moon orbits Earth. In turn, Earth and the other planets orbit the Sun. The space directly above our atmosphere is filled with artificial satellites in orbit. We examine the simplest of these orbits, the circular orbit, to understand the relationship between the speed and period of planets and satellites in relation to their positions and the bodies that they orbit.

Circular Orbits

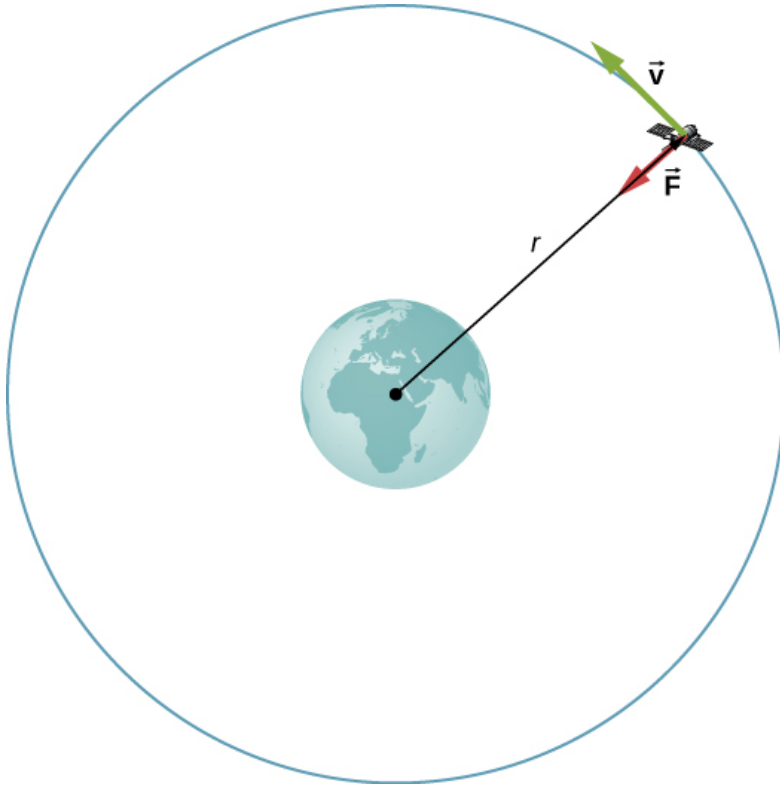
As noted at the beginning of this chapter, Nicolaus Copernicus first suggested that Earth and all other planets orbit the Sun in circles. He further noted that orbital periods increased with distance from the Sun. Later analysis by Kepler showed that these orbits are actually ellipses, but the orbits of most planets in the solar system are nearly circular. Earth's orbital distance from the Sun varies a mere 2%. The exception is the eccentric orbit of Mercury, whose orbital distance varies nearly 40%.

Determining the **orbital speed** and **orbital period** of a satellite is much easier for circular orbits, so we make that assumption in the derivation that follows. As we described in the previous section, an object with negative total energy is gravitationally bound and therefore is in orbit. Our computation for the special case of circular orbits will confirm this. We focus on objects orbiting Earth, but our results can be generalized for other cases.

Consider a satellite of mass m in a circular orbit about Earth at distance r from the center of Earth ([link](#)). It has centripetal acceleration directed toward the center of Earth. Earth's gravity is the only force acting, so Newton's second law gives

Equation:

$$\frac{GmM_E}{r^2} = ma_c = \frac{mv_{\text{orbit}}^2}{r}.$$



A satellite of mass m orbiting at radius r from the center of Earth. The gravitational force supplies the centripetal acceleration.

We solve for the speed of the orbit, noting that m cancels, to get the orbital speed

Note:
Equation:

$$v_{\text{orbit}} = \sqrt{\frac{GM_E}{r}}.$$

Consistent with what we saw in [\[link\]](#) and [\[link\]](#), m does not appear in [\[link\]](#). The value of g , the escape velocity, and orbital velocity depend only upon the distance from the center of the planet, and *not* upon the mass of the object being acted upon. Notice the similarity in the equations for v_{orbit} and v_{esc} . The escape velocity is exactly $\sqrt{2}$ times greater, about 40%, than the orbital velocity. This comparison was noted in [\[link\]](#), and it is true for a satellite at any radius.

To find the period of a circular orbit, we note that the satellite travels the circumference of the orbit $2\pi r$ in one period T . Using the definition of speed, we have $v_{\text{orbit}} = 2\pi r/T$. We substitute this into

[\[link\]](#) and rearrange to get

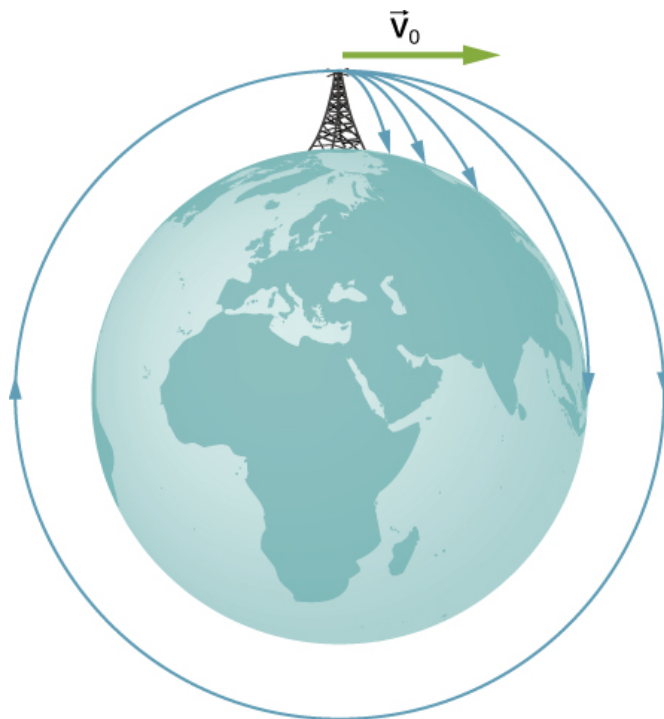
Note:

Equation:

$$T = 2\pi\sqrt{\frac{r^3}{GM_E}}.$$

We see in the next section that this represents Kepler's third law for the case of circular orbits. It also confirms Copernicus's observation that the period of a planet increases with increasing distance from the Sun. We need only replace M_E with M_{Sun} in [\[link\]](#).

We conclude this section by returning to our earlier discussion about astronauts in orbit appearing to be weightless, as if they were free-falling towards Earth. In fact, they are in free fall. Consider the trajectories shown in [\[link\]](#). (This figure is based on a drawing by Newton in his *Principia* and also appeared earlier in [Motion in Two and Three Dimensions](#).) All the trajectories shown that hit the surface of Earth have less than orbital velocity. The astronauts would accelerate toward Earth along the noncircular paths shown and feel weightless. (Astronauts actually train for life in orbit by riding in airplanes that free fall for 30 seconds at a time.) But with the correct orbital velocity, Earth's surface curves away from them at exactly the same rate as they fall toward Earth. Of course, staying the same distance from the surface is the point of a circular orbit.



A circular orbit is the result of choosing a tangential velocity such that Earth's surface curves away at the same rate as the object falls toward Earth.

We can summarize our discussion of orbiting satellites in the following Problem-Solving Strategy.

Note:

Orbits and Conservation of Energy

1. Determine whether the equations for speed, energy, or period are valid for the problem at hand. If not, start with the first principles we used to derive those equations.
2. To start from first principles, draw a free-body diagram and apply Newton's law of gravitation and Newton's second law.
3. Along with the definitions for speed and energy, apply Newton's second law of motion to the bodies of interest.

Example:

The International Space Station

Determine the orbital speed and period for the International Space Station (ISS).

Strategy

Since the ISS orbits $4.00 \times 10^2 \text{ km}$ above Earth's surface, the radius at which it orbits is $R_E + 4.00 \times 10^2 \text{ km}$. We use [\[link\]](#) and [\[link\]](#) to find the orbital speed and period, respectively.

Solution

Using [\[link\]](#), the orbital velocity is

Equation:

$$v_{\text{orbit}} = \sqrt{\frac{GM_E}{r}} = \sqrt{\frac{6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2 (5.96 \times 10^{24} \text{ kg})}{(6.36 \times 10^6 + 4.00 \times 10^5 \text{ m})}} = 7.67 \times 10^3 \text{ m/s}$$

which is about 17,000 mph. Using [\[link\]](#), the period is

Equation:

$$T = 2\pi \sqrt{\frac{r^3}{GM_E}} = 2\pi \sqrt{\frac{(6.37 \times 10^6 + 4.00 \times 10^5 \text{ m})^3}{(6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2)(5.96 \times 10^{24} \text{ kg})}} = 5.55 \times 10^3 \text{ s}$$

which is just over 90 minutes.

Significance

The ISS is considered to be in low Earth orbit (LEO). Nearly all satellites are in LEO, including most weather satellites. GPS satellites, at about 20,000 km, are considered medium Earth orbit. The higher the orbit, the more energy is required to put it there and the more energy is needed to reach it for repairs. Of particular interest are the satellites in geosynchronous orbit. All fixed satellite dishes on

the ground pointing toward the sky, such as TV reception dishes, are pointed toward geosynchronous satellites. These satellites are placed at the exact distance, and just above the equator, such that their period of orbit is 1 day. They remain in a fixed position relative to Earth's surface.

Note:

Exercise:

Problem:

Check Your Understanding By what factor must the radius change to reduce the orbital velocity of a satellite by one-half? By what factor would this change the period?

Solution:

In [\[link\]](#), the radius appears in the denominator inside the square root. So the radius must increase by a factor of 4, to decrease the orbital velocity by a factor of 2. The circumference of the orbit has also increased by this factor of 4, and so with half the orbital velocity, the period must be 8 times longer. That can also be seen directly from [\[link\]](#).

Example:

Determining the Mass of Earth

Determine the mass of Earth from the orbit of the Moon.

Strategy

We use [\[link\]](#), solve for M_E , and substitute for the period and radius of the orbit. The radius and period of the Moon's orbit was measured with reasonable accuracy thousands of years ago. From the astronomical data in [Appendix D](#), the period of the Moon is 27.3 days $= 2.36 \times 10^6$ s, and the average distance between the centers of Earth and the Moon is 384,000 km.

Solution

Solving for M_E ,

Equation:

$$T = 2\pi\sqrt{\frac{r^3}{GM_E}}$$
$$M_E = \frac{4\pi^2 r^3}{GT^2} = \frac{4\pi^2 (3.84 \times 10^8 \text{ m})^3}{(6.67 \times 10^{-11} \text{ N}\cdot\text{m}^2/\text{kg}^2)(2.36 \times 10^6 \text{ s})^2} = 6.01 \times 10^{24} \text{ kg}.$$

Significance

Compare this to the value of 5.96×10^{24} kg that we obtained in [\[link\]](#), using the value of g at the surface of Earth. Although these values are very close ($\sim 0.8\%$), both calculations use average values. The value of g varies from the equator to the poles by approximately 0.5%. But the Moon has an elliptical orbit in which the value of r varies just over 10%. (The apparent size of the full Moon actually varies by about this amount, but it is difficult to notice through casual observation as the time from one extreme to the other is many months.)

Note:

Exercise:**Problem:**

Check Your Understanding There is another consideration to this last calculation of M_E . We derived [\[link\]](#) assuming that the satellite orbits around the center of the astronomical body at the same radius used in the expression for the gravitational force between them. What assumption is made to justify this? Earth is about 81 times more massive than the Moon. Does the Moon orbit about the exact center of Earth?

Solution:

The assumption is that orbiting object is much less massive than the body it is orbiting. This is not really justified in the case of the Moon and Earth. Both Earth and the Moon orbit about their common center of mass. We tackle this issue in the next example.

Example:**Galactic Speed and Period**

Let's revisit [\[link\]](#). Assume that the Milky Way and Andromeda galaxies are in a circular orbit about each other. What would be the velocity of each and how long would their orbital period be? Assume the mass of each is 800 billion solar masses and their centers are separated by 2.5 million light years.

Strategy

We cannot use [\[link\]](#) and [\[link\]](#) directly because they were derived assuming that the object of mass m orbited about the center of a much larger planet of mass M . We determined the gravitational force in [\[link\]](#) using Newton's law of universal gravitation. We can use Newton's second law, applied to the centripetal acceleration of either galaxy, to determine their tangential speed. From that result we can determine the period of the orbit.

Solution

In [\[link\]](#), we found the force between the galaxies to be

Equation:

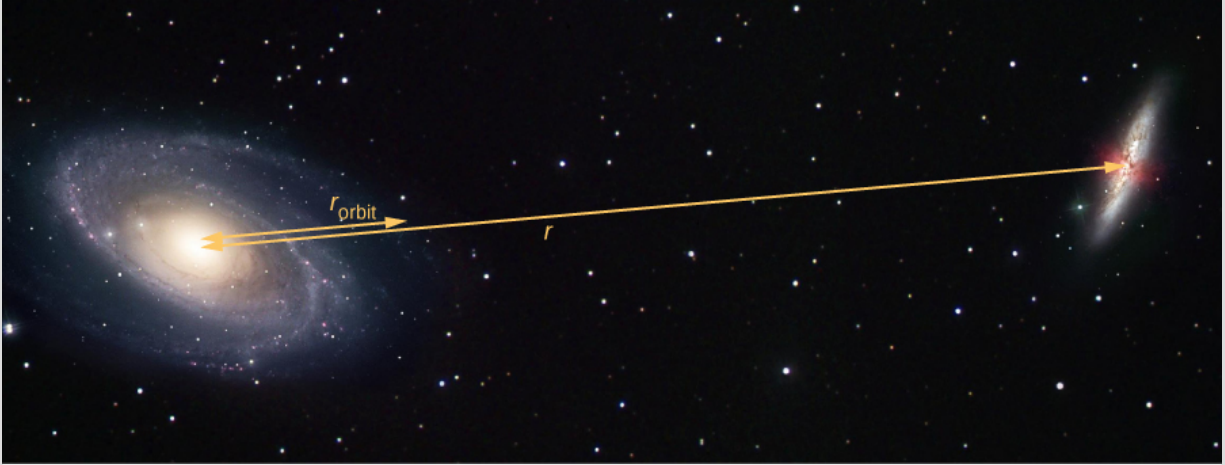
$$F_{12} = G \frac{m_1 m_2}{r^2} = (6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2) \frac{[(800 \times 10^9)(2.0 \times 10^{30} \text{ kg})]^2}{[(2.5 \times 10^6)(9.5 \times 10^{15} \text{ m})]^2} = 3.0 \times 10^{29} \text{ N}$$

and that the acceleration of each galaxy is

Equation:

$$a = \frac{F}{m} = \frac{3.0 \times 10^{29} \text{ N}}{(800 \times 10^9)(2.0 \times 10^{30} \text{ kg})} = 1.9 \times 10^{-13} \text{ m/s}^2.$$

Since the galaxies are in a circular orbit, they have centripetal acceleration. If we ignore the effect of other galaxies, then, as we learned in [Linear Momentum and Collisions](#) and [Fixed-Axis Rotation](#), the centers of mass of the two galaxies remain fixed. Hence, the galaxies must orbit about this common center of mass. For equal masses, the center of mass is exactly half way between them. So the radius of the orbit, r_{orbit} , is not the same as the distance between the galaxies, but one-half that value, or 1.25 million light-years. These two different values are shown in [\[link\]](#).



The distance between two galaxies, which determines the gravitational force between them, is r , and is different from r_{orbit} , which is the radius of orbit for each. For equal masses, $r_{\text{orbit}} = 1/2r$.
(credit: modification of work by Marc Van Norden)

Using the expression for centripetal acceleration, we have

Equation:

$$a_c = \frac{v_{\text{orbit}}^2}{r_{\text{orbit}}}$$

$$1.9 \times 10^{-13} \text{ m/s}^2 = \frac{v_{\text{orbit}}^2}{(1.25 \times 10^6)(9.5 \times 10^{15} \text{ m})}.$$

Solving for the orbit velocity, we have $v_{\text{orbit}} = 47 \text{ km/s}$. Finally, we can determine the period of the orbit directly from $T = 2\pi r/v_{\text{orbit}}$, to find that the period is $T = 1.6 \times 10^{18} \text{ s}$, about 50 billion years.

Significance

The orbital speed of 47 km/s might seem high at first. But this speed is comparable to the escape speed from the Sun, which we calculated in an earlier example. To give even more perspective, this period is nearly four times longer than the time that the Universe has been in existence.

In fact, the present relative motion of these two galaxies is such that they are expected to collide in about 4 billion years. Although the density of stars in each galaxy makes a direct collision of any two stars unlikely, such a collision will have a dramatic effect on the shape of the galaxies. Examples of such collisions are well known in astronomy.

Note:

Exercise:

Problem:

Check Your Understanding Galaxies are not single objects. How does the gravitational force of one galaxy exerted on the “closer” stars of the other galaxy compare to those farther away? What effect would this have on the shape of the galaxies themselves?

Solution:

The stars on the “inside” of each galaxy will be closer to the other galaxy and hence will feel a greater gravitational force than those on the outside. Consequently, they will have a greater acceleration. Even without this force difference, the inside stars would be orbiting at a smaller radius, and, hence, there would develop an elongation or stretching of each galaxy. The force difference only increases this effect.

Note:

See the [Sloan Digital Sky Survey page](#) for more information on colliding galaxies.

Note:

Use this [interactive simulation](#) to move the Sun, Earth, Moon, and space station to see the effects on their gravitational forces and orbital paths. Visualize the sizes and distances between different heavenly bodies, and turn off gravity to see what would happen without it.

Energy in Circular Orbits

In [Gravitational Potential Energy and Total Energy](#), we argued that objects are gravitationally bound if their total energy is negative. The argument was based on the simple case where the velocity was directly away or toward the planet. We now examine the total energy for a circular orbit and show that indeed, the total energy is negative. As we did earlier, we start with Newton’s second law applied to a circular orbit,

Equation:

$$\begin{aligned}\frac{GmM_E}{r^2} &= ma_c = \frac{mv^2}{r} \\ \frac{GmM_E}{r} &= mv^2.\end{aligned}$$

In the last step, we multiplied by r on each side. The right side is just twice the kinetic energy, so we have

Equation:

$$K = \frac{1}{2}mv^2 = \frac{GmM_E}{2r}.$$

The total energy is the sum of the kinetic and potential energies, so our final result is

Note:

Equation:

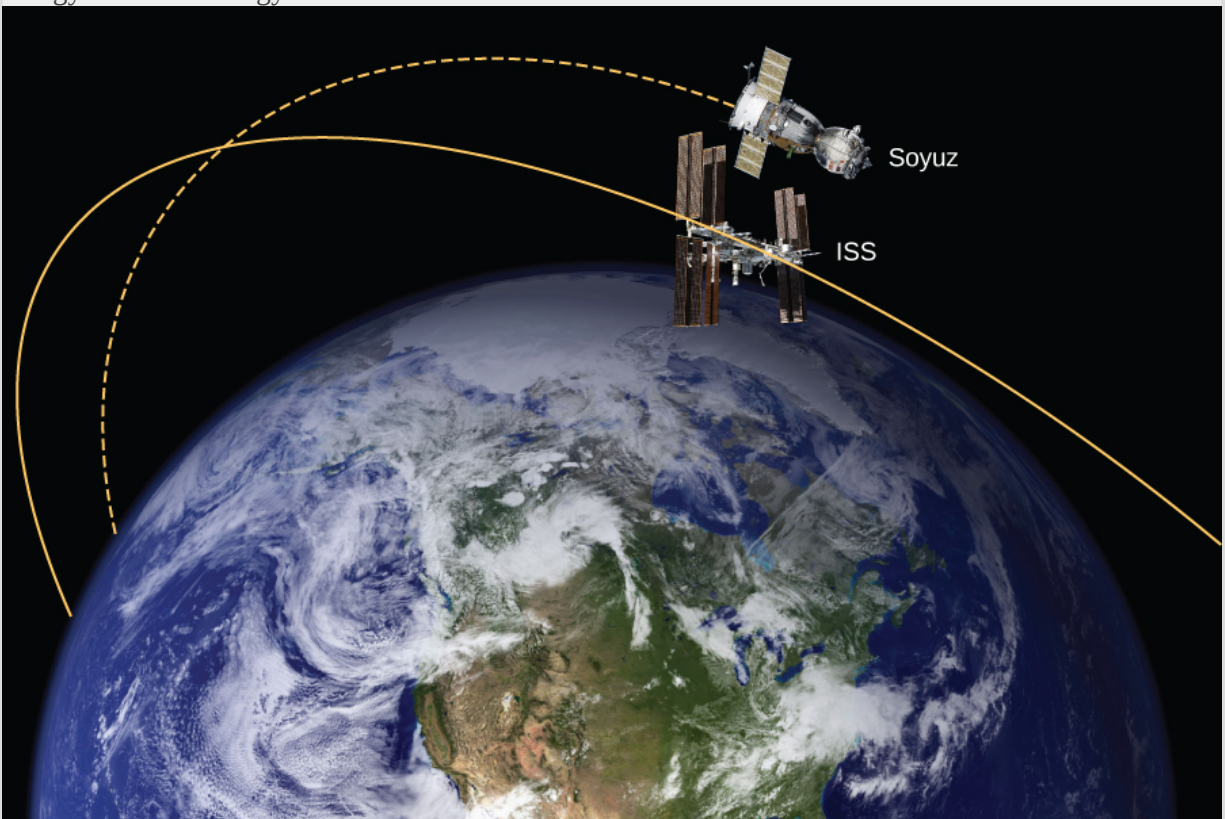
$$E = K + U = \frac{GmM_E}{2r} - \frac{GmM_E}{r} = -\frac{GmM_E}{2r}.$$

We can see that the total energy is negative, with the same magnitude as the kinetic energy. For circular orbits, the magnitude of the kinetic energy is exactly one-half the magnitude of the potential energy. Remarkably, this result applies to any two masses in circular orbits about their common center of mass, at a distance r from each other. The proof of this is left as an exercise. We will see in the next section that a very similar expression applies in the case of elliptical orbits.

Example:

Energy Required to Orbit

In [\[link\]](#), we calculated the energy required to simply lift the 9000-kg *Soyuz* vehicle from Earth's surface to the height of the ISS, 400 km above the surface. In other words, we found its *change* in potential energy. We now ask, what total energy change in the *Soyuz* vehicle is required to take it from Earth's surface and put it in orbit with the ISS for a rendezvous ([\[link\]](#))? How much of that total energy is kinetic energy?



The *Soyuz* in a rendezvous with the ISS. Note that this diagram is not to scale; the *Soyuz* is very small compared to the ISS and its orbit is much closer to Earth. (credit: modification of works by NASA)

Strategy

The energy required is the difference in the *Soyuz*'s total energy in orbit and that at Earth's surface. We can use [\[link\]](#) to find the total energy of the *Soyuz* at the ISS orbit. But the total energy at the surface is simply the potential energy, since it starts from rest. [Note that we *do not* use [\[link\]](#) at the surface, since we are not in orbit at the surface.] The kinetic energy can then be found from the

difference in the total energy change and the change in potential energy found in [\[link\]](#). Alternatively, we can use [\[link\]](#) to find v_{orbit} and calculate the kinetic energy directly from that. The total energy required is then the kinetic energy plus the change in potential energy found in [\[link\]](#).

Solution

From [\[link\]](#), the total energy of the *Soyuz* in the same orbit as the ISS is

Equation:

$$\begin{aligned} E_{\text{orbit}} &= K_{\text{orbit}} + U_{\text{orbit}} = -\frac{GmM_E}{2r} \\ &= \frac{(6.67 \times 10^{-11} \text{ N}\cdot\text{m}^2/\text{kg}^2)(9000 \text{ kg})(5.96 \times 10^{24} \text{ kg})}{2(6.36 \times 10^6 + 4.00 \times 10^5 \text{ m})} = -2.65 \times 10^{11} \text{ J.} \end{aligned}$$

The total energy at Earth's surface is

Equation:

$$\begin{aligned} E_{\text{surface}} &= K_{\text{surface}} + U_{\text{surface}} = 0 - \frac{GmM_E}{r} \\ &= -\frac{(6.67 \times 10^{-11} \text{ N}\cdot\text{m}^2/\text{kg}^2)(9000 \text{ kg})(5.96 \times 10^{24} \text{ kg})}{(6.36 \times 10^6 \text{ m})} \\ &= -5.63 \times 10^{11} \text{ J.} \end{aligned}$$

The change in energy is $\Delta E = E_{\text{orbit}} - E_{\text{surface}} = 2.98 \times 10^{11} \text{ J}$. To get the kinetic energy, we subtract the change in potential energy from [\[link\]](#), $\Delta U = 3.32 \times 10^{10} \text{ J}$. That gives us $K_{\text{orbit}} = 2.98 \times 10^{11} - 3.32 \times 10^{10} = 2.65 \times 10^{11} \text{ J}$. As stated earlier, the kinetic energy of a circular orbit is always one-half the magnitude of the potential energy, and the same as the magnitude of the total energy. Our result confirms this.

The second approach is to use [\[link\]](#) to find the orbital speed of the *Soyuz*, which we did for the ISS in [\[link\]](#).

Equation:

$$v_{\text{orbit}} = \sqrt{\frac{GM_E}{r}} = \sqrt{\frac{(6.67 \times 10^{-11} \text{ N}\cdot\text{m}^2/\text{kg}^2)(5.96 \times 10^{24} \text{ kg})}{(6.36 \times 10^6 + 4.00 \times 10^5 \text{ m})}} = 7.67 \times 10^3 \text{ m/s.}$$

So the kinetic energy of the *Soyuz* in orbit is

Equation:

$$K_{\text{orbit}} = \frac{1}{2}mv_{\text{orbit}}^2 = \frac{1}{2}(9000 \text{ kg})(7.67 \times 10^3 \text{ m/s})^2 = 2.65 \times 10^{11} \text{ J,}$$

the same as in the previous method. The total energy is just

Equation:

$$E_{\text{orbit}} = K_{\text{orbit}} + \Delta U = 2.65 \times 10^{11} + 3.32 \times 10^{10} = 2.95 \times 10^{11} \text{ J.}$$

Significance

The kinetic energy of the *Soyuz* is nearly eight times the change in its potential energy, or 90% of the total energy needed for the rendezvous with the ISS. And it is important to remember that this energy represents only the energy that must be given to the *Soyuz*. With our present rocket technology, the mass of the propulsion system (the rocket fuel, its container and combustion system) far exceeds that of the payload, and a tremendous amount of kinetic energy must be given to that mass. So the actual cost in energy is many times that of the change in energy of the payload itself.

Summary

- Orbital velocities are determined by the mass of the body being orbited and the distance from the center of that body, and not by the mass of a much smaller orbiting object.
- The period of the orbit is likewise independent of the orbiting object's mass.
- Bodies of comparable masses orbit about their common center of mass and their velocities and periods should be determined from Newton's second law and law of gravitation.

Conceptual Questions

Exercise:

Problem:

One student argues that a satellite in orbit is in free fall because the satellite keeps falling toward Earth. Another says a satellite in orbit is not in free fall because the acceleration due to gravity is not 9.80 m/s^2 . With whom do you agree with and why?

Exercise:

Problem:

Many satellites are placed in geosynchronous orbits. What is special about these orbits? For a global communication network, how many of these satellites would be needed?

Solution:

The period of the orbit must be 24 hours. But in addition, the satellite must be located in an equatorial orbit and orbiting in the same direction as Earth's rotation. All three criteria must be met for the satellite to remain in one position relative to Earth's surface. At least three satellites are needed, as two on opposite sides of Earth cannot communicate with each other. (This is not technically true, as a wavelength could be chosen that provides sufficient diffraction. But it would be totally impractical.)

Problems

Exercise:

Problem:

If a planet with 1.5 times the mass of Earth was traveling in Earth's orbit, what would its period be?

Exercise:

Problem:

Two planets in circular orbits around a star have speeds of v and $2v$. (a) What is the ratio of the orbital radii of the planets? (b) What is the ratio of their periods?

Solution:

a. 0.25; b. 0.125

Exercise:**Problem:**

Using the average distance of Earth from the Sun, and the orbital period of Earth, (a) find the centripetal acceleration of Earth in its motion about the Sun. (b) Compare this value to that of the centripetal acceleration at the equator due to Earth's rotation.

Exercise:**Problem:**

(a) What is the orbital radius of an Earth satellite having a period of 1.00 h? (b) What is unreasonable about this result?

Solution:

a. 5.08×10^3 km; b. This less than the radius of Earth.

Exercise:**Problem:**

Calculate the mass of the Sun based on data for Earth's orbit and compare the value obtained with the Sun's actual mass.

Exercise:**Problem:**

Find the mass of Jupiter based on the fact that Io, its innermost moon, has an average orbital radius of 421,700 km and a period of 1.77 days.

Solution:

1.89×10^{27} kg

Exercise:**Problem:**

Astronomical observations of our Milky Way galaxy indicate that it has a mass of about 8.0×10^{11} solar masses. A star orbiting on the galaxy's periphery is about 6.0×10^4 light-years from its center. (a) What should the orbital period of that star be? (b) If its period is 6.0×10^7 years instead, what is the mass of the galaxy? Such calculations are used to imply the existence of other matter, such as a very massive black hole at the center of the Milky Way.

Exercise:**Problem:**

(a) In order to keep a small satellite from drifting into a nearby asteroid, it is placed in orbit with a period of 3.02 hours and radius of 2.0 km. What is the mass of the asteroid? (b) Does this mass seem reasonable for the size of the orbit?

Solution:

a. 4.01×10^{13} kg; b. The satellite must be outside the radius of the asteroid, so it can't be larger than this. If it were this size, then its density would be about 1200 kg/m^3 . This is just above that of water, so this seems quite reasonable.

Exercise:

Problem:

The Moon and Earth rotate about their common center of mass, which is located about 4700 km from the center of Earth. (This is 1690 km below the surface.) (a) Calculate the acceleration due to the Moon's gravity at that point. (b) Calculate the centripetal acceleration of the center of Earth as it rotates about that point once each lunar month (about 27.3 d) and compare it with the acceleration found in part (a). Comment on whether or not they are equal and why they should or should not be.

Exercise:

Problem:

The Sun orbits the Milky Way galaxy once each 2.60×10^8 years, with a roughly circular orbit averaging a radius of 3.00×10^4 light-years. (A light-year is the distance traveled by light in 1 year.) Calculate the centripetal acceleration of the Sun in its galactic orbit. Does your result support the contention that a nearly inertial frame of reference can be located at the Sun? (b) Calculate the average speed of the Sun in its galactic orbit. Does the answer surprise you?

Solution:

a. $1.66 \times 10^{-10} \text{ m/s}^2$; Yes, the centripetal acceleration is so small it supports the contention that a nearly inertial frame of reference can be located at the Sun. b. $2.17 \times 10^5 \text{ m/s}$

Exercise:

Problem:

A geosynchronous Earth satellite is one that has an orbital period of precisely 1 day. Such orbits are useful for communication and weather observation because the satellite remains above the same point on Earth (provided it orbits in the equatorial plane in the same direction as Earth's rotation). Calculate the radius of such an orbit based on the data for Earth in [Appendix D](#).

Glossary

orbital period

time required for a satellite to complete one orbit

orbital speed

speed of a satellite in a circular orbit; it can be also be used for the instantaneous speed for noncircular orbits in which the speed is not constant

Kepler's Laws of Planetary Motion

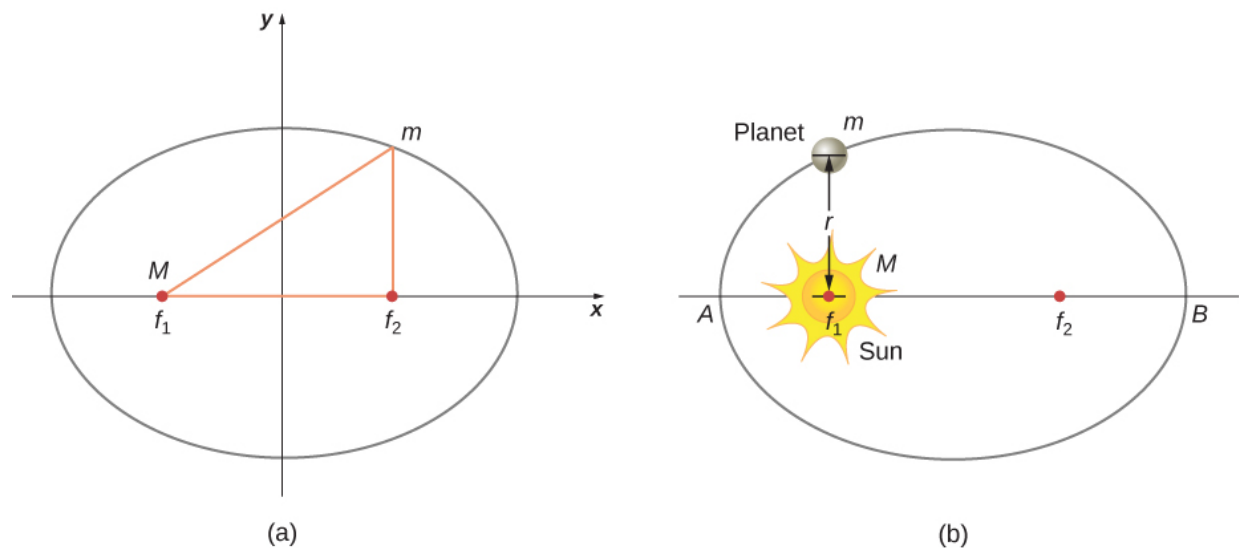
By the end of this section, you will be able to:

- Describe the conic sections and how they relate to orbital motion
- Describe how orbital velocity is related to conservation of angular momentum
- Determine the period of an elliptical orbit from its major axis

Using the precise data collected by Tycho Brahe, Johannes Kepler carefully analyzed the positions in the sky of all the known planets and the Moon, plotting their positions at regular intervals of time. From this analysis, he formulated three laws, which we address in this section.

Kepler's First Law

The prevailing view during the time of Kepler was that all planetary orbits were circular. The data for Mars presented the greatest challenge to this view and that eventually encouraged Kepler to give up the popular idea. **Kepler's first law** states that every planet moves along an ellipse, with the Sun located at a focus of the ellipse. An ellipse is defined as the set of all points such that the sum of the distance from each point to two foci is a constant. [\[link\]](#) shows an ellipse and describes a simple way to create it.



(a) An ellipse is a curve in which the sum of the distances from a point on the curve to two foci (f_1 and f_2) is a constant. From this definition, you can see that an ellipse can be created in the following way. Place a pin at each focus, then place a loop of string around a pencil and the pins. Keeping the string taut, move the pencil around in a complete circuit. If the two foci occupy the same place, the result is a circle—a special case of an ellipse. (b) For an elliptical orbit, if $m \ll M$, then m follows an elliptical path with M at one focus. More exactly, both m and M move in their own ellipse about the common center of mass.

For elliptical orbits, the point of closest approach of a planet to the Sun is called the **perihelion**. It is labeled point *A* in [\[link\]](#). The farthest point is the **aphelion** and is labeled point *B* in the figure. For the Moon's orbit about Earth, those points are called the perigee and apogee, respectively.

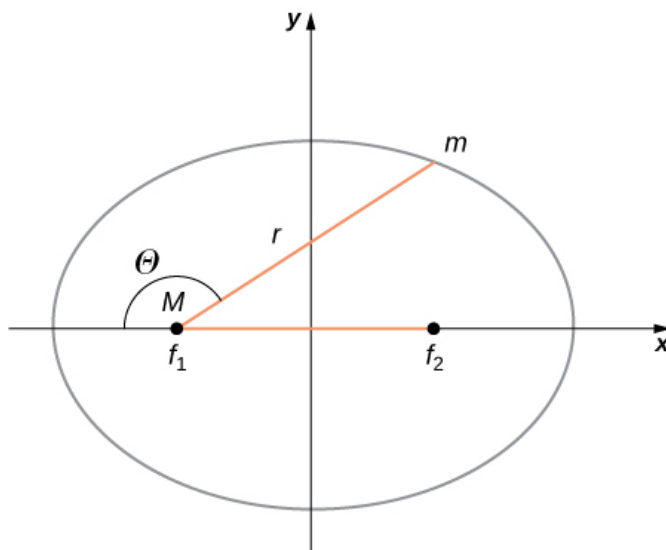
An ellipse has several mathematical forms, but all are a specific case of the more general equation for conic sections. There are four different conic sections, all given by the equation

Note:

Equation:

$$\frac{\alpha}{r} = 1 + e \cos \theta.$$

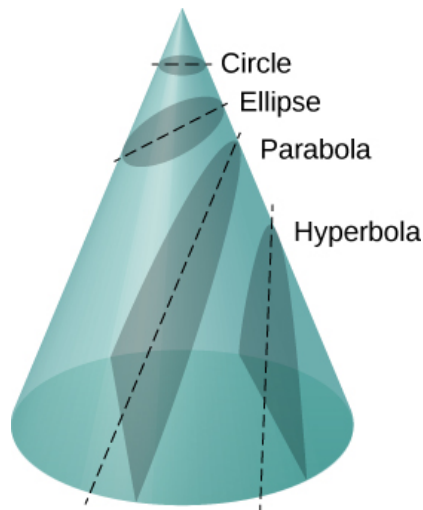
The variables r and θ are shown in [\[link\]](#) in the case of an ellipse. The constants α and e are determined by the total energy and angular momentum of the satellite at a given point. The constant e is called the eccentricity. The values of α and e determine which of the four conic sections represents the path of the satellite.



As before, the distance between the planet and the Sun is r , and the angle measured from the x -axis, which is along the major axis of the ellipse, is θ .

One of the real triumphs of Newton's law of universal gravitation, with the force proportional to the inverse of the distance squared, is that when it is combined with his second law, the solution for the path of any satellite is a conic section. Every path taken by m is one of the four conic sections: a circle

or an ellipse for bound or closed orbits, or a parabola or hyperbola for unbounded or open orbits. These conic sections are shown in [\[link\]](#).



All motion caused by an inverse square force is one of the four conic sections and is determined by the energy and direction of the moving body.

If the total energy is negative, then $0 \leq e < 1$, and [\[link\]](#) represents a bound or closed orbit of either an ellipse or a circle, where $e = 0$. [You can see from [\[link\]](#) that for $e = 0$, $r = \alpha$, and hence the radius is constant.] For ellipses, the eccentricity is related to how oblong the ellipse appears. A circle has zero eccentricity, whereas a very long, drawn-out ellipse has an eccentricity near one.

If the total energy is exactly zero, then $e = 1$ and the path is a parabola. Recall that a satellite with zero total energy has exactly the escape velocity. (The parabola is formed only by slicing the cone parallel to the tangent line along the surface.) Finally, if the total energy is positive, then $e > 1$ and the path is a hyperbola. These last two paths represent unbounded orbits, where m passes by M once and only once. This situation has been observed for several comets that approach the Sun and then travel away, never to return.

We have confined ourselves to the case in which the smaller mass (planet) orbits a much larger, and hence stationary, mass (Sun), but [\[link\]](#) also applies to any two gravitationally interacting masses. Each mass traces out the exact same-shaped conic section as the other. That shape is determined by the total energy and angular momentum of the system, with the center of mass of the system located at the focus. The ratio of the dimensions of the two paths is the inverse of the ratio of their masses.

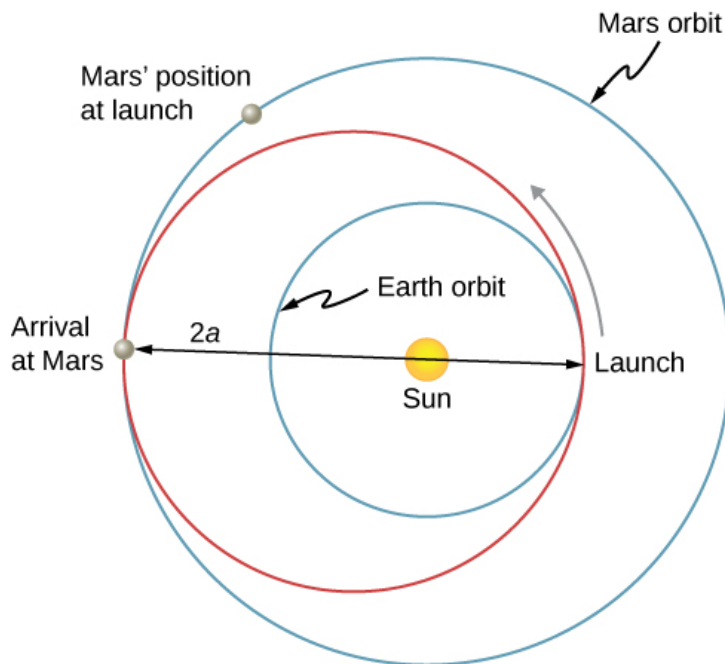
Note:

You can see an animation of two interacting objects at the *My Solar System* page at [Phet](#). Choose the Sun and Planet preset option. You can also view the more complicated multiple body problems as well. You may find the actual path of the Moon quite surprising, yet is obeying Newton's simple laws of motion.

Orbital Transfers

People have imagined traveling to the other planets of our solar system since they were discovered. But how can we best do this? The most efficient method was discovered in 1925 by Walter Hohmann, inspired by a popular science fiction novel of that time. The method is now called a Hohmann transfer. For the case of traveling between two circular orbits, the transfer is along a “transfer” ellipse that perfectly intercepts those orbits at the aphelion and perihelion of the ellipse. [\[link\]](#) shows the case for a trip from Earth's orbit to that of Mars. As before, the Sun is at the focus of the ellipse.

For any ellipse, the semi-major axis is defined as one-half the sum of the perihelion and the aphelion. In [\[link\]](#), the semi-major axis is the distance from the origin to either side of the ellipse along the x -axis, or just one-half the longest axis (called the major axis). Hence, to travel from one circular orbit of radius r_1 to another circular orbit of radius r_2 , the aphelion of the transfer ellipse will be equal to the value of the larger orbit, while the perihelion will be the smaller orbit. The semi-major axis, denoted a , is therefore given by $a = \frac{1}{2}(r_1 + r_2)$.



The transfer ellipse has its perihelion at Earth's orbit and aphelion at Mars' orbit.

Let's take the case of traveling from Earth to Mars. For the moment, we ignore the planets and assume we are alone in Earth's orbit and wish to move to Mars' orbit. From [\[link\]](#), the expression for total energy, we can see that the total energy for a spacecraft in the larger orbit (Mars) is greater (less negative) than that for the smaller orbit (Earth). To move onto the transfer ellipse from Earth's orbit, we will need to increase our kinetic energy, that is, we need a velocity boost. The most efficient method is a very quick acceleration along the circular orbital path, which is also along the path of the ellipse at that point. (In fact, the acceleration should be instantaneous, such that the circular and elliptical orbits are congruent during the acceleration. In practice, the finite acceleration is short enough that the difference is not a significant consideration.) Once you have arrived at Mars orbit, you will need another velocity boost to move into that orbit, or you will stay on the elliptical orbit and simply fall back to perihelion where you started. For the return trip, you simply reverse the process with a retro-boost at each transfer point.

To make the move onto the transfer ellipse and then off again, we need to know each circular orbit velocity and the transfer orbit velocities at perihelion and aphelion. The velocity boost required is simply the difference between the circular orbit velocity and the elliptical orbit velocity at each point. We can find the circular orbital velocities from [\[link\]](#). To determine the velocities for the ellipse, we state without proof (as it is beyond the scope of this course) that total energy for an elliptical orbit is **Equation:**

$$E = -\frac{GmM_S}{2a}$$

where M_S is the mass of the Sun and a is the semi-major axis. Remarkably, this is the same as [\[link\]](#) for circular orbits, but with the value of the semi-major axis replacing the orbital radius. Since we know the potential energy from [\[link\]](#), we can find the kinetic energy and hence the velocity needed for each point on the ellipse. We leave it as a challenge problem to find those transfer velocities for an Earth-to-Mars trip.

We end this discussion by pointing out a few important details. First, we have not accounted for the gravitational potential energy due to Earth and Mars, or the mechanics of landing on Mars. In practice, that must be part of the calculations. Second, timing is everything. You do not want to arrive at the orbit of Mars to find out it isn't there. We must leave Earth at precisely the correct time such that Mars will be at the aphelion of our transfer ellipse just as we arrive. That opportunity comes about every 2 years. And returning requires correct timing as well. The total trip would take just under 3 years! There are other options that provide for a faster transit, including a gravity assist flyby of Venus. But these other options come with an additional cost in energy and danger to the astronauts.

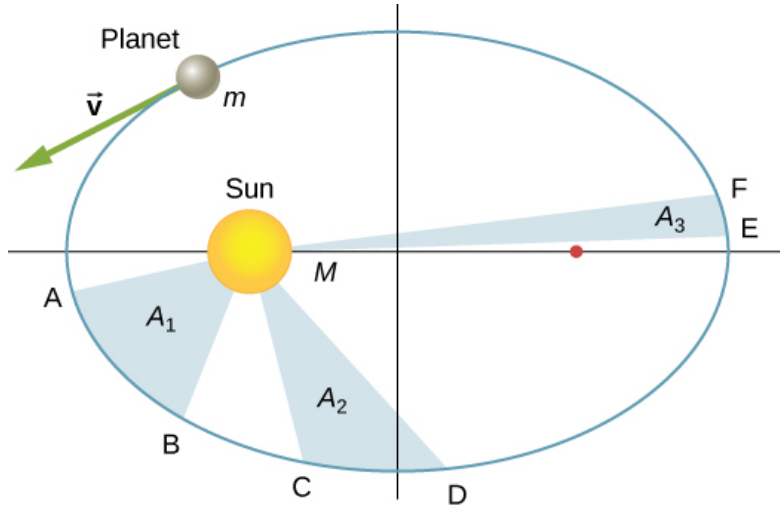
Note:

Visit this [site](#) for more details about planning a trip to Mars.

Kepler's Second Law

Kepler's second law states that a planet sweeps out equal areas in equal times, that is, the area divided by time, called the areal velocity, is constant. Consider [\[link\]](#). The time it takes a planet to move from position A to B , sweeping out area A_1 , is exactly the time taken to move from position C to D ,

sweeping area A_2 , and to move from E to F , sweeping out area A_3 . These areas are the same: $A_1 = A_2 = A_3$.



The shaded regions shown have equal areas and represent the same time interval.

Comparing the areas in the figure and the distance traveled along the ellipse in each case, we can see that in order for the areas to be equal, the planet must speed up as it gets closer to the Sun and slow down as it moves away. This behavior is completely consistent with our conservation equation, [\[link\]](#). But we will show that Kepler's second law is actually a consequence of the conservation of angular momentum, which holds for any system with only radial forces.

Recall the definition of angular momentum from [Angular Momentum](#), $\vec{L} = \vec{r} \times \vec{p}$. For the case of orbiting motion, \vec{L} is the angular momentum of the planet about the Sun, \vec{r} is the position vector of the planet measured from the Sun, and $\vec{p} = m\vec{v}$ is the instantaneous linear momentum at any point in the orbit. Since the planet moves along the ellipse, \vec{p} is always tangent to the ellipse.

We can resolve the linear momentum into two components: a radial component \vec{p}_{rad} along the line to the Sun, and a component \vec{p}_{perp} perpendicular to \vec{r} . The cross product for angular momentum can then be written as

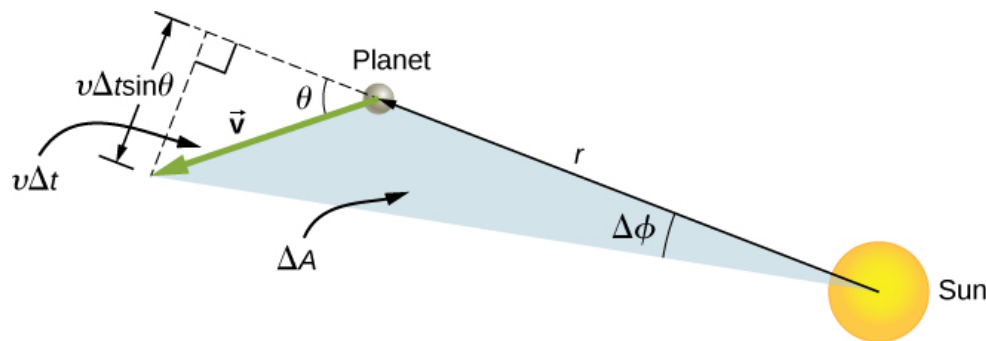
$$\vec{L} = \vec{r} \times \vec{p} = \vec{r} \times (\vec{p}_{\text{rad}} + \vec{p}_{\text{perp}}) = \vec{r} \times \vec{p}_{\text{rad}} + \vec{r} \times \vec{p}_{\text{perp}}.$$

The first term on the right is zero because \vec{r} is parallel to \vec{p}_{rad} , and in the second term \vec{r} is perpendicular to \vec{p}_{perp} , so the magnitude of the cross product reduces to $L = rp_{\text{perp}} = rmv_{\text{perp}}$. Note that the angular momentum does *not* depend upon p_{rad} . Since the gravitational force is only in the radial direction, it can change only p_{rad} and not p_{perp} ; hence, the angular momentum must remain constant.

Now consider [\[link\]](#). A small triangular area ΔA is swept out in time Δt . The velocity is along the path and it makes an angle θ with the radial direction. Hence, the perpendicular velocity is given by $v_{\text{perp}} = v \sin \theta$. The planet moves a distance $\Delta s = v \Delta t \sin \theta$ projected along the direction perpendicular to r . Since the area of a triangle is one-half the base (r) times the height (Δs), for a small displacement, the area is given by $\Delta A = \frac{1}{2} r \Delta s$. Substituting for Δs , multiplying by m in the numerator and denominator, and rearranging, we obtain

Equation:

$$\Delta A = \frac{1}{2} r \Delta s = \frac{1}{2} r (v \Delta t \sin \theta) = \frac{1}{2m} r (m v \sin \theta \Delta t) = \frac{1}{2m} r (m v_{\text{perp}} \Delta t) = \frac{L}{2m} \Delta t.$$



The element of area ΔA swept out in time Δt as the planet moves through angle $\Delta \phi$. The angle between the radial direction and \vec{v} is θ .

The areal velocity is simply the rate of change of area with time, so we have

Equation:

$$\text{areal velocity} = \frac{\Delta A}{\Delta t} = \frac{L}{2m}.$$

Since the angular momentum is constant, the areal velocity must also be constant. This is exactly Kepler's second law. As with Kepler's first law, Newton showed it was a natural consequence of his law of gravitation.

Note:

You can view an [animated version](#) of [\[link\]](#), and many other interesting animations as well, at the School of Physics (University of New South Wales) site.

Kepler's Third Law

Kepler's third law states that the square of the period is proportional to the cube of the semi-major axis of the orbit. In [Satellite Orbits and Energy](#), we derived Kepler's third law for the special case of a circular orbit. [\[link\]](#) gives us the period of a circular orbit of radius r about Earth:

Equation:

$$T = 2\pi \sqrt{\frac{r^3}{GM_E}}.$$

For an ellipse, recall that the semi-major axis is one-half the sum of the perihelion and the aphelion. For a circular orbit, the semi-major axis (a) is the same as the radius for the orbit. In fact, [\[link\]](#) gives us Kepler's third law if we simply replace r with a and square both sides.

Note:

Equation:

$$T^2 = \frac{4\pi^2}{GM} a^3$$

We have changed the mass of Earth to the more general M , since this equation applies to satellites orbiting any large mass.

Example:

Orbit of Halley's Comet

Determine the semi-major axis of the orbit of Halley's comet, given that it arrives at perihelion every 75.3 years. If the perihelion is 0.586 AU, what is the aphelion?

Strategy

We are given the period, so we can rearrange [\[link\]](#), solving for the semi-major axis. Since we know the value for the perihelion, we can use the definition of the semi-major axis, given earlier in this section, to find the aphelion. We note that 1 Astronomical Unit (AU) is the average radius of Earth's orbit and is defined to be $1 \text{ AU} = 1.50 \times 10^{11} \text{ m}$.

Solution

Rearranging [\[link\]](#) and inserting the values of the period of Halley's comet and the mass of the Sun, we have

Equation:

$$\begin{aligned} a &= \left(\frac{GM}{4\pi^2} T^2 \right)^{1/3} \\ &= \left(\frac{(6.67 \times 10^{-11} \text{ N}\cdot\text{m}^2/\text{kg}^2)(2.00 \times 10^{30} \text{ kg})}{4\pi^2} (75.3 \text{ yr} \times 365 \text{ days/yr} \times 24 \text{ hr/day} \times 3600 \text{ s/hr})^2 \right)^{1/3}. \end{aligned}$$

This yields a value of $2.67 \times 10^{12} \text{ m}$ or 17.8 AU for the semi-major axis.

The semi-major axis is one-half the sum of the aphelion and perihelion, so we have

Equation:

$$a = \frac{1}{2}(\text{aphelion} + \text{perihelion})$$

$$\text{aphelion} = 2a - \text{perihelion}.$$

Substituting for the values, we found for the semi-major axis and the value given for the perihelion, we find the value of the aphelion to be 35.0 AU.

Significance

Edmond Halley, a contemporary of Newton, first suspected that three comets, reported in 1531, 1607, and 1682, were actually the same comet. Before Tycho Brahe made measurements of comets, it was believed that they were one-time events, perhaps disturbances in the atmosphere, and that they were not affected by the Sun. Halley used Newton's new mechanics to predict his namesake comet's return in 1758.

Note:

Exercise:

Problem:

Check Your Understanding The nearly circular orbit of Saturn has an average radius of about 9.5 AU and has a period of 30 years, whereas Uranus averages about 19 AU and has a period of 84 years. Is this consistent with our results for Halley's comet?

Solution:

The semi-major axis for the highly elliptical orbit of Halley's comet is 17.8 AU and is the average of the perihelion and aphelion. This lies between the 9.5 AU and 19 AU orbital radii for Saturn and Uranus, respectively. The radius for a circular orbit is the same as the semi-major axis, and since the period increases with an increase of the semi-major axis, the fact that Halley's period is between the periods of Saturn and Uranus is expected.

Summary

- All orbital motion follows the path of a conic section. Bound or closed orbits are either a circle or an ellipse; unbounded or open orbits are either a parabola or a hyperbola.
- The areal velocity of any orbit is constant, a reflection of the conservation of angular momentum.
- The square of the period of an elliptical orbit is proportional to the cube of the semi-major axis of that orbit.

Conceptual Questions

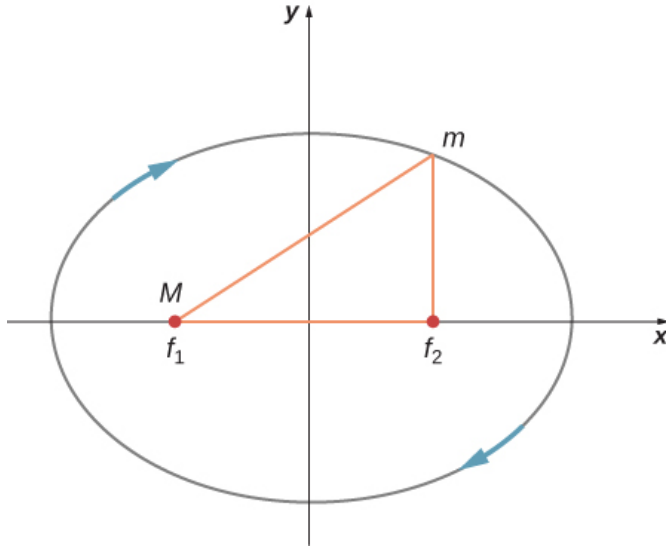
Exercise:

Problem: Are Kepler's laws purely descriptive, or do they contain causal information?

Exercise:

Problem:

In the diagram below for a satellite in an elliptical orbit about a much larger mass, indicate where its speed is the greatest and where it is the least. What conservation law dictates this behavior? Indicate the directions of the force, acceleration, and velocity at these points. Draw vectors for these same three quantities at the two points where the y -axis intersects (along the semi-minor axis) and from this determine whether the speed is increasing decreasing, or at a max/min.



Solution:

The speed is greatest where the satellite is closest to the large mass and least where farther away—at the periaapsis and apoapsis, respectively. It is conservation of angular momentum that governs this relationship. But it can also be gleaned from conservation of energy, the kinetic energy must be greatest where the gravitational potential energy is the least (most negative). The force, and hence acceleration, is always directed towards M in the diagram, and the velocity is always tangent to the path at all points. The acceleration vector has a tangential component along the direction of the velocity at the upper location on the y -axis; hence, the satellite is speeding up. Just the opposite is true at the lower position.

Problems**Exercise:****Problem:**

Calculate the mass of the Sun based on data for average Earth's orbit and compare the value obtained with the Sun's commonly listed value of 1.989×10^{30} kg.

Solution:

1.98×10^{30} kg; The values are the same within 0.05%.

Exercise:

Problem:

Io orbits Jupiter with an average radius of 421,700 km and a period of 1.769 days. Based upon these data, what is the mass of Jupiter?

Exercise:**Problem:**

The “mean” orbital radius listed for astronomical objects orbiting the Sun is typically not an integrated average but is calculated such that it gives the correct period when applied to the equation for circular orbits. Given that, what is the mean orbital radius in terms of aphelion and perihelion?

Solution:

Compare [\[link\]](#) and [\[link\]](#) to see that they differ only in that the circular radius, r , is replaced by the semi-major axis, a . Therefore, the mean radius is one-half the sum of the aphelion and perihelion, the same as the semi-major axis.

Exercise:**Problem:**

The perihelion of Halley’s comet is 0.586 AU and the aphelion is 17.8 AU. Given that its speed at perihelion is 55 km/s, what is the speed at aphelion ($1 \text{ AU} = 1.496 \times 10^{11} \text{ m}$)? (*Hint: You may use either conservation of energy or angular momentum, but the latter is much easier.*)

Exercise:**Problem:**

The perihelion of the comet Lagerkvist is 2.61 AU and it has a period of 7.36 years. Show that the aphelion for this comet is 4.95 AU.

Solution:

The semi-major axis, 3.78 AU is found from the equation for the period. This is one-half the sum of the aphelion and perihelion, giving an aphelion distance of 4.95 AU.

Exercise:**Problem:**

What is the ratio of the speed at perihelion to that at aphelion for the comet Lagerkvist in the previous problem?

Exercise:**Problem:**

Eros has an elliptical orbit about the Sun, with a perihelion distance of 1.13 AU and aphelion distance of 1.78 AU. What is the period of its orbit?

Solution:

1.75 years

Glossary

aphelion

farthest point from the Sun of an orbiting body; the corresponding term for the Moon's farthest point from Earth is the apogee

Kepler's first law

law stating that every planet moves along an ellipse, with the Sun located at a focus of the ellipse

Kepler's second law

law stating that a planet sweeps out equal areas in equal times, meaning it has a constant areal velocity

Kepler's third law

law stating that the square of the period is proportional to the cube of the semi-major axis of the orbit

perihelion

point of closest approach to the Sun of an orbiting body; the corresponding term for the Moon's closest approach to Earth is the perigee

Einstein's Theory of Gravity

By the end of this section, you will be able to:

- Describe how the theory of general relativity approaches gravitation
- Explain the principle of equivalence
- Calculate the Schwarzschild radius of an object
- Summarize the evidence for black holes

Newton's law of universal gravitation accurately predicts much of what we see within our solar system. Indeed, only Newton's laws have been needed to accurately send every space vehicle on its journey. The paths of Earth-crossing asteroids, and most other celestial objects, can be accurately determined solely with Newton's laws. Nevertheless, many phenomena have shown a discrepancy from what Newton's laws predict, including the orbit of Mercury and the effect that gravity has on light. In this section, we examine a different way of envisioning gravitation.

A Revolution in Perspective

In 1905, Albert Einstein published his theory of special relativity. This theory is discussed in great detail in [Relativity](#), so we say only a few words here. In this theory, no motion can exceed the speed of light—it is the speed limit of the Universe. This simple fact has been verified in countless experiments. However, it has incredible consequences—space and time are no longer absolute. Two people moving relative to one another do not agree on the length of objects or the passage of time. Almost all of the mechanics you learned in previous chapters, while remarkably accurate even for speeds of many thousands of miles per second, begin to fail when approaching the speed of light.

This speed limit on the Universe was also a challenge to the inherent assumption in Newton's law of gravitation that gravity is an **action-at-a-distance force**. That is, without physical contact, any change in the position of one mass is instantly communicated to all other masses. This assumption does not come from any first principle, as Newton's theory simply does not address the question. (The same was believed of electromagnetic forces, as well. It is fair to say that most scientists were not completely comfortable with the action-at-a-distance concept.)

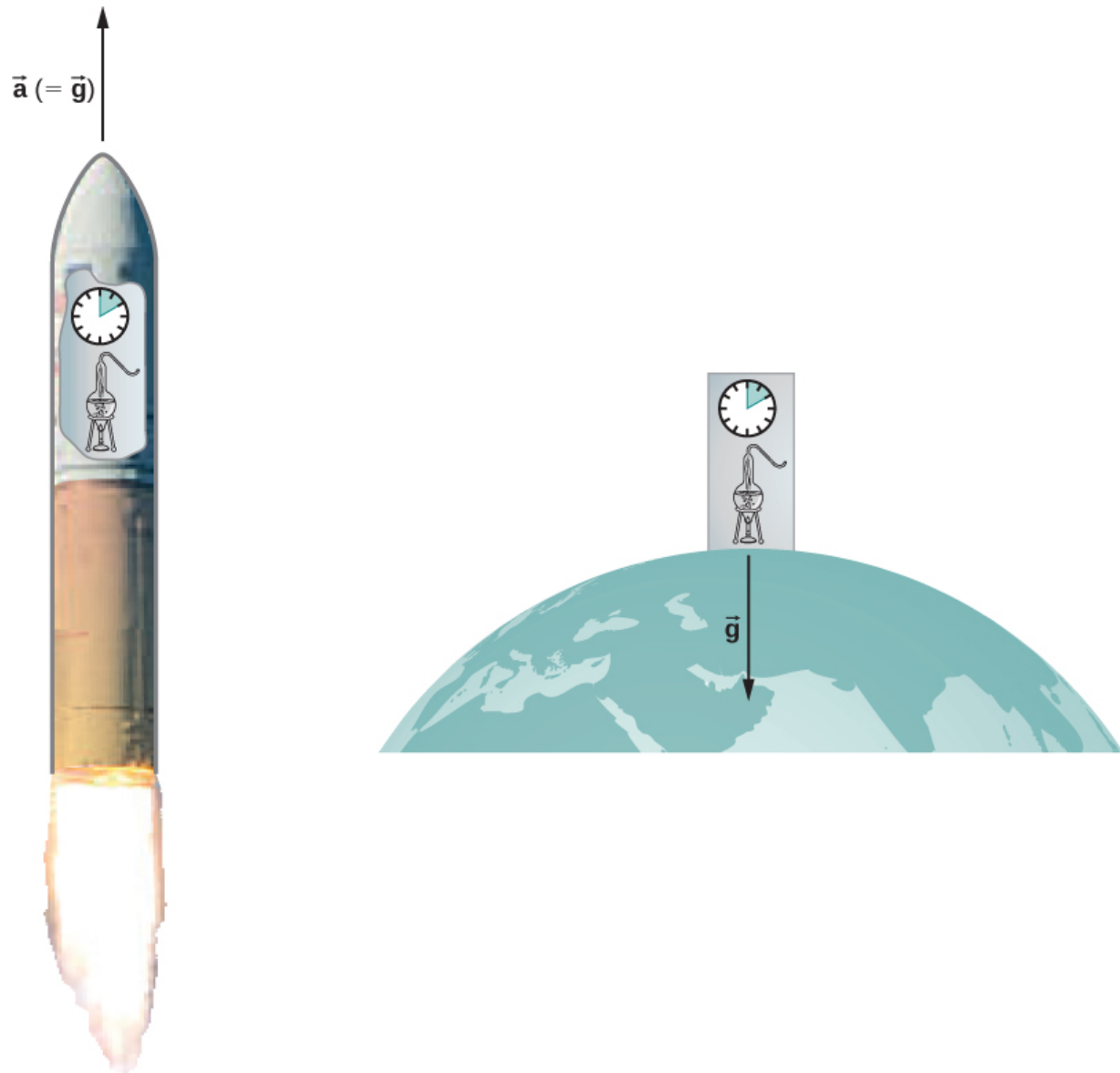
A second assumption also appears in Newton's law of gravitation [\[link\]](#). The masses are assumed to be exactly the same as those used in Newton's second law, $\vec{F} = m\vec{a}$. We made that assumption in many of our derivations in this chapter. Again, there is no underlying principle that this must be, but experimental results are consistent with this assumption. In Einstein's subsequent **theory of general relativity** (1916), both of

these issues were addressed. His theory was a theory of **space-time** geometry and how mass (and acceleration) distort and interact with that space-time. It was not a theory of gravitational forces. The mathematics of the general theory is beyond the scope of this text, but we can look at some underlying principles and their consequences.

The Principle of Equivalence

Einstein came to his general theory in part by wondering why someone who was free falling did not feel his or her weight. Indeed, it is common to speak of astronauts orbiting Earth as being weightless, despite the fact that Earth's gravity is still quite strong there. In Einstein's general theory, there is no difference between free fall and being weightless. This is called the **principle of equivalence**. The equally surprising corollary to this is that there is no difference between a uniform gravitational field and a uniform acceleration in the absence of gravity. Let's focus on this last statement. Although a perfectly uniform gravitational field is not feasible, we can approximate it very well.

Within a reasonably sized laboratory on Earth, the gravitational field \vec{g} is essentially uniform. The corollary states that any physical experiments performed there have the identical results as those done in a laboratory accelerating at $\vec{a} = \vec{g}$ in deep space, well away from all other masses. [\[link\]](#) illustrates the concept.



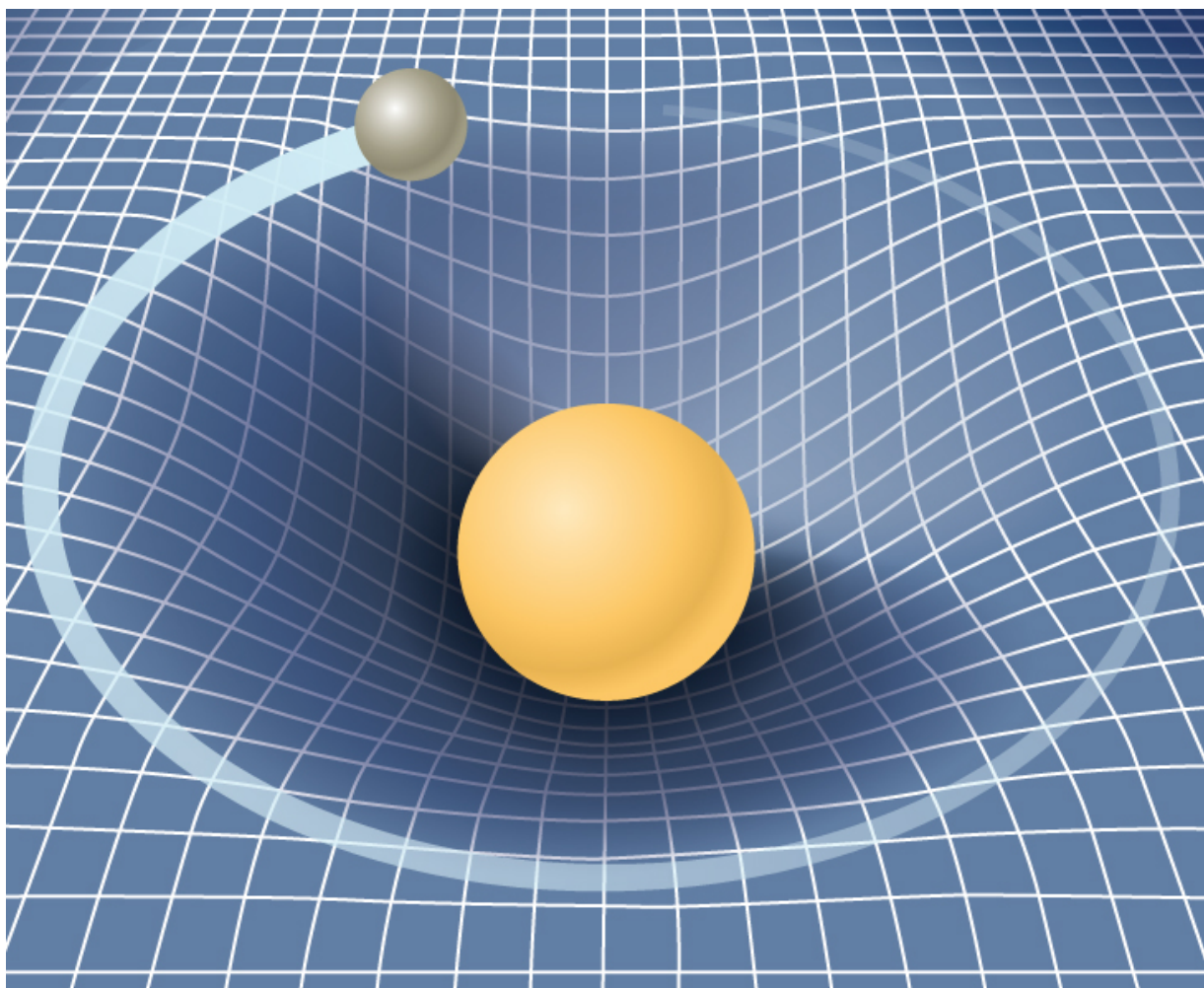
According to the principle of equivalence, the results of all experiments performed in a laboratory in a uniform gravitational field are identical to the results of the same experiments performed in a uniformly accelerating laboratory.

How can these two apparently fundamentally different situations be the same? The answer is that gravitation is not a force between two objects but is the result of each object responding to the effect that the other has on the space-time surrounding it. A uniform gravitational field and a uniform acceleration have exactly the same effect on space-time.

A Geometric Theory of Gravity

Euclidian geometry assumes a “flat” space in which, among the most commonly known attributes, a straight line is the shortest distance between two points, the sum of the angles of all triangles must be 180 degrees, and parallel lines never intersect. **Non-Euclidean geometry** was not seriously investigated until the nineteenth century, so it is not surprising that Euclidean space is inherently assumed in all of Newton’s laws.

The general theory of relativity challenges this long-held assumption. Only empty space is flat. The presence of mass—or energy, since relativity does not distinguish between the two—distorts or curves space and time, or space-time, around it. The motion of any other mass is simply a response to this curved space-time. [\[link\]](#) is a two-dimensional representation of a smaller mass orbiting in response to the distorted space created by the presence of a larger mass. In a more precise but confusing picture, we would also see space distorted by the orbiting mass, and both masses would be in motion in response to the total distortion of space. Note that the figure is a representation to help visualize the concept. These are distortions in our three-dimensional space and time. We do not see them as we would a dimple on a ball. We see the distortion only by careful measurements of the motion of objects and light as they move through space.



A smaller mass orbiting in the distorted space-time of a larger mass. In fact, all mass or energy distorts space-time.

For weak gravitational fields, the results of general relativity do not differ significantly from Newton's law of gravitation. But for intense gravitational fields, the results diverge, and general relativity has been shown to predict the correct results. Even in our Sun's relatively weak gravitational field at the distance of Mercury's orbit, we can observe the effect. Starting in the mid-1800s, Mercury's elliptical orbit has been carefully measured. However, although it is elliptical, its motion is complicated by the fact that the perihelion position of the ellipse slowly advances. Most of the advance is due to the gravitational pull of other planets, but a small portion of that advancement could not be accounted for by Newton's law. At one time, there was even a search for a "companion" planet that would explain the discrepancy. But general relativity correctly predicts the measurements. Since then, many measurements, such as the

deflection of light of distant objects by the Sun, have verified that general relativity correctly predicts the observations.

We close this discussion with one final comment. We have often referred to distortions of space-time or distortions in both space and time. In both special and general relativity, the dimension of time has equal footing with each spatial dimension (differing in its place in both theories only by an ultimately unimportant scaling factor). Near a very large mass, not only is the nearby space “stretched out,” but time is dilated or “slowed.” We discuss these effects more in the next section.

Black Holes

Einstein’s theory of gravitation is expressed in one deceptively simple-looking tensor equation (tensors are a generalization of scalars and vectors), which expresses how a mass determines the curvature of space-time around it. The solutions to that equation yield one of the most fascinating predictions: the **black hole**. The prediction is that if an object is sufficiently dense, it will collapse in upon itself and be surrounded by an **event horizon** from which nothing can escape. The name “black hole,” which was coined by astronomer John Wheeler in 1969, refers to the fact that light cannot escape such an object. Karl Schwarzschild was the first person to note this phenomenon in 1916, but at that time, it was considered mostly to be a mathematical curiosity.

Surprisingly, the idea of a massive body from which light cannot escape dates back to the late 1700s. Independently, John Michell and Pierre-Simon Laplace used Newton’s law of gravitation to show that light leaving the surface of a star with enough mass could not escape. Their work was based on the fact that the speed of light had been measured by Ole Rømer in 1676. He noted discrepancies in the data for the orbital period of the moon Io about Jupiter. Rømer realized that the difference arose from the relative positions of Earth and Jupiter at different times and that he could find the speed of light from that difference. Michell and Laplace both realized that since light had a finite speed, there could be a star massive enough that the escape speed from its surface could exceed that speed. Hence, light always would fall back to the star. Oddly, observers far enough away from the very largest stars would not be able to see them, yet they could see a smaller star from the same distance.

Recall that in [Gravitational Potential Energy and Total Energy](#), we found that the escape speed, given by [\[link\]](#), is independent of the mass of the object escaping. Even though the nature of light was not fully understood at the time, the mass of light, if it had any, was not relevant. Hence, [\[link\]](#) should be valid for light. Substituting c , the speed of light, for the escape velocity, we have

Equation:

$$v_{\text{esc}} = c = \sqrt{\frac{2GM}{R}}.$$

Thus, we only need values for R and M such that the escape velocity exceeds c , and then light will not be able to escape. Michell posited that if a star had the density of our Sun and a radius that extended just beyond the orbit of Mars, then light would not be able to escape from its surface. He also conjectured that we would still be able to detect such a star from the gravitational effect it would have on objects around it. This was an insightful conclusion, as this is precisely how we infer the existence of such objects today. While we have yet to, and may never, visit a black hole, the circumstantial evidence for them has become so compelling that few astronomers doubt their existence.

Before we examine some of that evidence, we turn our attention back to Schwarzschild's solution to the tensor equation from general relativity. In that solution arises a critical radius, now called the **Schwarzschild radius** (R_S). For any mass M , if that mass were compressed to the extent that its radius becomes less than the Schwarzschild radius, then the mass will collapse to a singularity, and anything that passes inside that radius cannot escape. Once inside R_S , the arrow of time takes all things to the singularity. (In a broad mathematical sense, a singularity is where the value of a function goes to infinity. In this case, it is a point in space of zero volume with a finite mass. Hence, the mass density and gravitational energy become infinite.) The Schwarzschild radius is given by

Note:

Equation:

$$R_S = \frac{2GM}{c^2}.$$

If you look at our escape velocity equation with $v_{\text{esc}} = c$, you will notice that it gives precisely this result. But that is merely a fortuitous accident caused by several incorrect assumptions. One of these assumptions is the use of the *incorrect* classical expression for the kinetic energy for light. Just how dense does an object have to be in order to turn into a black hole?

Example:**Calculating the Schwarzschild Radius**

Calculate the Schwarzschild radius for both the Sun and Earth. Compare the density of the nucleus of an atom to the density required to compress Earth's mass uniformly to its Schwarzschild radius. The density of a nucleus is about $2.3 \times 10^{17} \text{ kg/m}^3$.

Strategy

We use [\[link\]](#) for this calculation. We need only the masses of Earth and the Sun, which we obtain from the astronomical data given in [Appendix D](#).

Solution

Substituting the mass of the Sun, we have

Equation:

$$R_S = \frac{2GM}{c^2} = \frac{2(6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2)(1.99 \times 10^{30} \text{ kg})}{(3.0 \times 10^8 \text{ m/s})^2} = 2.95 \times 10^3 \text{ m}.$$

This is a diameter of only about 6 km. If we use the mass of Earth, we get $R_S = 8.85 \times 10^{-3} \text{ m}$. This is a diameter of less than 2 cm! If we pack Earth's mass into a sphere with the radius $R_S = 8.85 \times 10^{-3} \text{ m}$, we get a density of

Equation:

$$\rho = \frac{\text{mass}}{\text{volume}} = \frac{5.97 \times 10^{24} \text{ kg}}{(\frac{4}{3}\pi)(8.85 \times 10^{-3} \text{ m})^3} = 2.06 \times 10^{30} \text{ kg/m}^3.$$

Significance

A **neutron star** is the most compact object known—outside of a black hole itself. The neutron star is composed of neutrons, with the density of an atomic nucleus, and, like many black holes, is believed to be the remnant of a supernova—a star that explodes at the end of its lifetime. To create a black hole from Earth, we would have to compress it to a density thirteen orders of magnitude greater than that of a neutron star. This process would require unimaginable force. There is no known mechanism that could cause an Earth-sized object to become a black hole. For the Sun, you should be able to show that it would have to be compressed to a density only about 80 times that of a nucleus. (Note: Once the mass is compressed within its Schwarzschild radius, general relativity dictates that it will collapse to a singularity. These calculations merely show the density we must achieve to initiate that collapse.)

Note:**Exercise:**

Problem:

Check Your Understanding Consider the density required to make Earth a black hole compared to that required for the Sun. What conclusion can you draw from this comparison about what would be required to create a black hole? Would you expect the Universe to have many black holes with small mass?

Solution:

Given the incredible density required to force an Earth-sized body to become a black hole, we do not expect to see such small black holes. Even a body with the mass of our Sun would have to be compressed by a factor of 80 beyond that of a neutron star. It is believed that stars of this size cannot become black holes. However, for stars with a few solar masses, it is believed that gravitational collapse at the end of a star's life could form a black hole. As we will discuss later, it is now believed that black holes are common at the center of galaxies. These galactic black holes typically contain the mass of many millions of stars.

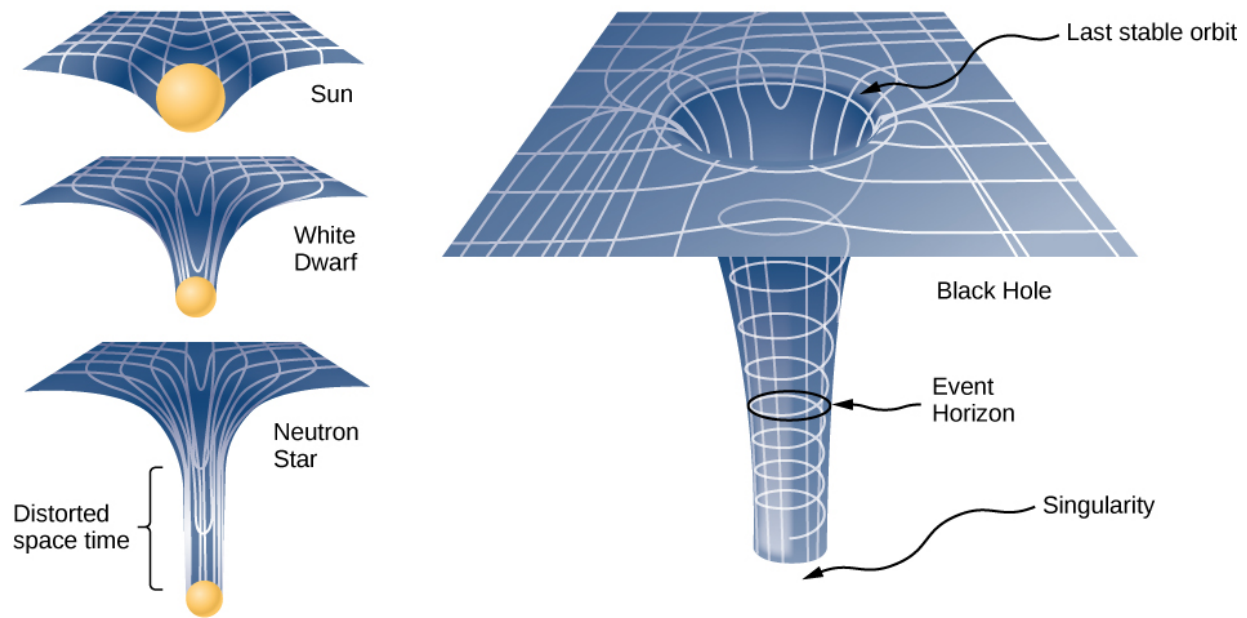
The event horizon

The Schwarzschild radius is also called the event horizon of a black hole. We noted that both space and time are stretched near massive objects, such as black holes. [\[link\]](#) illustrates that effect on space. The distortion caused by our Sun is actually quite small, and the diagram is exaggerated for clarity. Consider the neutron star, described in [\[link\]](#). Although the distortion of space-time at the surface of a neutron star is very high, the radius is still larger than its Schwarzschild radius. Objects could still escape from its surface.

However, if a neutron star gains additional mass, it would eventually collapse, shrinking beyond the Schwarzschild radius. Once that happens, the entire mass would be pulled, inevitably, to a singularity. In the diagram, space is stretched to infinity. Time is also stretched to infinity. As objects fall toward the event horizon, we see them approaching ever more slowly, but never reaching the event horizon. As outside observers, we never see objects pass through the event horizon—effectively, time is stretched to a stop.

Note:

Visit this [site](#) to view an animated example of these spatial distortions.



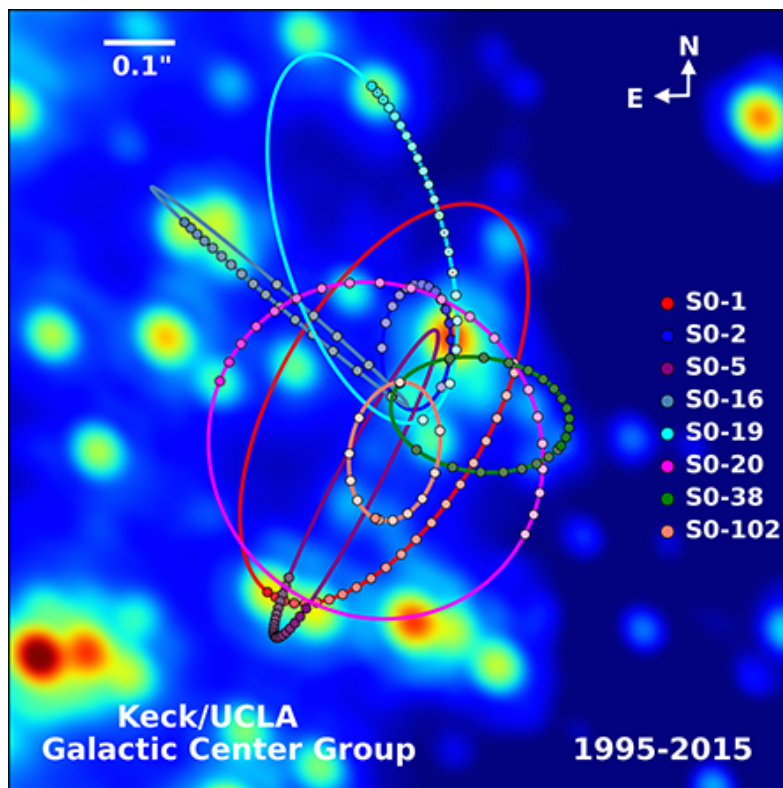
The space distortion becomes more noticeable around increasingly larger masses. Once the mass density reaches a critical level, a black hole forms and the fabric of space-time is torn. The curvature of space is greatest at the surface of each of the first three objects shown and is finite. The curvature then decreases (not shown) to zero as you move to the center of the object. But the black hole is different. The curvature becomes infinite: The surface has collapsed to a singularity, and the cone extends to infinity. (Note: These diagrams are not to any scale.) (credit: modification of work by NASA)

The evidence for black holes

Not until the 1960s, when the first neutron star was discovered, did interest in the existence of black holes become renewed. Evidence for black holes is based upon several types of observations, such as radiation analysis of X-ray binaries, gravitational lensing of the light from distant galaxies, and the motion of visible objects around invisible partners. We will focus on these later observations as they relate to what we have learned in this chapter. Although light cannot escape from a black hole for us to see, we can nevertheless see the gravitational effect of the black hole on surrounding masses.

The closest, and perhaps most dramatic, evidence for a black hole is at the center of our Milky Way galaxy. The UCLA Galactic Group, using data obtained by the W. M.

Keck telescopes, has determined the orbits of several stars near the center of our galaxy. Some of that data is shown in [\[link\]](#). The orbits of two stars are highlighted. From measurements of the periods and sizes of their orbits, it is estimated that they are orbiting a mass of approximately 4 million solar masses. Note that the mass must reside in the region created by the intersection of the ellipses of the stars. The region in which that mass must reside would fit inside the orbit of Mercury—yet nothing is seen there in the visible spectrum.



Paths of stars orbiting about a mass at the center of our Milky Way galaxy. From their motion, it is estimated that a black hole of about 4 million solar masses resides at the center. (credit: modification of work by UCLA Galactic Center Group – W.M. Keck Observatory Laser Team)

The physics of stellar creation and evolution is well established. The ultimate source of energy that makes stars shine is the self-gravitational energy that triggers fusion. The general behavior is that the more massive a star, the brighter it shines and the

shorter it lives. The logical inference is that a mass that is 4 million times the mass of our Sun, confined to a very small region, and that cannot be seen, has no viable interpretation other than a black hole. Extragalactic observations strongly suggest that black holes are common at the center of galaxies.

Note:

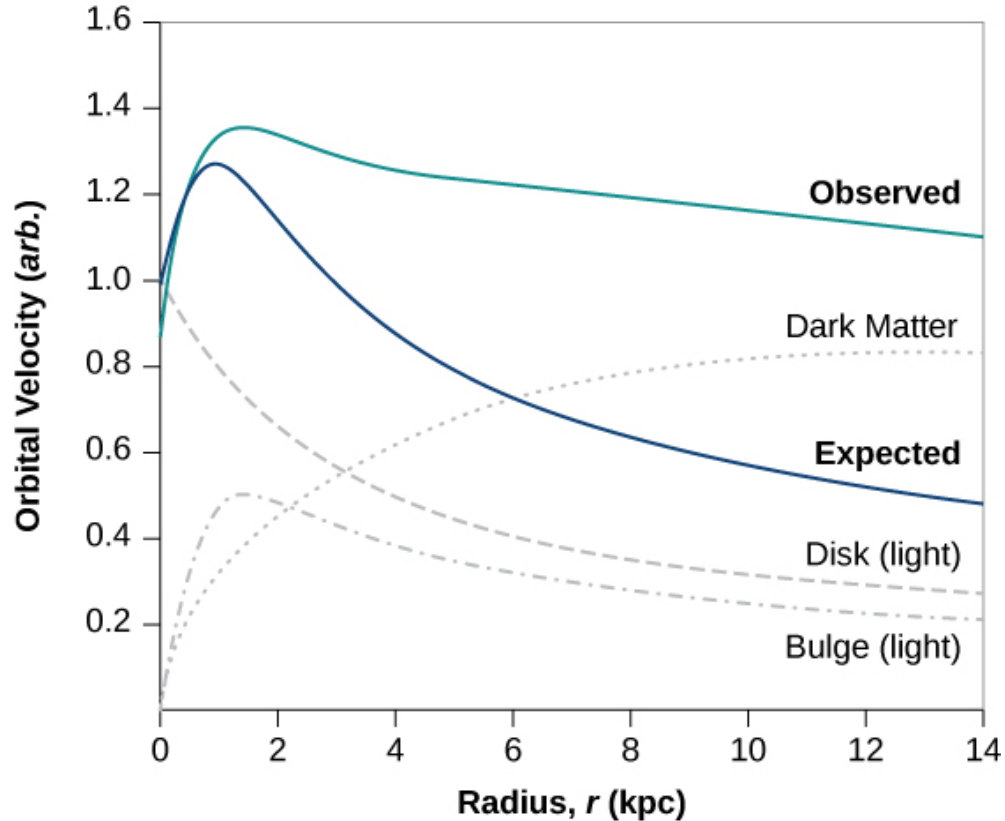
Visit the [UCLA Galactic Center Group main page](#) for information on X-ray binaries and gravitational lensing. Visit this [page](#) to view a three-dimensional visualization of the stars orbiting near the center of our galaxy, where the animation is near the bottom of the page.

Dark matter

Stars orbiting near the very heart of our galaxy provide strong evidence for a black hole there, but the orbits of stars far from the center suggest another intriguing phenomenon that is observed indirectly as well. Recall from [Gravitation Near Earth's Surface](#) that we can consider the mass for spherical objects to be located at a point at the center for calculating their gravitational effects on other masses. Similarly, we can treat the total mass that lies within the orbit of any star in our galaxy as being located at the center of the Milky Way disc. We can estimate that mass from counting the visible stars and include in our estimate the mass of the black hole at the center as well.

But when we do that, we find the orbital speed of the stars is far too fast to be caused by that amount of matter. [\[link\]](#) shows the orbital velocities of stars as a function of their distance from the center of the Milky Way. The blue line represents the velocities we would expect from our estimates of the mass, whereas the green curve is what we get from direct measurements. Apparently, there is a lot of matter we don't see, estimated to be about five times as much as what we do see, so it has been dubbed dark matter. Furthermore, the velocity profile does not follow what we expect from the observed distribution of visible stars. Not only is the estimate of the total mass inconsistent with the data, but the expected distribution is inconsistent as well. And this phenomenon is not restricted to our galaxy, but seems to be a feature of all galaxies. In fact, the issue was first noted in the 1930s when galaxies within clusters were measured to be orbiting about the center of mass of those clusters faster than they should based upon visible mass estimates.

Galaxy Rotation Curve



The blue curve shows the expected orbital velocity of stars in the Milky Way based upon the visible stars we can see. The green curve shows that the actual velocities are higher, suggesting additional matter that cannot be seen. (credit: modification of work by Matthew Newby)

There are two prevailing ideas of what this matter could be—WIMPs and MACHOs. WIMPs stands for weakly interacting massive particles. These particles (neutrinos are one example) interact very weakly with ordinary matter and, hence, are very difficult to detect directly. MACHOs stands for massive compact halo objects, which are composed of ordinary baryonic matter, such as neutrons and protons. There are unresolved issues with both of these ideas, and far more research will be needed to solve the mystery.

Summary

- According to the theory of general relativity, gravity is the result of distortions in space-time created by mass and energy.
- The principle of equivalence states that both mass and acceleration distort space-time and are indistinguishable in comparable circumstances.
- Black holes, the result of gravitational collapse, are singularities with an event horizon that is proportional to their mass.
- Evidence for the existence of black holes is still circumstantial, but the amount of that evidence is overwhelming.

Key Equations

Newton's law of gravitation	$\vec{\mathbf{F}}_{12} = G \frac{m_1 m_2}{r^2} \hat{\mathbf{r}}_{12}$
Acceleration due to gravity at the surface of Earth	$g = G \frac{M_E}{r_2^2}$
Gravitational potential energy beyond Earth	$U = -\frac{GM_E m}{r}$
Conservation of energy	$\frac{1}{2} m v_1^2 - \frac{GMm}{r_1} = \frac{1}{2} m v_2^2 - \frac{GMm}{r_2}$
Escape velocity	$v_{\text{esc}} = \sqrt{\frac{2GM}{R}}$
Orbital speed	$v_{\text{orbit}} = \sqrt{\frac{GM_E}{r}}$
Orbital period	$T = 2\pi \sqrt{\frac{r^3}{GM_E}}$
Energy in circular orbit	$E = K + U = -\frac{GmM_E}{2r}$
Conic sections	$\frac{\alpha}{r} = 1 + e \cos \theta$
Kepler's third law	$T^2 = \frac{4\pi^2}{GM} a^3$

Schwarzschild radius

$$R_S = \frac{2GM}{c^2}$$

Conceptual Questions

Exercise:

Problem:

The principle of equivalence states that all experiments done in a lab in a uniform gravitational field cannot be distinguished from those done in a lab that is not in a gravitational field but is uniformly accelerating. For the latter case, consider what happens to a laser beam at some height shot perfectly horizontally to the floor, across the accelerating lab. (View this from a nonaccelerating frame outside the lab.) Relative to the height of the laser, where will the laser beam hit the far wall? What does this say about the effect of a gravitational field on light? Does the fact that light has no mass make any difference to the argument?

Solution:

The laser beam will hit the far wall at a lower elevation than it left, as the floor is accelerating upward. Relative to the lab, the laser beam “falls.” So we would expect this to happen in a gravitational field. The mass of light, or even an object with mass, is not relevant.

Exercise:

Problem:

As a person approaches the Schwarzschild radius of a black hole, outside observers see all the processes of that person (their clocks, their heart rate, etc.) slowing down, and coming to a halt as they reach the Schwarzschild radius. (The person falling into the black hole sees their own processes unaffected.) But the speed of light is the same everywhere for all observers. What does this say about space as you approach the black hole?

Problems

Exercise:

Problem:

What is the Schwarzschild radius for the black hole at the center of our galaxy if it has the mass of 4 million solar masses?

Solution:

$$1.19 \times 10^7 \text{ km}$$

Exercise:**Problem:**

What would be the Schwarzschild radius, in light years, if our Milky Way galaxy of 100 billion stars collapsed into a black hole? Compare this to our distance from the center, about 13,000 light years.

Additional Problems**Exercise:****Problem:**

A neutron star is a cold, collapsed star with nuclear density. A particular neutron star has a mass twice that of our Sun with a radius of 12.0 km. (a) What would be the weight of a 100-kg astronaut on standing on its surface? (b) What does this tell us about landing on a neutron star?

Solution:

a. $1.85 \times 10^{14} \text{ N}$; b. Don't do it!

Exercise:**Problem:**

(a) How far from the center of Earth would the net gravitational force of Earth and the Moon on an object be zero? (b) Setting the *magnitudes* of the forces equal should result in two answers from the quadratic. Do you understand why there are two positions, but only one where the net force is zero?

Exercise:**Problem:**

How far from the center of the Sun would the net gravitational force of Earth and the Sun on a spaceship be zero?

Solution:

$$1.49 \times 10^8 \text{ km}$$

Exercise:**Problem:**

Calculate the values of g at Earth's surface for the following changes in Earth's properties: (a) its mass is doubled and its radius is halved; (b) its mass density is doubled and its radius is unchanged; (c) its mass density is halved and its mass is unchanged.

Exercise:**Problem:**

Suppose you can communicate with the inhabitants of a planet in another solar system. They tell you that on their planet, whose diameter and mass are 5.0×10^3 km and 3.6×10^{23} kg, respectively, the record for the high jump is 2.0 m. Given that this record is close to 2.4 m on Earth, what would you conclude about your extraterrestrial friends' jumping ability?

Solution:

The value of g for this planet is 3.8 m/s^2 , which is about one-fourth that of Earth. So they are weak high jumpers.

Exercise:**Problem:**

(a) Suppose that your measured weight at the equator is one-half your measured weight at the pole on a planet whose mass and diameter are equal to those of Earth. What is the rotational period of the planet? (b) Would you need to take the shape of this planet into account?

Exercise:**Problem:**

A body of mass 100 kg is weighed at the North Pole and at the equator with a spring scale. What is the scale reading at these two points? Assume that $g = 9.83 \text{ m/s}^2$ at the pole.

Solution:

At the North Pole, 983 N; at the equator, 980 N

Exercise:

Problem:

Find the speed needed to escape from the solar system starting from the surface of Earth. Assume there are no other bodies involved and do not account for the fact that Earth is moving in its orbit. [Hint: [link](#) does not apply. Use [link](#) and include the potential energy of both Earth and the Sun.

Exercise:**Problem:**

Consider the previous problem and include the fact that Earth has an orbital speed about the Sun of 29.8 km/s. (a) What speed relative to Earth would be needed and in what direction should you leave Earth? (b) What will be the shape of the trajectory?

Solution:

a. The escape velocity is still 43.6 km/s. By launching from Earth in the direction of Earth's tangential velocity, you need $43.6 - 29.8 = 13.8$ km/s relative to Earth. b. The total energy is zero and the trajectory is a parabola.

Exercise:**Problem:**

A comet is observed 1.50 AU from the Sun with a speed of 24.3 km/s. Is this comet in a bound or unbound orbit?

Exercise:**Problem:**

An asteroid has speed 15.5 km/s when it is located 2.00 AU from the sun. At its closest approach, it is 0.400 AU from the Sun. What is its speed at that point?

Solution:

61.5 km/s

Exercise:

Problem:

Space debris left from old satellites and their launchers is becoming a hazard to other satellites. (a) Calculate the speed of a satellite in an orbit 900 km above Earth's surface. (b) Suppose a loose rivet is in an orbit of the same radius that intersects the satellite's orbit at an angle of 90° . What is the velocity of the rivet relative to the satellite just before striking it? (c) If its mass is 0.500 g, and it comes to rest inside the satellite, how much energy in joules is generated by the collision? (Assume the satellite's velocity does not change appreciably, because its mass is much greater than the rivet's.)

Exercise:**Problem:**

A satellite of mass 1000 kg is in circular orbit about Earth. The radius of the orbit of the satellite is equal to two times the radius of Earth. (a) How far away is the satellite? (b) Find the kinetic, potential, and total energies of the satellite.

Solution:

a. 1.3×10^7 m; b. 1.56×10^{10} J; -3.12×10^{10} J; -1.56×10^{10} J

Exercise:**Problem:**

After Ceres was promoted to a dwarf planet, we now recognize the largest known asteroid to be Vesta, with a mass of 2.67×10^{20} kg and a diameter ranging from 578 km to 458 km. Assuming that Vesta is spherical with radius 520 km, find the approximate escape velocity from its surface.

Exercise:**Problem:**

(a) Given the asteroid Vesta which has a diameter of 520 km and mass of 2.67×10^{20} kg, what would be the orbital period for a space probe in a circular orbit of 10.0 km from its surface? (b) Why is this calculation marginally useful at best?

Solution:

a. 6.24×10^3 s or about 1.8 hours. This was using the 520 km average diameter.
b. Vesta is clearly not very spherical, so you would need to be above the largest dimension, nearly 580 km. More importantly, the nonspherical nature would

disturb the orbit very quickly, so this calculation would not be very accurate even for one orbit.

Exercise:

Problem:

What is the orbital velocity of our solar system about the center of the Milky Way? Assume that the mass within a sphere of radius equal to our distance away from the center is about a 100 billion solar masses. Our distance from the center is 27,000 light years.

Exercise:

Problem:

(a) Using the information in the previous problem, what velocity do you need to escape the Milky Way galaxy from our present position? (b) Would you need to accelerate a spaceship to this speed relative to Earth?

Solution:

a. 323 km/s; b. No, you need only the difference between the solar system's orbital speed and escape speed, so about $323 - 228 = 95$ km/s.

Exercise:

Problem:

Circular orbits in [\[link\]](#) for conic sections must have eccentricity zero. From this, and using Newton's second law applied to centripetal acceleration, show that the value of α in [\[link\]](#) is given by $\alpha = \frac{L^2}{GMm^2}$ where L is the angular momentum of the orbiting body. The value of α is constant and given by this expression regardless of the type of orbit.

Exercise:

Problem:

Show that for eccentricity equal to one in [\[link\]](#) for conic sections, the path is a parabola. Do this by substituting Cartesian coordinates, x and y , for the polar coordinates, r and θ , and showing that it has the general form for a parabola, $x = ay^2 + by + c$.

Solution:

Setting $e = 1$, we have $\frac{\alpha}{r} = 1 + \cos\theta \rightarrow \alpha = r + r\cos\theta = r + x$; hence, $r^2 = x^2 + y^2 = (\alpha - x)^2$. Expand and collect to show $x = \frac{1}{-2\alpha}y^2 + \frac{\alpha}{2}$.

Exercise:

Problem:

Using the technique shown in [Satellite Orbits and Energy](#), show that two masses m_1 and m_2 in circular orbits about their common center of mass, will have total energy $E = K + E = K_1 + K_2 - \frac{Gm_1m_2}{r} = -\frac{Gm_1m_2}{2r}$. We have shown the kinetic energy of both masses explicitly. (*Hint: The masses orbit at radii r_1 and r_2 , respectively, where $r = r_1 + r_2$. Be sure not to confuse the radius needed for centripetal acceleration with that for the gravitational force.*)

Exercise:

Problem:

Given the perihelion distance, p , and aphelion distance, q , for an elliptical orbit, show that the velocity at perihelion, v_p , is given by $v_p = \sqrt{\frac{2GM_{\text{Sun}}}{(q+p)} \frac{q}{p}}$. (*Hint: Use conservation of angular momentum to relate v_p and v_q , and then substitute into the conservation of energy equation.*)

Solution:

Substitute directly into the energy equation using $pv_p = qv_q$ from conservation of angular momentum, and solve for v_p .

Exercise:

Problem:

Comet P/1999 R1 has a perihelion of 0.0570 AU and aphelion of 4.99 AU. Using the results of the previous problem, find its speed at aphelion. (*Hint: The expression is for the perihelion. Use symmetry to rewrite the expression for aphelion.*)

Challenge Problems

Exercise:

Problem:

A tunnel is dug through the center of a perfectly spherical and airless planet of radius R . Using the expression for g derived in [Gravitation Near Earth's Surface](#) for a uniform density, show that a particle of mass m dropped in the tunnel will execute simple harmonic motion. Deduce the period of oscillation of m and show that it has the same period as an orbit at the surface.

Solution:

$g = \frac{4}{3} G \rho \pi r \rightarrow F = mg = \left[\frac{4}{3} G m \rho \pi \right] r$, and from $F = m \frac{d^2 r}{dt^2}$, we get $\frac{d^2 r}{dt^2} = \left[\frac{4}{3} G \rho \pi \right] r$ where the first term is ω^2 . Then $T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{3}{4G\rho\pi}}$ and if we substitute $\rho = \frac{M}{4/3\pi R^3}$, we get the same expression as for the period of orbit R .

Exercise:**Problem:**

Following the technique used in [Gravitation Near Earth's Surface](#), find the value of g as a function of the radius r from the center of a spherical shell planet of constant density ρ with inner and outer radii R_{in} and R_{out} . Find g for both $R_{\text{in}} < r < R_{\text{out}}$ and for $r < R_{\text{in}}$. Assuming the inside of the shell is kept airless, describe travel inside the spherical shell planet.

Exercise:**Problem:**

Show that the areal velocity for a circular orbit of radius r about a mass M is $\frac{\Delta A}{\Delta t} = \frac{1}{2} \sqrt{GM r}$. Does your expression give the correct value for Earth's areal velocity about the Sun?

Solution:

Using the mass of the Sun and Earth's orbital radius, the equation gives $2.24 \times 10^{15} \text{ m}^2/\text{s}$. The value of $\pi R_{\text{ES}}^2 / (1 \text{ year})$ gives the same value.

Exercise:

Problem:

Show that the period of orbit for two masses, m_1 and m_2 , in circular orbits of radii r_1 and r_2 , respectively, about their common center-of-mass, is given by $T = 2\pi\sqrt{\frac{r^3}{G(m_1+m_2)}}$ where $r = r_1 + r_2$. (Hint: The masses orbit at radii r_1 and r_2 , respectively where $r = r_1 + r_2$. Use the expression for the center-of-mass to relate the two radii and note that the two masses must have equal but opposite momenta. Start with the relationship of the period to the circumference and speed of orbit for one of the masses. Use the result of the previous problem using momenta in the expressions for the kinetic energy.)

Exercise:**Problem:**

Show that for small changes in height h , such that $h \ll R_E$, [\[link\]](#) reduces to the expression $\Delta U = mgh$.

Solution:

$\Delta U = U_f - U_i = -\frac{GM_E m}{r_f} + \frac{GM_E m}{r_i} = GM_E m \left(\frac{r_f - r_i}{r_f r_i} \right)$ where $h = r_f - r_i$.
If $h \ll R_E$, then $r_f r_i \approx R_E^2$, and upon substitution, we have

$\Delta U = GM_E m \left(\frac{h}{R_E^2} \right) = m \left(\frac{GM_E}{R_E^2} \right) h$ where we recognize the expression with the parenthesis as the definition of g .

Exercise:**Problem:**

Using [\[link\]](#), carefully sketch a free body diagram for the case of a simple pendulum hanging at latitude λ , labeling all forces acting on the point mass, m . Set up the equations of motion for equilibrium, setting one coordinate in the direction of the centripetal acceleration (toward P in the diagram), the other perpendicular to that. Show that the deflection angle ε , defined as the angle between the pendulum string and the radial direction toward the center of Earth, is given by the expression below. What is the deflection angle at latitude 45 degrees? Assume that Earth is a perfect sphere. $\tan(\lambda + \varepsilon) = \frac{g}{(g - \omega^2 R_E)} \tan \lambda$, where ω is the angular velocity of Earth.

Exercise:

Problem:

(a) Show that tidal force on a small object of mass m , defined as the *difference* in the gravitational force that would be exerted on m at a distance at the near and the far side of the object, due to the gravitation at a distance R from M , is given by $F_{\text{tidal}} = \frac{2GMm}{R^3} \Delta r$ where Δr is the distance between the near and far side and $\Delta r \ll R$. (b) Assume you are falling feet first into the black hole at the center of our galaxy. It has mass of 4 million solar masses. What would be the difference between the force at your head and your feet at the Schwarzschild radius (event horizon)? Assume your feet and head each have mass 5.0 kg and are 2.0 m apart. Would you survive passing through the event horizon?

Solution:

a. Find the difference in force,

$$F_{\text{tidal}} = \frac{2GMm}{R^3} \Delta r;$$

b. For the case given, using the Schwarzschild radius from a previous problem, we have a tidal force of 9.5×10^{-3} N. This won't even be noticed!

Exercise:**Problem:**

Find the Hohmann transfer velocities, $\Delta v_{\text{EllipseEarth}}$ and $\Delta v_{\text{EllipseMars}}$, needed for a trip to Mars. Use [\[link\]](#) to find the circular orbital velocities for Earth and Mars. Using [\[link\]](#) and the total energy of the ellipse (with semi-major axis a), given by $E = -\frac{GmM_s}{2a}$, find the velocities at Earth (perihelion) and at Mars (aphelion) required to be on the transfer ellipse. The difference, Δv , at each point is the velocity boost or transfer velocity needed.

Glossary

action-at-a-distance force

type of force exerted without physical contact

black hole

mass that becomes so dense, that it collapses in on itself, creating a singularity at the center surround by an event horizon

event horizon

location of the Schwarzschild radius and is the location near a black hole from within which no object, even light, can escape

neutron star

most compact object known—outside of a black hole itself

non-Euclidean geometry

geometry of curved space, describing the relationships among angles and lines on the surface of a sphere, hyperboloid, etc.

principle of equivalence

part of the general theory of relativity, it states that there no difference between free fall and being weightless, or a uniform gravitational field and uniform acceleration

Schwarzschild radius

critical radius (R_S) such that if a mass were compressed to the extent that its radius becomes less than the Schwarzschild radius, then the mass will collapse to a singularity, and anything that passes inside that radius cannot escape

space-time

concept of space-time is that time is essentially another coordinate that is treated the same way as any individual spatial coordinate; in the equations that represent both special and general relativity, time appears in the same context as do the spatial coordinates

theory of general relativity

Einstein's theory for gravitation and accelerated reference frames; in this theory, gravitation is the result of mass and energy distorting the space-time around it; it is also often referred to as Einstein's theory of gravity

Introduction

class="introduction"

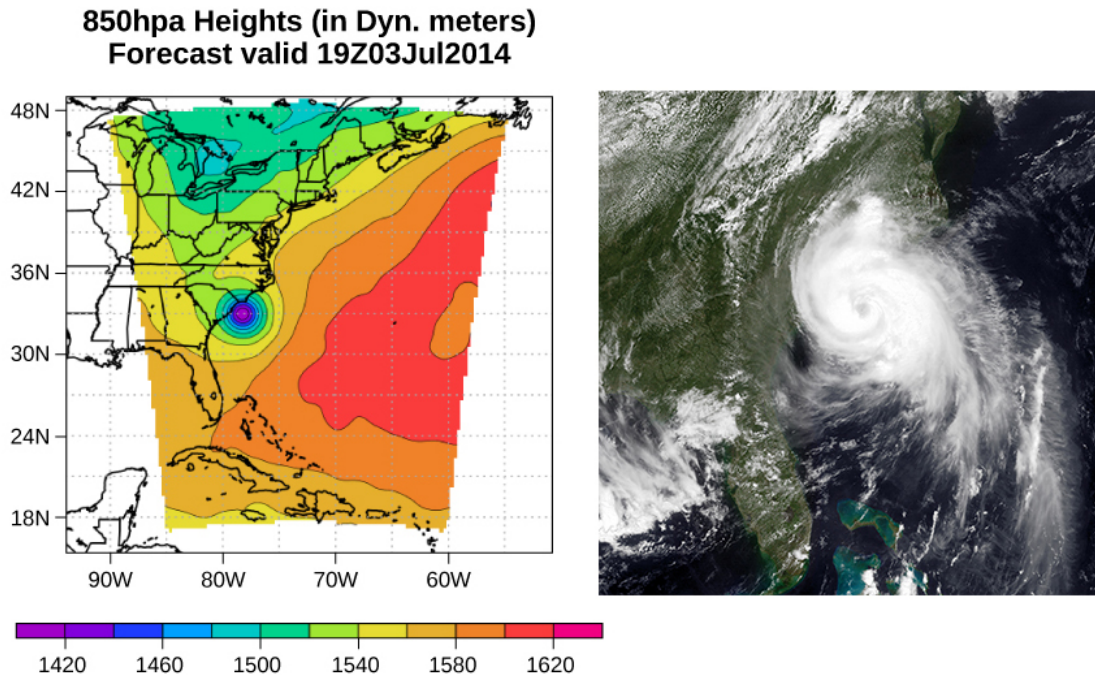
This pressure map (left) and satellite photo (right) were used to model the path and impact of Hurricane Arthur as it traveled up the East Coast of the United States in July 2014.

Computer models use force and energy equations to predict developing weather patterns.

Scientists numerically integrate these time-dependent equations, along with the energy budgets of long- and short-wave solar energy, to model changes

in the
atmosphere.
The pressure
map on the left
was created
using the
Weather
Research and
Forecasting
Model
designed at the
National
Center for
Atmospheric
Research. The
colors
represent the
height of the
850-mbar
pressure
surface. (credit
left:
modification of
work by The
National
Center for
Atmospheric
Research;
credit right:
modification of
work by NRL
Monterey
Marine
Meteorology
Division, The
National
Oceanic and

Atmospheric Administration)



Picture yourself walking along a beach on the eastern shore of the United States. The air smells of sea salt and the sun warms your body. Suddenly, an alert appears on your cell phone. A tropical depression has formed into a hurricane. Atmospheric pressure has fallen to nearly 15% below average. As a result, forecasters expect torrential rainfall, winds in excess of 100 mph, and millions of dollars in damage. As you prepare to evacuate, you wonder: How can such a small drop in pressure lead to such a severe change in the weather?

Pressure is a physical phenomenon that is responsible for much more than just the weather. Changes in pressure cause ears to “pop” during takeoff in an airplane. Changes in pressure can also cause scuba divers to suffer a sometimes fatal disorder known as the “bends,” which occurs when nitrogen dissolved in the water of the body at extreme depths returns to a gaseous state in the body as the diver surfaces. Pressure lies at the heart of the phenomena called buoyancy, which causes hot air balloons to rise and

ships to float. Before we can fully understand the role that pressure plays in these phenomena, we need to discuss the states of matter and the concept of density.

Fluids, Density, and Pressure

By the end of this section, you will be able to:

- State the different phases of matter
- Describe the characteristics of the phases of matter at the molecular or atomic level
- Distinguish between compressible and incompressible materials
- Define density and its related SI units
- Compare and contrast the densities of various substances
- Define pressure and its related SI units
- Explain the relationship between pressure and force
- Calculate force given pressure and area

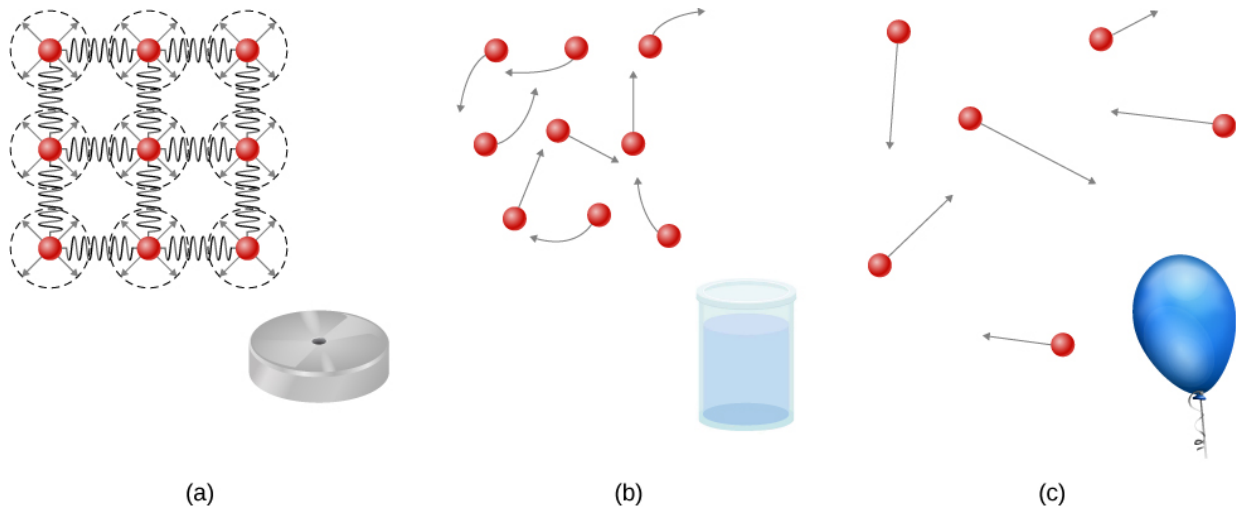
Matter most commonly exists as a solid, liquid, or gas; these states are known as the three common phases of matter. We will look at each of these phases in detail in this section.

Characteristics of Solids

Solids are rigid and have specific shapes and definite volumes. The atoms or molecules in a solid are in close proximity to each other, and there is a significant force between these molecules. Solids will take a form determined by the nature of these forces between the molecules. Although true solids are not incompressible, it nevertheless requires a large force to change the shape of a solid. In some cases, the force between molecules can cause the molecules to organize into a lattice as shown in [\[link\]](#). The structure of this three-dimensional lattice is represented as molecules connected by rigid bonds (modeled as stiff springs), which allow limited freedom for movement. Even a large force produces only small displacements in the atoms or molecules of the lattice, and the solid maintains its shape. Solids also resist shearing forces. (Shearing forces are forces applied tangentially to a surface, as described in [Static Equilibrium and Elasticity](#).)

Characteristics of Fluids

Liquids and gases are considered to be **fluids** because they yield to shearing forces, whereas solids resist them. Like solids, the molecules in a liquid are bonded to neighboring molecules, but possess many fewer of these bonds. The molecules in a liquid are not locked in place and can move with respect to each other. The distance between molecules is similar to the distances in a solid, and so liquids have definite volumes, but the shape of a liquid changes, depending on the shape of its container. Gases are not bonded to neighboring atoms and can have large separations between molecules. Gases have neither specific shapes nor definite volumes, since their molecules move to fill the container in which they are held ([\[link\]](#)).



(a) Atoms in a solid are always in close contact with neighboring atoms, held in place by forces represented here by springs. (b) Atoms in a liquid are also in close contact but can slide over one another. Forces between the atoms strongly resist attempts to compress the atoms. (c) Atoms in a gas move about freely and are separated by large distances. A gas must be held in a closed container to prevent it from expanding freely and escaping.

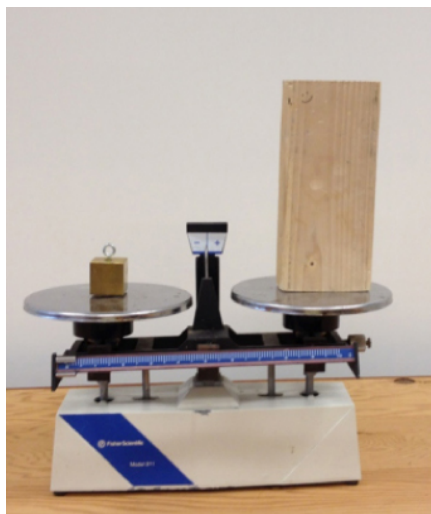
Liquids deform easily when stressed and do not spring back to their original shape once a force is removed. This occurs because the atoms or molecules in a liquid are free to slide about and change neighbors. That is, liquids flow (so they are a type of fluid), with the molecules held together by mutual attraction. When a liquid is placed in a container with no lid, it remains in the container. Because the atoms are closely packed, liquids, like solids, resist compression; an extremely large force is necessary to change the volume of a liquid.

In contrast, atoms in gases are separated by large distances, and the forces between atoms in a gas are therefore very weak, except when the atoms collide with one another. This makes gases relatively easy to compress and allows them to flow (which makes them fluids). When placed in an open container, gases, unlike liquids, will escape.

In this chapter, we generally refer to both gases and liquids simply as fluids, making a distinction between them only when they behave differently. There exists one other phase of matter, plasma, which exists at very high temperatures. At high temperatures, molecules may disassociate into atoms, and atoms disassociate into electrons (with negative charges) and protons (with positive charges), forming a plasma. Plasma will not be discussed in depth in this chapter because plasma has very different properties from the three other common phases of matter, discussed in this chapter, due to the strong electrical forces between the charges.

Density

Suppose a block of brass and a block of wood have exactly the same mass. If both blocks are dropped in a tank of water, why does the wood float and the brass sink ([link](#))? This occurs because the brass has a greater density than water, whereas the wood has a lower density than water.



(a)



(b)

(a) A block of brass and a block of wood both have the same weight and mass, but the block of wood has a much greater volume. (b) When placed in a fish tank filled with water, the cube of brass sinks and the block of wood floats. (The block of wood is the same in both pictures; it was turned on its side to fit on the scale.) (credit: modification of works by Joseph J. Trout, Stockton University)

Density is an important characteristic of substances. It is crucial, for example, in determining whether an object sinks or floats in a fluid.

Note:

Density

The average density of a substance or object is defined as its mass per unit volume,

Equation:

$$\rho = \frac{m}{V}$$

where the Greek letter ρ (rho) is the symbol for density, m is the mass, and V is the volume.

The SI unit of density is kg/m^3 . [\[link\]](#) lists some representative values. The cgs unit of density is the gram per cubic centimeter, g/cm^3 , where

Equation:

$$1 \text{ g/cm}^3 = 1000 \text{ kg/m}^3.$$

The metric system was originally devised so that water would have a density of 1 g/cm^3 , equivalent to 10^3 kg/m^3 . Thus, the basic mass unit, the kilogram, was first devised to be the mass of 1000 mL of

water, which has a volume of 1000 cm^3 .

Solids (0.0°C)		Liquids (0.0°C)		Gases (0.0°C , 101.3 kPa)	
Substance	$\rho(\text{kg/m}^3)$	Substance	$\rho(\text{kg/m}^3)$	Substance	$\rho(\text{kg/m}^3)$
Aluminum	2.70×10^3	Benzene	8.79×10^2	Air	1.29×10^0
Bone	1.90×10^3	Blood	1.05×10^3	Carbon dioxide	1.98×10^0
Brass	8.44×10^3	Ethyl alcohol	8.06×10^2	Carbon monoxide	1.25×10^0
Concrete	2.40×10^3	Gasoline	6.80×10^2	Helium	1.80×10^{-1}
Copper	8.92×10^3	Glycerin	1.26×10^3	Hydrogen	9.00×10^{-2}
Cork	2.40×10^2	Mercury	1.36×10^4	Methane	7.20×10^{-2}
Earth's crust	3.30×10^3	Olive oil	9.20×10^2	Nitrogen	1.25×10^0
Glass	2.60×10^3			Nitrous oxide	1.98×10^0
Gold	1.93×10^4			Oxygen	1.43×10^0
Granite	2.70×10^3				
Iron	7.86×10^3				
Lead	1.13×10^4				
Oak	7.10×10^2				
Pine	3.73×10^2				
Platinum	2.14×10^4				
Polystyrene	1.00×10^2				
Tungsten	1.93×10^4				

Solids (0.0 °C)		Liquids (0.0 °C)		Gases (0.0 °C, 101.3 kPa)	
Uranium	1.87×10^3				

Densities of Some Common Substances

As you can see by examining [\[link\]](#), the density of an object may help identify its composition. The density of gold, for example, is about 2.5 times the density of iron, which is about 2.5 times the density of aluminum. Density also reveals something about the phase of the matter and its substructure. Notice that the densities of liquids and solids are roughly comparable, consistent with the fact that their atoms are in close contact. The densities of gases are much less than those of liquids and solids, because the atoms in gases are separated by large amounts of empty space. The gases are displayed for a standard temperature of 0.0 °C and a standard pressure of 101.3 kPa, and there is a strong dependence of the densities on temperature and pressure. The densities of the solids and liquids displayed are given for the standard temperature of 0.0 °C and the densities of solids and liquids depend on the temperature. The density of solids and liquids normally increase with decreasing temperature.

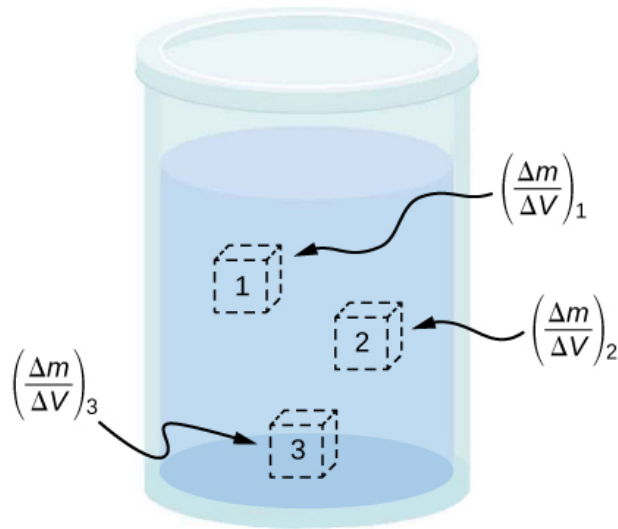
[\[link\]](#) shows the density of water in various phases and temperature. The density of water increases with decreasing temperature, reaching a maximum at 4.0 °C, and then decreases as the temperature falls below 4.0 °C. This behavior of the density of water explains why ice forms at the top of a body of water.

Substance	$\rho(\text{kg/m}^3)$
Ice (0 °C)	9.17×10^2
Water (0 °C)	9.998×10^2
Water (4 °C)	1.000×10^3
Water (20 °C)	9.982×10^2
Water (100 °C)	9.584×10^2
Steam (100 °C, 101.3 kPa)	1.670×10^2
Sea water (0 °C)	1.030×10^3

Densities of Water

The density of a substance is not necessarily constant throughout the volume of a substance. If the density is constant throughout a substance, the substance is said to be a homogeneous substance. A solid iron bar is an example of a homogeneous substance. The density is constant throughout, and the density of any sample of the substance is the same as its average density. If the density of a substance were not constant, the substance is said to be a heterogeneous substance. A chunk of Swiss cheese is an example

of a heterogeneous material containing both the solid cheese and gas-filled voids. The density at a specific location within a heterogeneous material is called *local density*, and is given as a function of location, $\rho = \rho(x, y, z)$ ([link](#)).



Density may vary throughout a heterogeneous mixture. Local density at a point is obtained from dividing mass by volume in a small volume around a given point.

Local density can be obtained by a limiting process, based on the average density in a small volume around the point in question, taking the limit where the size of the volume approaches zero,

Note:

Equation:

$$\rho = \lim_{\Delta V \rightarrow 0} \frac{\Delta m}{\Delta V}$$

where ρ is the density, m is the mass, and V is the volume.

Since gases are free to expand and contract, the densities of the gases vary considerably with temperature, whereas the densities of liquids vary little with temperature. Therefore, the densities of liquids are often treated as constant, with the density equal to the average density.

Density is a dimensional property; therefore, when comparing the densities of two substances, the units must be taken into consideration. For this reason, a more convenient, dimensionless quantity called the **specific gravity** is often used to compare densities. Specific gravity is defined as the ratio of the density of the material to the density of water at 4.0 °C and one atmosphere of pressure, which is 1000 kg/m³:

Equation:

$$\text{Specific gravity} = \frac{\text{Density of material}}{\text{Density of water}}.$$

The comparison uses water because the density of water is 1 g/cm^3 , which was originally used to define the kilogram. Specific gravity, being dimensionless, provides a ready comparison among materials without having to worry about the unit of density. For instance, the density of aluminum is 2.7 g/cm^3 (2700 kg/m^3), but its specific gravity is 2.7 , regardless of the unit of density. Specific gravity is a particularly useful quantity with regard to buoyancy, which we will discuss later in this chapter.

Pressure

You have no doubt heard the word ‘pressure’ used in relation to blood (high or low blood pressure) and in relation to weather (high- and low-pressure weather systems). These are only two of many examples of pressure in fluids. (Recall that we introduced the idea of pressure in [Static Equilibrium and Elasticity](#), in the context of bulk stress and strain.)

Note:**Pressure**

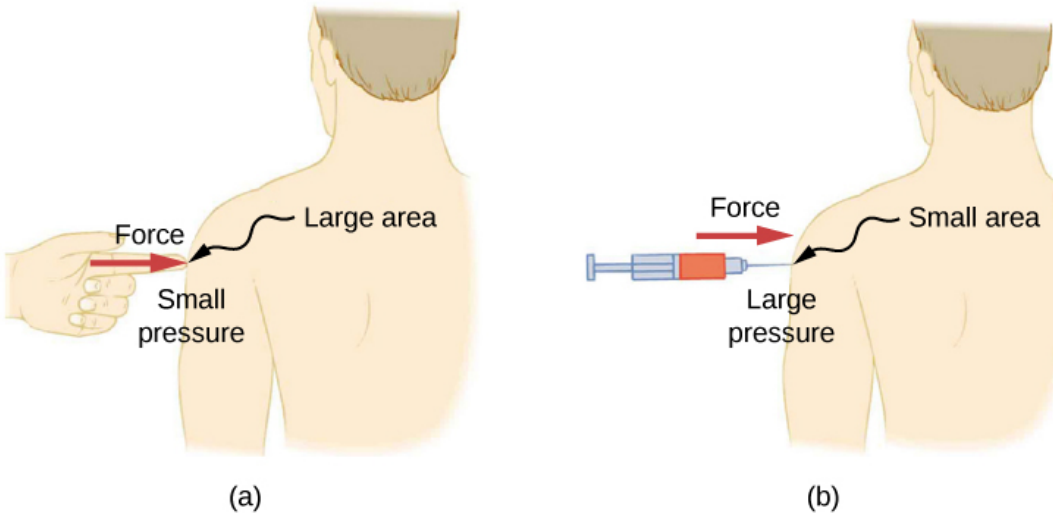
Pressure (p) is defined as the normal force F per unit area A over which the force is applied, or

Equation:

$$p = \frac{F}{A}.$$

To define the pressure at a specific point, the pressure is defined as the force dF exerted by a fluid over an infinitesimal element of area dA containing the point, resulting in $p = \frac{dF}{dA}$.

A given force can have a significantly different effect, depending on the area over which the force is exerted. For instance, a force applied to an area of 1 mm^2 has a pressure that is 100 times as great as the same force applied to an area of 1 cm^2 . That is why a sharp needle is able to poke through skin when a small force is exerted, but applying the same force with a finger does not puncture the skin ([link](#)).



(a) A person being poked with a finger might be irritated, but the force has little lasting effect. (b) In contrast, the same force applied to an area the size of the sharp end of a needle is enough to break the skin.

Note that although force is a vector, pressure is a scalar. Pressure is a scalar quantity because it is defined to be proportional to the magnitude of the force acting perpendicular to the surface area. The SI unit for pressure is the *pascal* (Pa), named after the French mathematician and physicist Blaise Pascal (1623–1662), where

Equation:

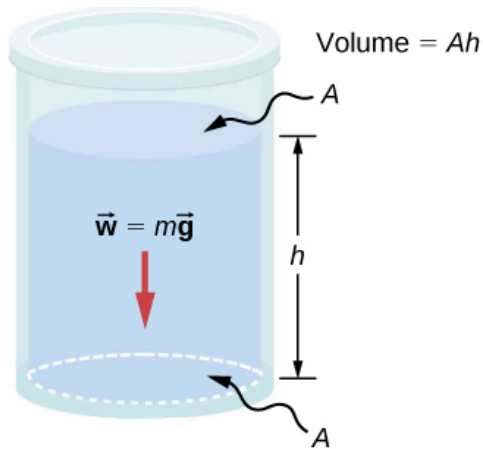
$$1 \text{ Pa} = 1 \text{ N/m}^2.$$

Several other units are used for pressure, which we discuss later in the chapter.

Variation of pressure with depth in a fluid of constant density

Pressure is defined for all states of matter, but it is particularly important when discussing fluids. An important characteristic of fluids is that there is no significant resistance to the component of a force applied parallel to the surface of a fluid. The molecules of the fluid simply flow to accommodate the horizontal force. A force applied perpendicular to the surface compresses or expands the fluid. If you try to compress a fluid, you find that a reaction force develops at each point inside the fluid in the outward direction, balancing the force applied on the molecules at the boundary.

Consider a fluid of constant density as shown in [\[link\]](#). The pressure at the bottom of the container is due to the pressure of the atmosphere (p_0) plus the pressure due to the weight of the fluid. The pressure due to the fluid is equal to the weight of the fluid divided by the area. The weight of the fluid is equal to its mass times the acceleration due to gravity.



The bottom of this container supports the entire weight of the fluid in it. The vertical sides cannot exert an upward force on the fluid (since it cannot withstand a shearing force), so the bottom must support it all.

Since the density is constant, the weight can be calculated using the density:

Equation:

$$w = mg = \rho Vg = \rho Ahg.$$

The pressure at the bottom of the container is therefore equal to atmospheric pressure added to the weight of the fluid divided by the area:

Equation:

$$p = p_0 + \frac{\rho Ahg}{A} = p_0 + \rho hg.$$

This equation is only good for pressure at a depth for a fluid of constant density.

Note:

Pressure at a Depth for a Fluid of Constant Density

The pressure at a depth in a fluid of constant density is equal to the pressure of the atmosphere plus the pressure due to the weight of the fluid, or

Equation:

$$p = p_0 + \rho hg,$$

Where p is the pressure at a particular depth, p_0 is the pressure of the atmosphere, ρ is the density of the fluid, g is the acceleration due to gravity, and h is the depth.

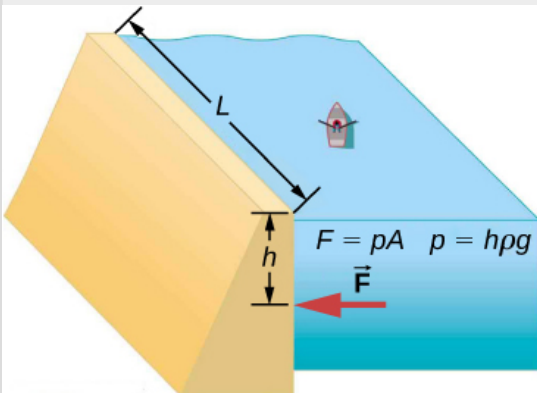


The Three Gorges Dam, erected on the Yangtze River in central China in 2008, created a massive reservoir that displaced more than one million people. (credit: modification of work by “Le Grand Portage”/Flickr)

Example:

What Force Must a Dam Withstand?

Consider the pressure and force acting on the dam retaining a reservoir of water ([link](#)). Suppose the dam is 500-m wide and the water is 80.0-m deep at the dam, as illustrated below. (a) What is the average pressure on the dam due to the water? (b) Calculate the force exerted against the dam.



The average pressure p due to the weight of the water is the pressure at the average depth h of 40.0 m, since pressure increases linearly with depth. The force exerted on the dam by the water is the average pressure times the area of contact, $F = pA$.

solution

- a. The average pressure due to the weight of a fluid is

Equation:

$$p = h\rho g.$$

Entering the density of water from [\[link\]](#) and taking h to be the average depth of 40.0 m, we obtain
Equation:

$$\begin{aligned} p &= (40.0 \text{ m}) \left(10^3 \frac{\text{kg}}{\text{m}^3} \right) \left(9.80 \frac{\text{m}}{\text{s}^2} \right) \\ &= 3.92 \times 10^5 \frac{\text{N}}{\text{m}^2} = 392 \text{ kPa}. \end{aligned}$$

b. We have already found the value for p . The area of the dam is

Equation:

$$A = 80.0 \text{ m} \times 500 \text{ m} = 4.00 \times 10^4 \text{ m}^2,$$

so that

Equation:

$$\begin{aligned} F &= (3.92 \times 10^5 \text{ N/m}^2)(4.00 \times 10^4 \text{ m}^2) \\ &= 1.57 \times 10^{10} \text{ N}. \end{aligned}$$

Significance

Although this force seems large, it is small compared with the $1.96 \times 10^{13} \text{ N}$ weight of the water in the reservoir. In fact, it is only 0.0800% of the weight.

Note:

Exercise:

Problem:

Check Your Understanding If the reservoir in [\[link\]](#) covered twice the area, but was kept to the same depth, would the dam need to be redesigned?

Solution:

The pressure found in part (a) of the example is completely independent of the width and length of the lake; it depends only on its average depth at the dam. Thus, the force depends only on the water's average depth and the dimensions of the dam, not on the horizontal extent of the reservoir. In the diagram, note that the thickness of the dam increases with depth to balance the increasing force due to the increasing pressure.

Pressure in a static fluid in a uniform gravitational field

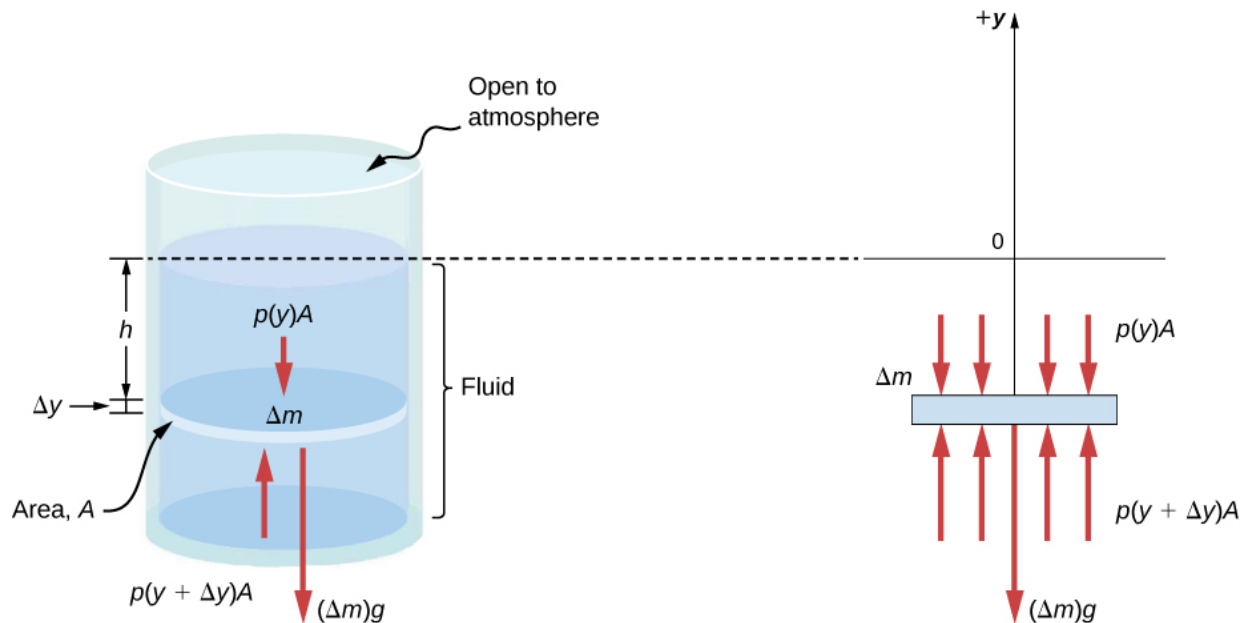
A *static fluid* is a fluid that is not in motion. At any point within a static fluid, the pressure on all sides must be equal—otherwise, the fluid at that point would react to a net force and accelerate.

The pressure at any point in a static fluid depends only on the depth at that point. As discussed, pressure in a fluid near Earth varies with depth due to the weight of fluid above a particular level. In the above examples, we assumed density to be constant and the average density of the fluid to be a good representation of the density. This is a reasonable approximation for liquids like water, where large

forces are required to compress the liquid or change the volume. In a swimming pool, for example, the density is approximately constant, and the water at the bottom is compressed very little by the weight of the water on top. Traveling up in the atmosphere is quite a different situation, however. The density of the air begins to change significantly just a short distance above Earth's surface.

To derive a formula for the variation of pressure with depth in a tank containing a fluid of density ρ on the surface of Earth, we must start with the assumption that the density of the fluid is not constant. Fluid located at deeper levels is subjected to more force than fluid nearer to the surface due to the weight of the fluid above it. Therefore, the pressure calculated at a given depth is different than the pressure calculated using a constant density.

Imagine a thin element of fluid at a depth h , as shown in [\[link\]](#). Let the element have a cross-sectional area A and height Δy . The forces acting upon the element are due to the pressures $p(y)$ above and $p(y + \Delta y)$ below it. The weight of the element itself is also shown in the free-body diagram.



Forces on a mass element inside a fluid. The weight of the element itself is shown in the free-body diagram.

Since the element of fluid between y and $y + \Delta y$ is not accelerating, the forces are balanced. Using a Cartesian y -axis oriented up, we find the following equation for the y -component:

Equation:

$$p(y + \Delta y)A - p(y)A - g\Delta m = 0 (\Delta y < 0).$$

Note that if the element had a non-zero y -component of acceleration, the right-hand side would not be zero but would instead be the mass times the y -acceleration. The mass of the element can be written in terms of the density of the fluid and the volume of the elements:

Equation:

$$\Delta m = |\rho A \Delta y| = -\rho A \Delta y \quad (\Delta y < 0).$$

Putting this expression for Δm into [\[link\]](#) and then dividing both sides by $A \Delta y$, we find

Equation:

$$\frac{p(y + \Delta y) - p(y)}{\Delta y} = -\rho g.$$

Taking the limit of the infinitesimally thin element $\Delta y \rightarrow 0$, we obtain the following differential equation, which gives the variation of pressure in a fluid:

Note:

Equation:

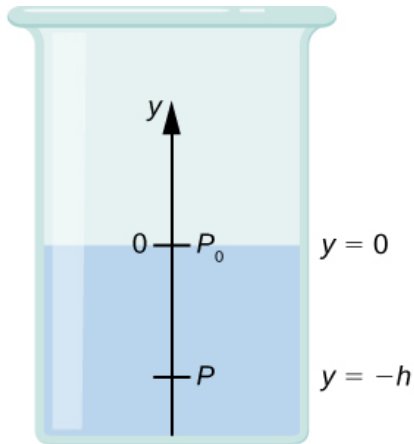
$$\frac{dp}{dy} = -\rho g.$$

This equation tells us that the rate of change of pressure in a fluid is proportional to the density of the fluid. The solution of this equation depends upon whether the density ρ is constant or changes with depth; that is, the function $\rho(y)$.

If the range of the depth being analyzed is not too great, we can assume the density to be constant. But if the range of depth is large enough for the density to vary appreciably, such as in the case of the atmosphere, there is significant change in density with depth. In that case, we cannot use the approximation of a constant density.

Pressure in a fluid with a constant density

Let's use [\[link\]](#) to work out a formula for the pressure at a depth h from the surface in a tank of a liquid such as water, where the density of the liquid can be taken to be constant.



We need to integrate [\[link\]](#) from $y = 0$, where the pressure is atmospheric pressure (p_0), to $y = -h$, the y-coordinate of the depth:

Equation:

$$\begin{aligned}\int_{p_0}^p dp &= - \int_0^{-h} \rho g dy \\ p - p_0 &= \rho gh \\ p &= p_0 + \rho gh.\end{aligned}$$

Hence, pressure at a depth of fluid on the surface of Earth is equal to the atmospheric pressure plus ρgh if the density of the fluid is constant over the height, as we found previously.

Note that the pressure in a fluid depends only on the depth from the surface and not on the shape of the container. Thus, in a container where a fluid can freely move in various parts, the liquid stays at the same level in every part, regardless of the shape, as shown in [\[link\]](#).



If a fluid can flow freely between parts of a container, it rises to the same height in each part. In the container pictured, the pressure at the bottom of each column is the same; if it were not the same, the

fluid would flow until the pressures became equal.

Variation of atmospheric pressure with height

The change in atmospheric pressure with height is of particular interest. Assuming the temperature of air to be constant, and that the ideal gas law of thermodynamics describes the atmosphere to a good approximation, we can find the variation of atmospheric pressure with height, when the temperature is constant. (We discuss the ideal gas law in a later chapter, but we assume you have some familiarity with it from high school and chemistry.) Let $p(y)$ be the atmospheric pressure at height y . The density ρ at y , the temperature T in the Kelvin scale (K), and the mass m of a molecule of air are related to the absolute pressure by the ideal gas law, in the form

Equation:

$$p = \rho \frac{k_B T}{m} \text{ (atmosphere),}$$

where k_B is Boltzmann's constant, which has a value of $1.38 \times 10^{-23} \text{ J/K}$.

You may have encountered the ideal gas law in the form $pV = nRT$, where n is the number of moles and R is the gas constant. Here, the same law has been written in a different form, using the density ρ instead of volume V . Therefore, if pressure p changes with height, so does the density ρ . Using density from the ideal gas law, the rate of variation of pressure with height is given as

Equation:

$$\frac{dp}{dy} = -p \left(\frac{mg}{k_B T} \right),$$

where constant quantities have been collected inside the parentheses. Replacing these constants with a single symbol α , the equation looks much simpler:

Equation:

$$\begin{aligned} \frac{dp}{dy} &= -\alpha p \\ \frac{dp}{p} &= -\alpha dy \\ \int_{p_0}^{p(y)} \frac{dp}{p} &= \int_0^y -\alpha dy \\ [\ln(p)]_{p_0}^{p(y)} &= [-\alpha y]_0^y \\ \ln(p) - \ln(p_0) &= -\alpha y \\ \ln\left(\frac{p}{p_0}\right) &= -\alpha y \end{aligned}$$

This gives the solution

Equation:

$$p(y) = p_0 \exp(-\alpha y).$$

Thus, atmospheric pressure drops exponentially with height, since the y -axis is pointed up from the ground and y has positive values in the atmosphere above sea level. The pressure drops by a factor of $\frac{1}{e}$ when the height is $\frac{1}{\alpha}$, which gives us a physical interpretation for α : The constant $\frac{1}{\alpha}$ is a length scale that characterizes how pressure varies with height and is often referred to as the pressure scale height.

We can obtain an approximate value of α by using the mass of a nitrogen molecule as a proxy for an air molecule. At temperature 27°C , or 300 K , we find

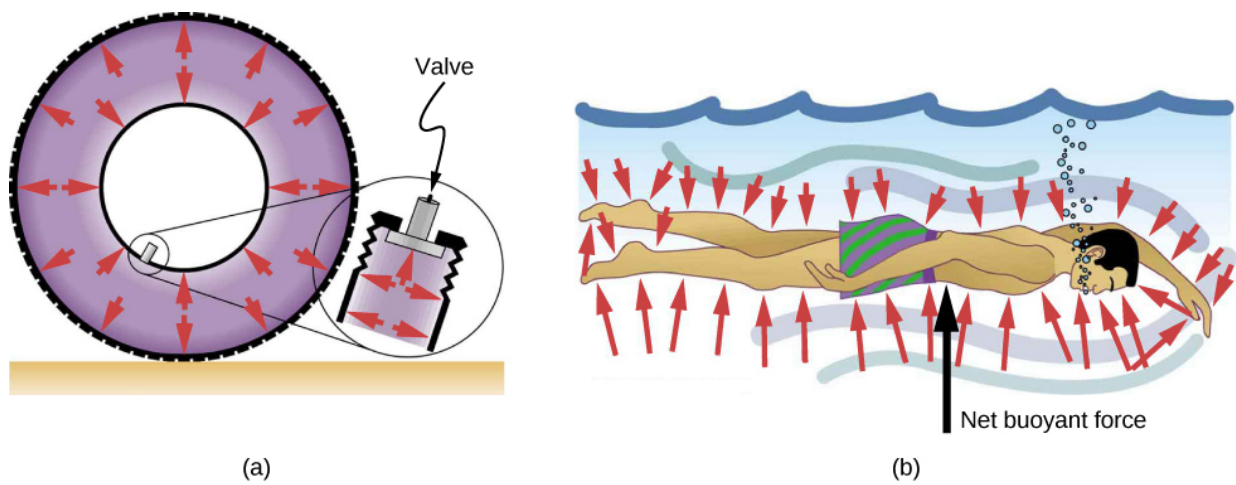
Equation:

$$\alpha = -\frac{mg}{k_B T} = \frac{4.8 \times 10^{-26} \text{ kg} \times 9.81 \text{ m/s}^2}{1.38 \times 10^{-23} \text{ J/K} \times 300 \text{ K}} = \frac{1}{8800 \text{ m}}.$$

Therefore, for every 8800 meters, the air pressure drops by a factor $1/e$, or approximately one-third of its value. This gives us only a rough estimate of the actual situation, since we have assumed both a constant temperature and a constant g over such great distances from Earth, neither of which is correct in reality.

Direction of pressure in a fluid

Fluid pressure has no direction, being a scalar quantity, whereas the forces due to pressure have well-defined directions: They are always exerted perpendicular to any surface. The reason is that fluids cannot withstand or exert shearing forces. Thus, in a static fluid enclosed in a tank, the force exerted on the walls of the tank is exerted perpendicular to the inside surface. Likewise, pressure is exerted perpendicular to the surfaces of any object within the fluid. [\[link\]](#) illustrates the pressure exerted by air on the walls of a tire and by water on the body of a swimmer.



(a) Pressure inside this tire exerts forces perpendicular to all surfaces it contacts. The arrows

represent directions and magnitudes of the forces exerted at various points. (b) Pressure is exerted perpendicular to all sides of this swimmer, since the water would flow into the space he occupies if he were not there. The arrows represent the directions and magnitudes of the forces exerted at various points on the swimmer. Note that the forces are larger underneath, due to greater depth, giving a net upward or buoyant force. The net vertical force on the swimmer is equal to the sum of the buoyant force and the weight of the swimmer.

Summary

- A fluid is a state of matter that yields to sideways or shearing forces. Liquids and gases are both fluids. Fluid statics is the physics of stationary fluids.
- Density is the mass per unit volume of a substance or object, defined as $\rho = m/V$. The SI unit of density is kg/m^3 .
- Pressure is the force per unit perpendicular area over which the force is applied, $p = F/A$. The SI unit of pressure is the pascal: $1 \text{ Pa} = 1 \text{ N/m}^2$.
- Pressure due to the weight of a liquid of constant density is given by $p = \rho gh$, where p is the pressure, h is the depth of the liquid, ρ is the density of the liquid, and g is the acceleration due to gravity.

Conceptual Questions

Exercise:

Problem:

Which of the following substances are fluids at room temperature and atmospheric pressure: air, mercury, water, glass?

Solution:

Mercury and water are liquid at room temperature and atmospheric pressure. Air is a gas at room temperature and atmospheric pressure. Glass is an amorphous solid (non-crystalline) material at room temperature and atmospheric pressure. At one time, it was thought that glass flowed, but flowed very slowly. This theory came from the observation that old glass planes were thicker at the bottom. It is now thought unlikely that this theory is accurate.

Exercise:

Problem: Why are gases easier to compress than liquids and solids?

Exercise:

Problem: Explain how the density of air varies with altitude.

Solution:

The density of air decreases with altitude. For a column of air of a constant temperature, the density decreases exponentially with altitude. This is a fair approximation, but since the temperature does change with altitude, it is only an approximation.

Exercise:**Problem:**

The image shows a glass of ice water filled to the brim. Will the water overflow when the ice melts? Explain your answer.

**Exercise:**

Problem: How is pressure related to the sharpness of a knife and its ability to cut?

Solution:

Pressure is force divided by area. If a knife is sharp, the force applied to the cutting surface is divided over a smaller area than the same force applied with a dull knife. This means that the pressure would be greater for the sharper knife, increasing its ability to cut.

Exercise:

Problem: Why is a force exerted by a static fluid on a surface always perpendicular to the surface?

Exercise:**Problem:**

Imagine that in a remote location near the North Pole, a chunk of ice floats in a lake. Next to the lake, a glacier with the same volume as the floating ice sits on land. If both chunks of ice should melt due to rising global temperatures, and the melted ice all goes into the lake, which one would cause the level of the lake to rise the most? Explain.

Solution:

If the two chunks of ice had the same volume, they would produce the same volume of water. The glacier would cause the greatest rise in the lake, however, because part of the floating chunk of ice is already submerged in the lake, and is thus already contributing to the lake's level.

Exercise:

Problem:

In ballet, dancing *en pointe* (on the tips of the toes) is much harder on the toes than normal dancing or walking. Explain why, in terms of pressure.

Exercise:**Problem:**

Atmospheric pressure exerts a large force (equal to the weight of the atmosphere above your body—about 10 tons) on the top of your body when you are lying on the beach sunbathing. Why are you able to get up?

Solution:

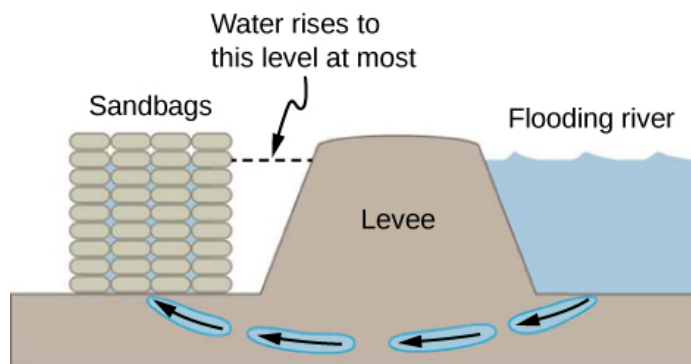
The pressure is acting all around your body, assuming you are not in a vacuum.

Exercise:

Problem: Why does atmospheric pressure decrease more rapidly than linearly with altitude?

Exercise:**Problem:**

The image shows how sandbags placed around a leak outside a river levee can effectively stop the flow of water under the levee. Explain how the small amount of water inside the column of sandbags is able to balance the much larger body of water behind the levee.

**Solution:**

Because the river level is very high, it has started to leak under the levee. Sandbags are placed around the leak, and the water held by them rises until it is the same level as the river, at which point the water there stops rising. The sandbags will absorb water until the water reaches the height of the water in the levee.

Exercise:

Problem: Is there a net force on a dam due to atmospheric pressure? Explain your answer.

Exercise:

Problem:

Does atmospheric pressure add to the gas pressure in a rigid tank? In a toy balloon? When, in general, does atmospheric pressure not affect the total pressure in a fluid?

Solution:

Atmospheric pressure does not affect the gas pressure in a rigid tank, but it does affect the pressure inside a balloon. In general, atmospheric pressure affects fluid pressure unless the fluid is enclosed in a rigid container.

Exercise:**Problem:**

You can break a strong wine bottle by pounding a cork into it with your fist, but the cork must press directly against the liquid filling the bottle—there can be no air between the cork and liquid. Explain why the bottle breaks only if there is no air between the cork and liquid.

Problems**Exercise:****Problem:**

Gold is sold by the troy ounce (31.103 g). What is the volume of 1 troy ounce of pure gold?

Solution:

1.610 cm^3

Exercise:**Problem:**

Mercury is commonly supplied in flasks containing 34.5 kg (about 76 lb.). What is the volume in liters of this much mercury?

Exercise:**Problem:**

What is the mass of a deep breath of air having a volume of 2.00 L? Discuss the effect taking such a breath has on your body's volume and density.

Solution:

The mass is 2.58 g. The volume of your body increases by the volume of air you inhale. The average density of your body decreases when you take a deep breath because the density of air is substantially smaller than the average density of the body.

Exercise:

Problem:

A straightforward method of finding the density of an object is to measure its mass and then measure its volume by submerging it in a graduated cylinder. What is the density of a 240-g rock that displaces 89.0 cm³ of water? (Note that the accuracy and practical applications of this technique are more limited than a variety of others that are based on Archimedes' principle.)

Exercise:**Problem:**

Suppose you have a coffee mug with a circular cross-section and vertical sides (uniform radius). What is its inside radius if it holds 375 g of coffee when filled to a depth of 7.50 cm? Assume coffee has the same density as water.

Solution:

3.99 cm

Exercise:**Problem:**

A rectangular gasoline tank can hold 50.0 kg of gasoline when full. What is the depth of the tank if it is 0.500-m wide by 0.900-m long? (b) Discuss whether this gas tank has a reasonable volume for a passenger car.

Exercise:**Problem:**

A trash compactor can compress its contents to 0.350 times their original volume. Neglecting the mass of air expelled, by what factor is the density of the rubbish increased?

Solution:

2.86 times denser

Exercise:**Problem:**

A 2.50-kg steel gasoline can holds 20.0 L of gasoline when full. What is the average density of the full gas can, taking into account the volume occupied by steel as well as by gasoline?

Exercise:**Problem:**

What is the density of 18.0-karat gold that is a mixture of 18 parts gold, 5 parts silver, and 1 part copper? (These values are parts by mass, not volume.) Assume that this is a simple mixture having an average density equal to the weighted densities of its constituents.

Solution:

15.6 g/cm³

Exercise:**Problem:**

The tip of a nail exerts tremendous pressure when hit by a hammer because it exerts a large force over a small area. What force must be exerted on a nail with a circular tip of 1.00-mm diameter to create a pressure of $3.00 \times 10^9 \text{ N/m}^2$? (This high pressure is possible because the hammer striking the nail is brought to rest in such a short distance.)

Exercise:**Problem:**

A glass tube contains mercury. What would be the height of the column of mercury which would create pressure equal to 1.00 atm?

Solution:

$$0.760 \text{ m} = 76.0 \text{ cm} = 760 \text{ mm}$$

Exercise:**Problem:**

The greatest ocean depths on Earth are found in the Marianas Trench near the Philippines. Calculate the pressure due to the ocean at the bottom of this trench, given its depth is 11.0 km and assuming the density of seawater is constant all the way down.

Exercise:

Problem: Verify that the SI unit of $h\rho g$ is N/m^2 .

Solution:

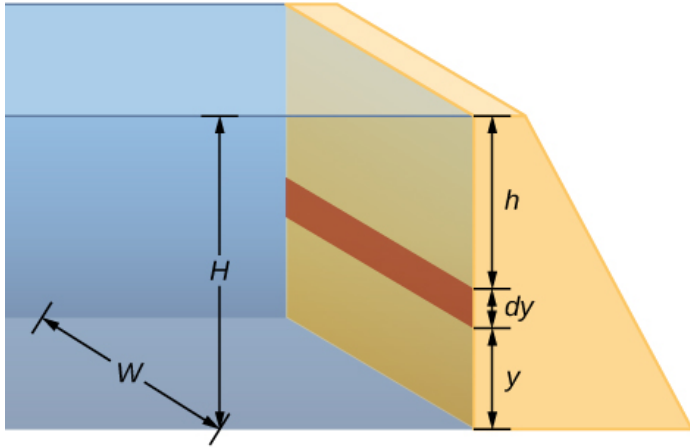
proof

Exercise:**Problem:**

What pressure is exerted on the bottom of a gas tank that is 0.500-m wide and 0.900-m long and can hold 50.0 kg of gasoline when full?

Exercise:**Problem:**

A dam is used to hold back a river. The dam has a height $H = 12 \text{ m}$ and a width $W = 10 \text{ m}$. Assume that the density of the water is $\rho = 1000 \text{ kg/m}^3$. (a) Determine the net force on the dam. (b) Why does the thickness of the dam increase with depth?



Solution:

- Pressure at $h = 7.06 \times 10^6 \text{ N}$;
- The pressure increases as the depth increases, so the dam must be built thicker toward the bottom to withstand the greater pressure.

Glossary**density**

mass per unit volume of a substance or object

fluids

liquids and gases; a fluid is a state of matter that yields to shearing forces

pressure

force per unit area exerted perpendicular to the area over which the force acts

specific gravity

ratio of the density of an object to a fluid (usually water)

Measuring Pressure

By the end of this section, you will be able to:

- Define gauge pressure and absolute pressure
- Explain various methods for measuring pressure
- Understand the working of open-tube barometers
- Describe in detail how manometers and barometers operate

In the preceding section, we derived a formula for calculating the variation in pressure for a fluid in hydrostatic equilibrium. As it turns out, this is a very useful calculation. Measurements of pressure are important in daily life as well as in science and engineering applications. In this section, we discuss different ways that pressure can be reported and measured.

Gauge Pressure vs. Absolute Pressure

Suppose the pressure gauge on a full scuba tank reads 3000 psi, which is approximately 207 atmospheres. When the valve is opened, air begins to escape because the pressure inside the tank is greater than the atmospheric pressure outside the tank. Air continues to escape from the tank until the pressure inside the tank equals the pressure of the atmosphere outside the tank. At this point, the pressure gauge on the tank reads zero, even though the pressure inside the tank is actually 1 atmosphere—the same as the air pressure outside the tank.

Most pressure gauges, like the one on the scuba tank, are calibrated to read zero at atmospheric pressure. Pressure readings from such gauges are called **gauge pressure**, which is the pressure relative to the atmospheric pressure. When the pressure inside the tank is greater than atmospheric pressure, the gauge reports a positive value.

Some gauges are designed to measure negative pressure. For example, many physics experiments must take place in a vacuum chamber, a rigid chamber from which some of the air is pumped out. The pressure inside the vacuum chamber is less than atmospheric pressure, so the pressure gauge on the chamber reads a negative value.

Unlike gauge pressure, **absolute pressure** accounts for atmospheric pressure, which in effect adds to the pressure in any fluid not enclosed in a rigid container.

Note:

Absolute Pressure

The absolute pressure, or total pressure, is the sum of gauge pressure and atmospheric pressure:

Equation:

$$p_{\text{abs}} = p_{\text{g}} + p_{\text{atm}}$$

where p_{abs} is absolute pressure, p_{g} is gauge pressure, and p_{atm} is atmospheric pressure.

For example, if a tire gauge reads 34 psi, then the absolute pressure is 34 psi plus 14.7 psi (p_{atm} in psi), or 48.7 psi (equivalent to 336 kPa).

In most cases, the absolute pressure in fluids cannot be negative. Fluids push rather than pull, so the smallest absolute pressure in a fluid is zero (a negative absolute pressure is a pull). Thus, the smallest possible gauge pressure is $p_{\text{g}} = -p_{\text{atm}}$ (which makes p_{abs} zero). There is no theoretical limit to how large a gauge pressure can be.

Measuring Pressure

A host of devices are used for measuring pressure, ranging from tire gauges to blood pressure monitors. Many other types of pressure gauges are commonly used to test the pressure of fluids, such as mechanical pressure gauges. We will explore some of these in this section.

Any property that changes with pressure in a known way can be used to construct a pressure gauge. Some of the most common types include strain gauges, which use the change in the shape of a material with pressure;

capacitance pressure gauges, which use the change in electric capacitance due to shape change with pressure; piezoelectric pressure gauges, which generate a voltage difference across a piezoelectric material under a pressure difference between the two sides; and ion gauges, which measure pressure by ionizing molecules in highly evacuated chambers. Different pressure gauges are useful in different pressure ranges and under different physical situations. Some examples are shown in [\[link\]](#).



(a)



(b)



(c)

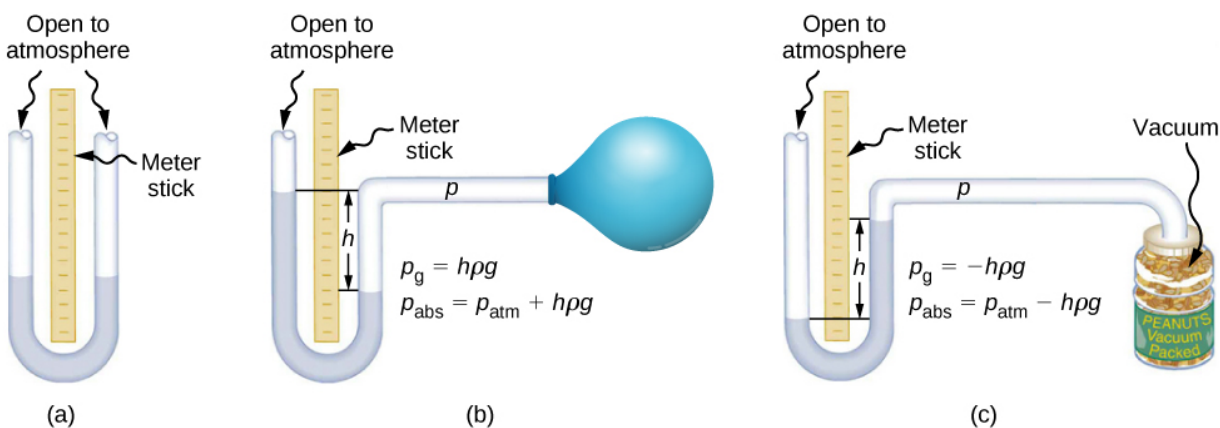
(a) Gauges are used to measure and monitor pressure in gas cylinders. Compressed gases are used in many industrial as well as medical applications. (b) Tire pressure gauges come in many different models, but all are meant for the same purpose: to measure the internal pressure of the tire. This enables the driver to keep the tires inflated at optimal pressure for load weight and driving conditions. (c) An ionization gauge is a high-sensitivity device used to monitor the pressure of gases in an enclosed system. Neutral gas molecules are ionized by the release of electrons, and the current is translated into a pressure reading. Ionization gauges are commonly used in industrial applications that rely on vacuum systems.

Manometers

One of the most important classes of pressure gauges applies the property that pressure due to the weight of a fluid of constant density is given by

$p = h\rho g$. The U-shaped tube shown in [\[link\]](#) is an example of a *manometer*; in part (a), both sides of the tube are open to the atmosphere, allowing atmospheric pressure to push down on each side equally so that its effects cancel.

A manometer with only one side open to the atmosphere is an ideal device for measuring gauge pressures. The gauge pressure is $p_g = h\rho g$ and is found by measuring h . For example, suppose one side of the U-tube is connected to some source of pressure p_{abs} , such as the balloon in part (b) of the figure or the vacuum-packed peanut jar shown in part (c). Pressure is transmitted undiminished to the manometer, and the fluid levels are no longer equal. In part (b), p_{abs} is greater than atmospheric pressure, whereas in part (c), p_{abs} is less than atmospheric pressure. In both cases, p_{abs} differs from atmospheric pressure by an amount $h\rho g$, where ρ is the density of the fluid in the manometer. In part (b), p_{abs} can support a column of fluid of height h , so it must exert a pressure $h\rho g$ greater than atmospheric pressure (the gauge pressure p_g is positive). In part (c), atmospheric pressure can support a column of fluid of height h , so p_{abs} is less than atmospheric pressure by an amount $h\rho g$ (the gauge pressure p_g is negative).



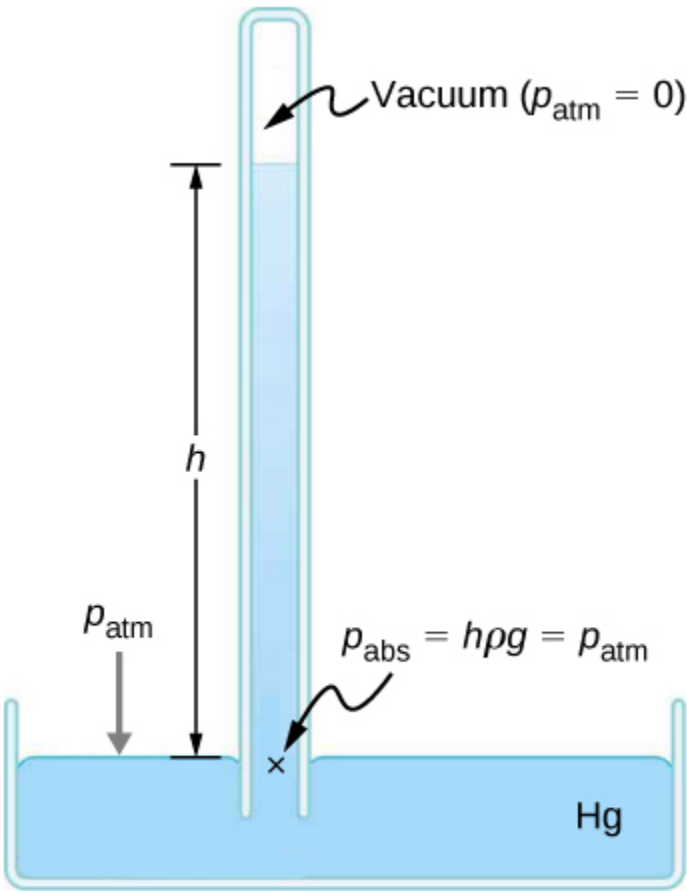
An open-tube manometer has one side open to the atmosphere. (a) Fluid depth must be the same on both sides, or the pressure each side exerts at the bottom will be unequal and liquid will flow from the deeper side. (b) A positive gauge pressure $p_g = h\rho g$ transmitted to one side of the manometer can support a column of fluid of height h . (c)

Similarly, atmospheric pressure is greater than a negative gauge pressure p_g by an amount $h\rho g$. The jar's rigidity prevents atmospheric pressure from being transmitted to the peanuts.

Barometers

Manometers typically use a U-shaped tube of a fluid (often mercury) to measure pressure. A *barometer* (see [\[link\]](#)) is a device that typically uses a single column of mercury to measure atmospheric pressure. The barometer, invented by the Italian mathematician and physicist Evangelista Torricelli (1608–1647) in 1643, is constructed from a glass tube closed at one end and filled with mercury. The tube is then inverted and placed in a pool of mercury. This device measures atmospheric pressure, rather than gauge pressure, because there is a nearly pure vacuum above the mercury in the tube. The height of the mercury is such that $h\rho g = p_{\text{atm}}$. When atmospheric pressure varies, the mercury rises or falls.

Weather forecasters closely monitor changes in atmospheric pressure (often reported as barometric pressure), as rising mercury typically signals improving weather and falling mercury indicates deteriorating weather. The barometer can also be used as an altimeter, since average atmospheric pressure varies with altitude. Mercury barometers and manometers are so common that units of mm Hg are often quoted for atmospheric pressure and blood pressures.



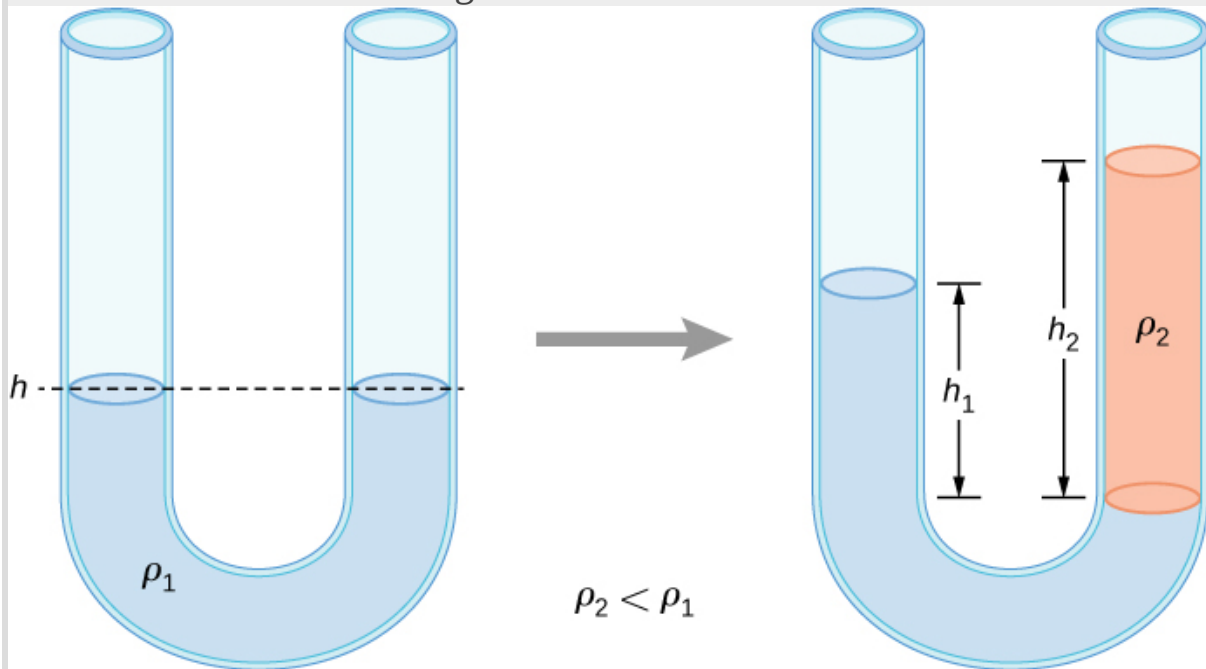
A mercury barometer measures atmospheric pressure. The pressure due to the mercury's weight, $h\rho g$, equals atmospheric pressure. The atmosphere is able to force mercury in the tube to a height h because the pressure above the mercury is zero.

Example:

Fluid Heights in an Open U-Tube

A U-tube with both ends open is filled with a liquid of density ρ_1 to a height h on both sides ([link](#)). A liquid of density $\rho_2 < \rho_1$ is poured into one side and Liquid 2 settles on top of Liquid 1. The heights on the two

sides are different. The height to the top of Liquid 2 from the interface is h_2 and the height to the top of Liquid 1 from the level of the interface is h_1 . Derive a formula for the height difference.



Two liquids of different densities are shown in a U-tube.

Strategy

The pressure at points at the same height on the two sides of a U-tube must be the same as long as the two points are in the same liquid. Therefore, we consider two points at the same level in the two arms of the tube: One point is the interface on the side of the Liquid 2 and the other is a point in the arm with Liquid 1 that is at the same level as the interface in the other arm. The pressure at each point is due to atmospheric pressure plus the weight of the liquid above it.

Equation:

Pressure on the side with Liquid 1 = $p_0 + \rho_1 g h_1$

Pressure on the side with Liquid 2 = $p_0 + \rho_2 g h_2$

Solution

Since the two points are in Liquid 1 and are at the same height, the pressure at the two points must be the same. Therefore, we have

Equation:

$$p_0 + \rho_1 g h_1 = p_0 + \rho_2 g h_2.$$

Hence,

Equation:

$$\rho_1 h_1 = \rho_2 h_2.$$

This means that the difference in heights on the two sides of the U-tube is

Equation:

$$h_2 - h_1 = \left(1 - \frac{\rho_2}{\rho_1}\right) h_2.$$

The result makes sense if we set $\rho_2 = \rho_1$, which gives $h_2 = h_1$. If the two sides have the same density, they have the same height.

Note:

Exercise:

Problem:

Check Your Understanding Mercury is a hazardous substance. Why do you suppose mercury is typically used in barometers instead of a safer fluid such as water?

Solution:

The density of mercury is 13.6 times greater than the density of water. It takes approximately 76 cm (29.9 in.) of mercury to measure the pressure of the atmosphere, whereas it would take approximately 10 m (34 ft.) of water.

Units of pressure

As stated earlier, the SI unit for pressure is the pascal (Pa), where

Equation:

$$1 \text{ Pa} = 1 \text{ N/m}^2.$$

In addition to the pascal, many other units for pressure are in common use ([link](#)). In meteorology, atmospheric pressure is often described in the unit of millibars (mbar), where

Equation:

$$1000 \text{ mbar} = 1 \times 10^5 \text{ Pa}.$$

The millibar is a convenient unit for meteorologists because the average atmospheric pressure at sea level on Earth is $1.013 \times 10^5 \text{ Pa} = 1013 \text{ mbar} = 1 \text{ atm}$. Using the equations derived when considering pressure at a depth in a fluid, pressure can also be measured as millimeters or inches of mercury. The pressure at the bottom of a 760-mm column of mercury at 0°C in a container where the top part is evacuated is equal to the atmospheric pressure. Thus, 760 mm Hg is also used in place of 1 atmosphere of pressure. In vacuum physics labs, scientists often use another unit called the torr, named after Torricelli, who, as we have just seen, invented the mercury manometer for measuring pressure. One torr is equal to a pressure of 1 mm Hg.

Unit	Definition
SI unit: the Pascal	$1 \text{ Pa} = 1 \text{ N/m}^2$

Unit	Definition
English unit: pounds per square inch (lb/in. ² or psi)	1 psi = 6.895×10^3 Pa
Other units of pressure	1 atm = 760 mmHg = 1.013×10^5 Pa = 14.7 psi = 29.9 inches of Hg = 1013 mbar
	1 bar = 10^5 Pa
	1 torr = 1 mm Hg = 133.3 Pa

Summary of the Units of Pressure

Summary

- Gauge pressure is the pressure relative to atmospheric pressure.
- Absolute pressure is the sum of gauge pressure and atmospheric pressure.
- Open-tube manometers have U-shaped tubes and one end is always open. They are used to measure pressure. A mercury barometer is a device that measures atmospheric pressure.
- The SI unit of pressure is the pascal (Pa), but several other units are commonly used.

Conceptual Questions

Exercise:

Problem:

Explain why the fluid reaches equal levels on either side of a manometer if both sides are open to the atmosphere, even if the tubes are of different diameters.

Solution:

The pressure of the atmosphere is due to the weight of the air above. The pressure, force per area, on the manometer will be the same at the same depth of the atmosphere.

Problems**Exercise:****Problem:**

Find the gauge and absolute pressures in the balloon and peanut jar shown in [\[link\]](#), assuming the manometer connected to the balloon uses water and the manometer connected to the jar contains mercury. Express in units of centimeters of water for the balloon and millimeters of mercury for the jar, taking $h = 0.0500\text{m}$ for each.

Exercise:**Problem:**

How tall must a water-filled manometer be to measure blood pressure as high as 300 mm Hg?

Solution:

4.08 m

Exercise:

Problem:

Assuming bicycle tires are perfectly flexible and support the weight of bicycle and rider by pressure alone, calculate the total area of the tires in contact with the ground if a bicycle and rider have a total mass of 80.0 kg, and the gauge pressure in the tires is $3.50 \times 10^5 \text{ Pa}$.

Glossary

absolute pressure

sum of gauge pressure and atmospheric pressure

gauge pressure

pressure relative to atmospheric pressure

Pascal's Principle and Hydraulics

By the end of this section, you will be able to:

- State Pascal's principle
- Describe applications of Pascal's principle
- Derive relationships between forces in a hydraulic system

In 1653, the French philosopher and scientist Blaise Pascal published his *Treatise on the Equilibrium of Liquids*, in which he discussed principles of static fluids. A static fluid is a fluid that is not in motion. When a fluid is not flowing, we say that the fluid is in static equilibrium. If the fluid is water, we say it is in **hydrostatic equilibrium**. For a fluid in static equilibrium, the net force on any part of the fluid must be zero; otherwise the fluid will start to flow.

Pascal's observations—since proven experimentally—provide the foundation for hydraulics, one of the most important developments in modern mechanical technology. Pascal observed that a change in pressure applied to an enclosed fluid is transmitted undiminished throughout the fluid and to the walls of its container. Because of this, we often know more about pressure than other physical quantities in fluids. Moreover, Pascal's principle implies that the total pressure in a fluid is the sum of the pressures from different sources. A good example is the fluid at a depth depends on the depth of the fluid and the pressure of the atmosphere.

Pascal's Principle

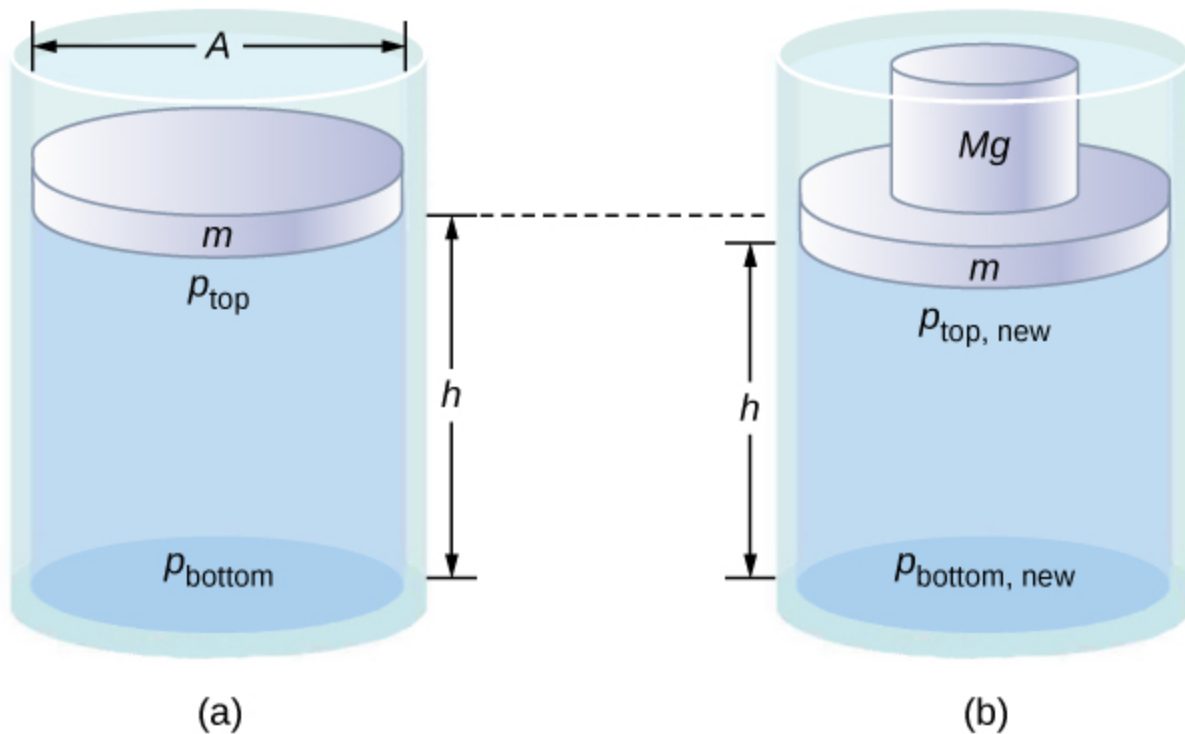
Pascal's principle (also known as Pascal's law) states that when a change in pressure is applied to an enclosed fluid, it is transmitted undiminished to all portions of the fluid and to the walls of its container. In an enclosed fluid, since atoms of the fluid are free to move about, they transmit pressure to all parts of the fluid *and* to the walls of the container. Any change in pressure is transmitted undiminished.

Note that this principle does not say that the pressure is the same at all points of a fluid—which is not true, since the pressure in a fluid near Earth

varies with height. Rather, this principle applies to the *change* in pressure. Suppose you place some water in a cylindrical container of height H and cross-sectional area A that has a movable piston of mass m ([link](#)). Adding weight Mg at the top of the piston increases the pressure at the top by Mg/A , since the additional weight also acts over area A of the lid:

Equation:

$$\Delta p_{\text{top}} = \frac{Mg}{A}.$$



Pressure in a fluid changes when the fluid is compressed. (a) The pressure at the top layer of the fluid is different from pressure at the bottom layer. (b) The increase in pressure by adding weight to the piston is the same everywhere, for example,

$$p_{\text{top new}} - p_{\text{top}} = p_{\text{bottom new}} - p_{\text{bottom}}.$$

According to Pascal's principle, the pressure at all points in the water changes by the same amount, Mg/A . Thus, the pressure at the bottom also increases by Mg/A . The pressure at the bottom of the container is equal to the sum of the atmospheric pressure, the pressure due the fluid, and the pressure supplied by the mass. The change in pressure at the bottom of the container due to the mass is

Equation:

$$\Delta p_{\text{bottom}} = \frac{Mg}{A}.$$

Since the pressure changes are the same everywhere in the fluid, we no longer need subscripts to designate the pressure change for top or bottom:

Equation:

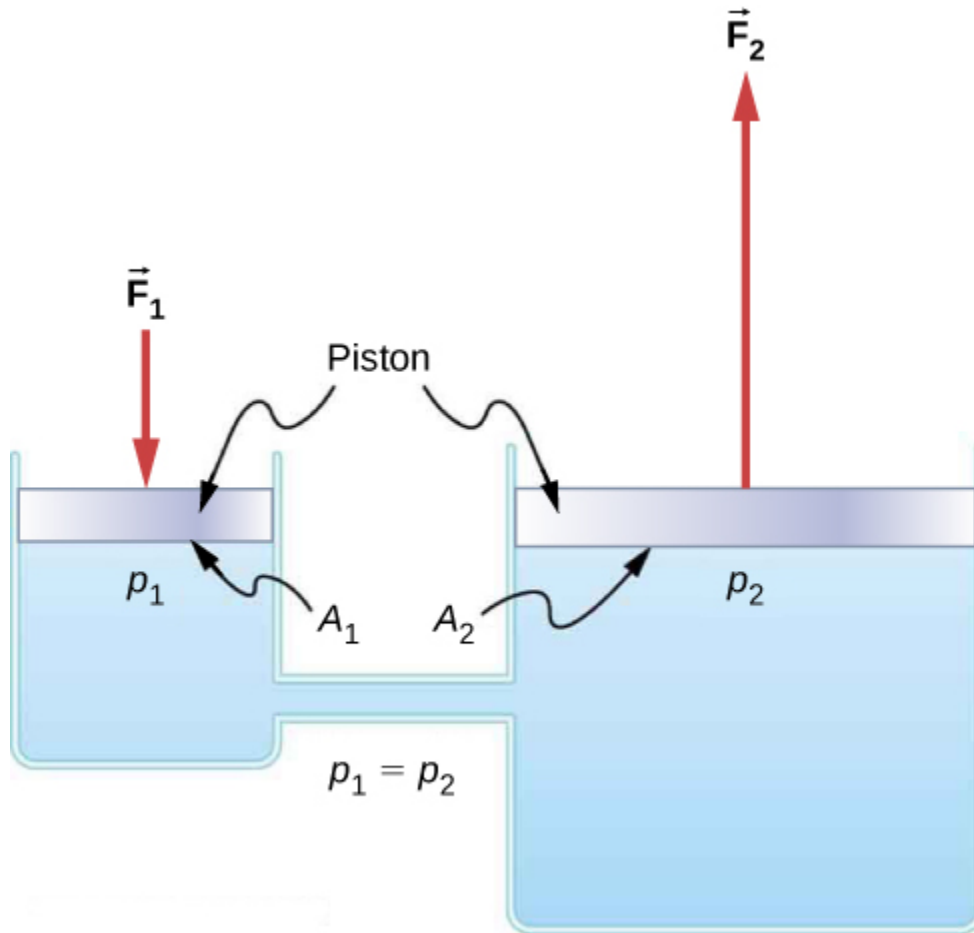
$$\Delta p = \Delta p_{\text{top}} = \Delta p_{\text{bottom}} = \Delta p_{\text{everywhere}}.$$

Note:

Pascal's Barrel is a great demonstration of Pascal's principle. Watch a [simulation](#) of Pascal's 1646 experiment, in which he demonstrated the effects of changing pressure in a fluid.

Applications of Pascal's Principle and Hydraulic Systems

Hydraulic systems are used to operate automotive brakes, hydraulic jacks, and numerous other mechanical systems ([\[link\]](#)).



A typical hydraulic system with two fluid-filled cylinders, capped with pistons and connected by a tube called a hydraulic line. A downward force \vec{F}_1 on the left piston creates a change in pressure that is transmitted undiminished to all parts of the enclosed fluid. This results in an upward force \vec{F}_2 on the right piston that is larger than \vec{F}_1 because the right piston has a larger surface area.

We can derive a relationship between the forces in this simple hydraulic system by applying Pascal's principle. Note first that the two pistons in the system are at the same height, so there is no difference in pressure due to a difference in depth. The pressure due to F_1 acting on area A_1 is simply

Equation:

$$p_1 = \frac{F_1}{A_1}, \text{ as defined by } p = \frac{F}{A}.$$

According to Pascal's principle, this pressure is transmitted undiminished throughout the fluid and to all walls of the container. Thus, a pressure p_2 is felt at the other piston that is equal to p_1 . That is, $p_1 = p_2$. However, since $p_2 = F_2/A_2$, we see that

Note:**Equation:**

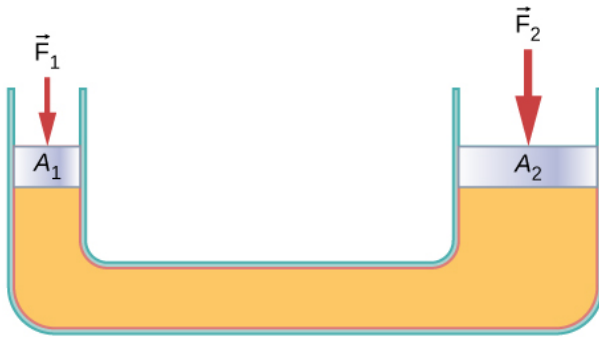
$$\frac{F_1}{A_1} = \frac{F_2}{A_2}.$$

This equation relates the ratios of force to area in any hydraulic system, provided that the pistons are at the same vertical height and that friction in the system is negligible.

Hydraulic systems can increase or decrease the force applied to them. To make the force larger, the pressure is applied to a larger area. For example, if a 100-N force is applied to the left cylinder in [\[link\]](#) and the right cylinder has an area five times greater, then the output force is 500 N. Hydraulic systems are analogous to simple levers, but they have the advantage that pressure can be sent through tortuously curved lines to several places at once.

The **hydraulic jack** is such a hydraulic system. A hydraulic jack is used to lift heavy loads, such as the ones used by auto mechanics to raise an automobile. It consists of an incompressible fluid in a U-tube fitted with a movable piston on each side. One side of the U-tube is narrower than the

other. A small force applied over a small area can balance a much larger force on the other side over a larger area ([link](#)).



(a)



(b)

(a) A hydraulic jack operates by applying forces (F_1 , F_2) to an incompressible fluid in a U-tube, using a movable piston (A_1 , A_2) on each side of the tube. (b) Hydraulic jacks are commonly used by car mechanics to lift vehicles so that repairs and maintenance can be performed. (credit b: modification of work by Jane Whitney)

From Pascal's principle, it can be shown that the force needed to lift the car is less than the weight of the car:

Equation:

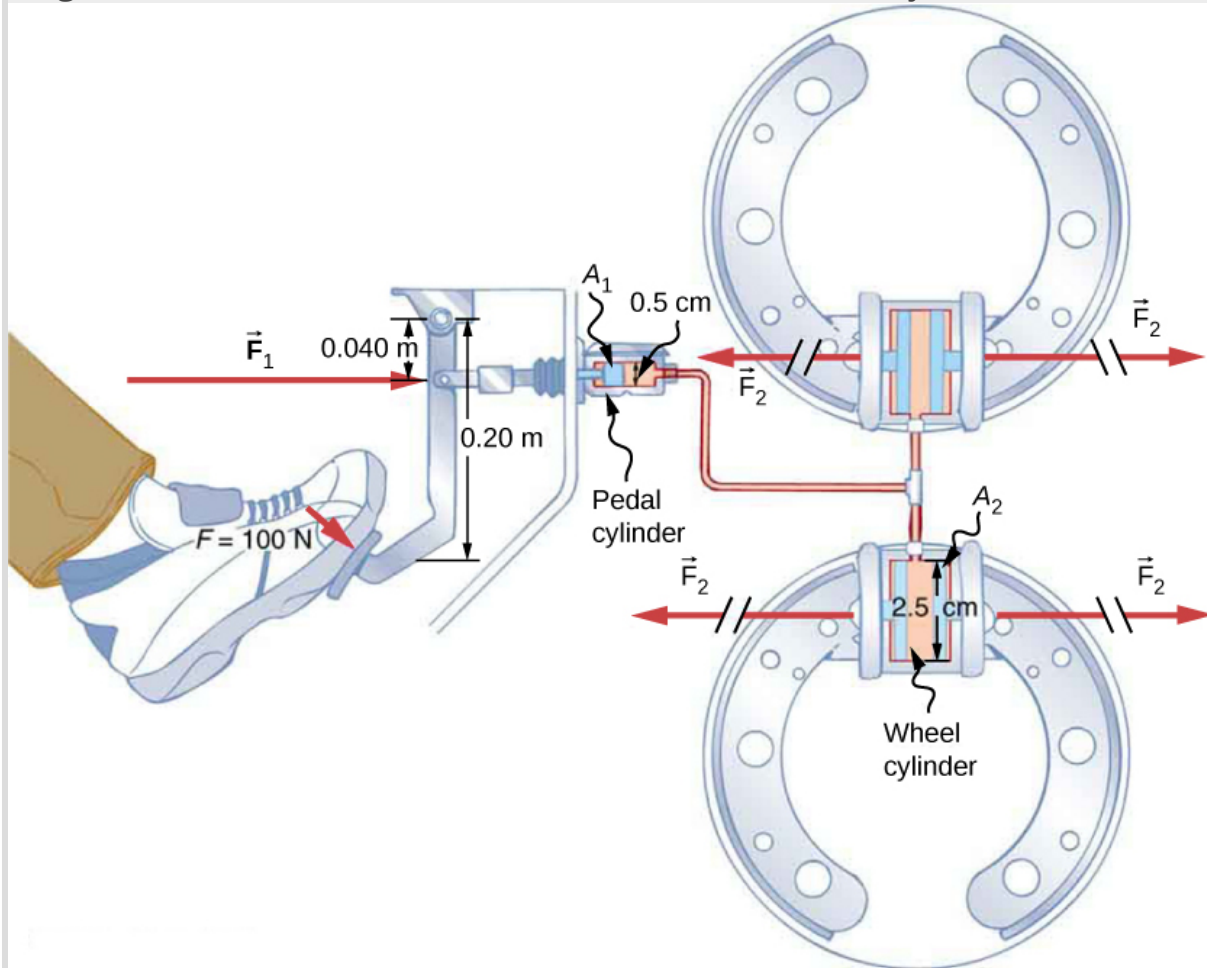
$$F_1 = \frac{A_1}{A_2} F_2,$$

where F_1 is the force applied to lift the car, A_1 is the cross-sectional area of the smaller piston, A_2 is the cross sectional area of the larger piston, and F_2 is the weight of the car.

Example:

Calculating Force on Wheel Cylinders: Pascal Puts on the Brakes

Consider the automobile hydraulic system shown in [\[link\]](#). Suppose a force of 100 N is applied to the brake pedal, which acts on the pedal cylinder (acting as a “master” cylinder) through a lever. A force of 500 N is exerted on the pedal cylinder. Pressure created in the pedal cylinder is transmitted to the four wheel cylinders. The pedal cylinder has a diameter of 0.500 cm and each wheel cylinder has a diameter of 2.50 cm. Calculate the magnitude of the force F_2 created at each of the wheel cylinders.



Hydraulic brakes use Pascal's principle. The driver pushes the brake pedal, exerting a force that is increased by the simple lever and again by the hydraulic system. Each of the identical wheel cylinders receives the same pressure and, therefore, creates the same force output F_2 . The circular cross-sectional areas of the pedal and wheel cylinders are represented by A_1 and A_2 , respectively.

Strategy

We are given the force F_1 applied to the pedal cylinder. The cross-sectional areas A_1 and A_2 can be calculated from their given diameters. Then we can use the following relationship to find the force F_2 :

Equation:

$$\frac{F_1}{A_1} = \frac{F_2}{A_2}.$$

Manipulate this algebraically to get F_2 on one side and substitute known values.

Solution

Pascal's principle applied to hydraulic systems is given by $\frac{F_1}{A_1} = \frac{F_2}{A_2}$:

Equation:

$$\begin{aligned} F_2 &= \frac{A_2}{A_1} F_1 = \frac{\pi r_2^2}{\pi r_1^2} F_1 \\ &= \frac{(1.25 \text{ cm})^2}{(0.250 \text{ cm})^2} \times 500 \text{ N} = 1.25 \times 10^4 \text{ N}. \end{aligned}$$

Significance

This value is the force exerted by each of the four wheel cylinders. Note that we can add as many wheel cylinders as we wish. If each has a 2.50-cm diameter, each will exert $1.25 \times 10^4 \text{ N}$. A simple hydraulic system, as an example of a simple machine, can increase force but cannot do more work than is done on it. Work is force times distance moved, and the wheel cylinder moves through a smaller distance than the pedal cylinder.

Furthermore, the more wheels added, the smaller the distance each one moves. Many hydraulic systems—such as power brakes and those in bulldozers—have a motorized pump that actually does most of the work in the system.

Note:**Exercise:**

Problem:

Check Your Understanding Would a hydraulic press still operate properly if a gas is used instead of a liquid?

Solution:

Yes, it would still work, but since a gas is compressible, it would not operate as efficiently. When the force is applied, the gas would first compress and warm. Hence, the air in the brake lines must be bled out in order for the brakes to work properly.

Summary

- Pressure is force per unit area.
- A change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container.
- A hydraulic system is an enclosed fluid system used to exert forces.

Conceptual Questions

Exercise:**Problem:**

Suppose the master cylinder in a hydraulic system is at a greater height than the cylinder it is controlling. Explain how this will affect the force produced at the cylinder that is being controlled.

Problems

Exercise:

Problem:

How much pressure is transmitted in the hydraulic system considered in [\[link\]](#)? Express your answer in atmospheres.

Solution:

251 atm

Exercise:**Problem:**

What force must be exerted on the master cylinder of a hydraulic lift to support the weight of a 2000-kg car (a large car) resting on a second cylinder? The master cylinder has a 2.00-cm diameter and the second cylinder has a 24.0-cm diameter.

Exercise:**Problem:**

A host pours the remnants of several bottles of wine into a jug after a party. The host then inserts a cork with a 2.00-cm diameter into the bottle, placing it in direct contact with the wine. The host is amazed when the host pounds the cork into place and the bottom of the jug (with a 14.0-cm diameter) breaks away. Calculate the extra force exerted against the bottom if he pounded the cork with a 120-N force.

Solution:

5.76×10^3 N extra force

Exercise:

Problem:

A certain hydraulic system is designed to exert a force 100 times as large as the one put into it. (a) What must be the ratio of the area of the cylinder that is being controlled to the area of the master cylinder? (b) What must be the ratio of their diameters? (c) By what factor is the distance through which the output force moves reduced relative to the distance through which the input force moves? Assume no losses due to friction.

Exercise:**Problem:**

Verify that work input equals work output for a hydraulic system assuming no losses due to friction. Do this by showing that the distance the output force moves is reduced by the same factor that the output force is increased. Assume the volume of the fluid is constant. What effect would friction within the fluid and between components in the system have on the output force? How would this depend on whether or not the fluid is moving?

Solution:

If the system is not moving, the friction would not play a role. With friction, we know there are losses, so that $W_o = W_i - W_f$; therefore, the work output is less than the work input. In other words, to account for friction, you would need to push harder on the input piston than was calculated.

Glossary

hydraulic jack

simple machine that uses cylinders of different diameters to distribute force

hydrostatic equilibrium

state at which water is not flowing, or is static

Pascal's principle

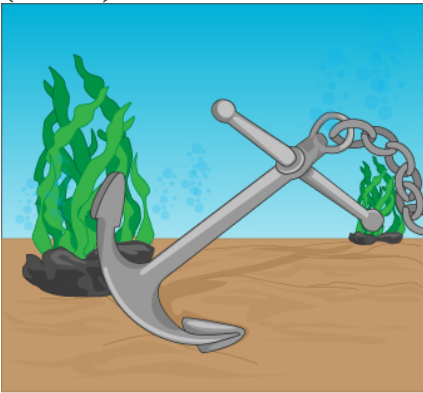
change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container

Archimedes' Principle and Buoyancy

By the end of this section, you will be able to:

- Define buoyant force
- State Archimedes' principle
- Describe the relationship between density and Archimedes' principle

When placed in a fluid, some objects float due to a buoyant force. Where does this buoyant force come from? Why is it that some things float and others do not? Do objects that sink get any support at all from the fluid? Is your body buoyed by the atmosphere, or are only helium balloons affected ([link](#))?



(a)



(b)



(c)

(a) Even objects that sink, like this anchor, are partly supported by water when submerged. (b) Submarines have adjustable density (ballast tanks) so that they may float or sink as desired. (c) Helium-filled balloons tug upward on their strings, demonstrating air's buoyant effect. (credit b: modification of work by Allied Navy; credit c: modification of work by "Crystl"/Flickr)

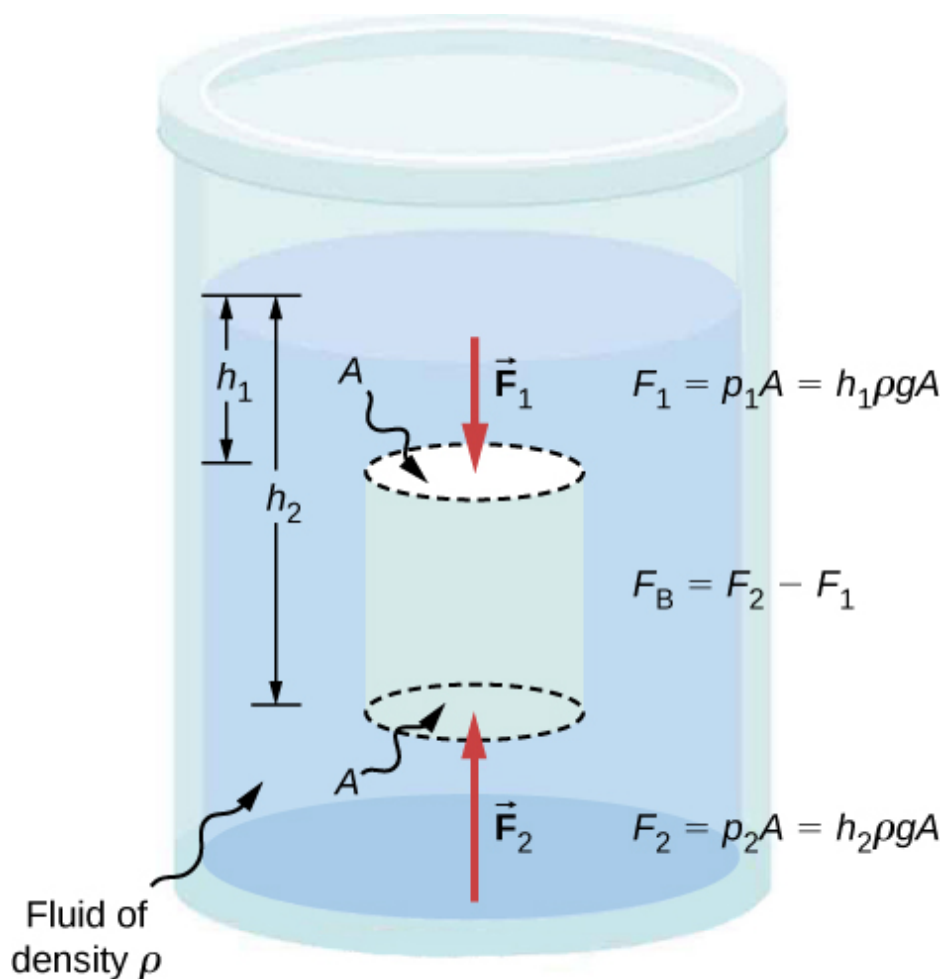
Answers to all these questions, and many others, are based on the fact that pressure increases with depth in a fluid. This means that the upward force on the bottom of an object in a fluid is greater than the downward force on top of the object. There is an upward force, or **buoyant force**, on any object in any fluid ([link](#)). If the buoyant force is greater than the object's weight, the

object rises to the surface and floats. If the buoyant force is less than the object's weight, the object sinks. If the buoyant force equals the object's weight, the object can remain suspended at its present depth. The buoyant force is always present, whether the object floats, sinks, or is suspended in a fluid.

Note:

Buoyant Force

The buoyant force is the upward force on any object in any fluid.



Pressure due to the weight of a fluid increases with depth because $p = h\rho g$. This change in pressure and

associated upward force on the bottom of the cylinder are greater than the downward force on the top of the cylinder. The differences in the force results in the buoyant force F_B . (Horizontal forces cancel.)

Archimedes' Principle

Just how large a force is buoyant force? To answer this question, think about what happens when a submerged object is removed from a fluid, as in [\[link\]](#). If the object were not in the fluid, the space the object occupied would be filled by fluid having a weight w_{fl} . This weight is supported by the surrounding fluid, so the buoyant force must equal w_{fl} , the weight of the fluid displaced by the object.

Note:

Archimedes' Principle

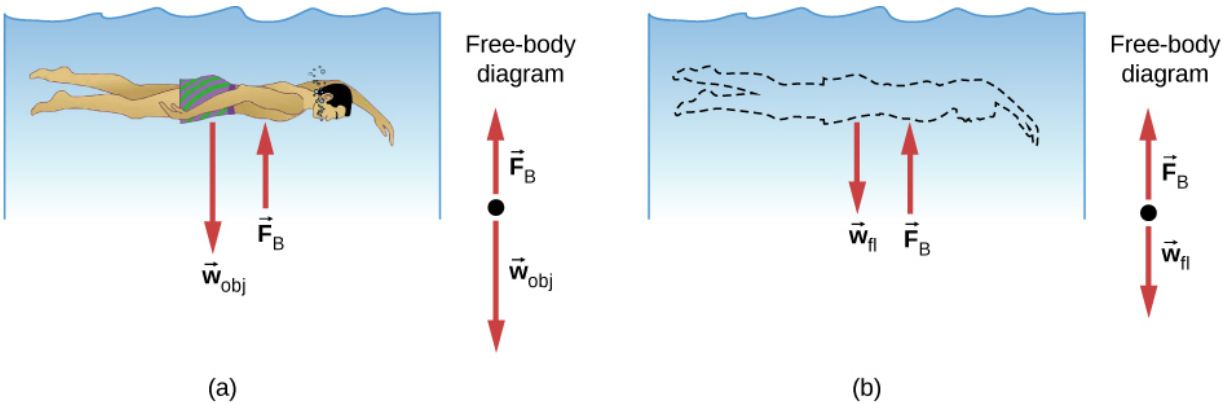
The buoyant force on an object equals the weight of the fluid it displaces. In equation form, **Archimedes' principle** is

Equation:

$$F_B = w_{fl},$$

where F_B is the buoyant force and w_{fl} is the weight of the fluid displaced by the object.

This principle is named after the Greek mathematician and inventor Archimedes (ca. 287–212 BCE), who stated this principle long before concepts of force were well established.



(a) An object submerged in a fluid experiences a buoyant force F_B . If F_B is greater than the weight of the object, the object rises. If F_B is less than the weight of the object, the object sinks. (b) If the object is removed, it is replaced by fluid having weight w_{fl} . Since this weight is supported by surrounding fluid, the buoyant force must equal the weight of the fluid displaced.

Archimedes' principle refers to the force of buoyancy that results when a body is submerged in a fluid, whether partially or wholly. The force that provides the pressure of a fluid acts on a body perpendicular to the surface of the body. In other words, the force due to the pressure at the bottom is pointed up, while at the top, the force due to the pressure is pointed down; the forces due to the pressures at the sides are pointing into the body.

Since the bottom of the body is at a greater depth than the top of the body, the pressure at the lower part of the body is higher than the pressure at the upper part, as shown in [\[link\]](#). Therefore a net upward force acts on the body. This upward force is the force of buoyancy, or simply *buoyancy*.

Note:

The exclamation “Eureka” (meaning “I found it”) has often been credited to Archimedes as he made the discovery that would lead to Archimedes’ principle. Some say it all started in a bathtub. To hear this story, watch this [video](#) or explore [Scientific American](#) to learn more.

Density and Archimedes' Principle

If you drop a lump of clay in water, it will sink. But if you mold the same lump of clay into the shape of a boat, it will float. Because of its shape, the clay boat displaces more water than the lump and experiences a greater buoyant force, even though its mass is the same. The same is true of steel ships.

The average density of an object is what ultimately determines whether it floats. If an object's average density is less than that of the surrounding fluid, it will float. The reason is that the fluid, having a higher density, contains more mass and hence more weight in the same volume. The buoyant force, which equals the weight of the fluid displaced, is thus greater than the weight of the object. Likewise, an object denser than the fluid will sink.

The extent to which a floating object is submerged depends on how the object's density compares to the density of the fluid. In [\[link\]](#), for example, the unloaded ship has a lower density and less of it is submerged compared with the same ship when loaded. We can derive a quantitative expression for the fraction submerged by considering density. The fraction submerged is the ratio of the volume submerged to the volume of the object, or

Equation:

$$\text{fraction submerged} = \frac{V_{\text{sub}}}{V_{\text{obj}}} = \frac{V_{\text{fl}}}{V_{\text{obj}}}.$$

The volume submerged equals the volume of fluid displaced, which we call V_{fl} . Now we can obtain the relationship between the densities by substituting $\rho = \frac{m}{V}$ into the expression. This gives

Equation:

$$\frac{V_{\text{fl}}}{V_{\text{obj}}} = \frac{m_{\text{fl}}/\rho_{\text{fl}}}{m_{\text{obj}}/\rho_{\text{obj}}},$$

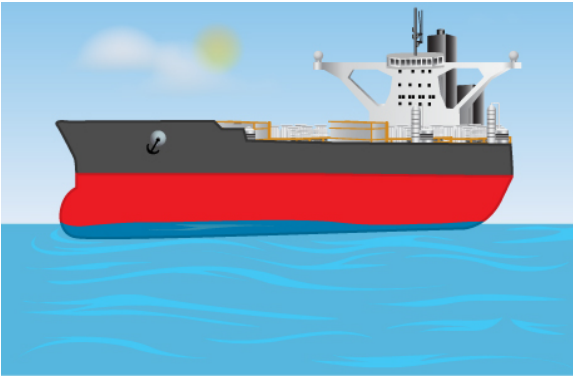
where ρ_{obj} is the average density of the object and ρ_{fl} is the density of the fluid. Since the object floats, its mass and that of the displaced fluid are

equal, so they cancel from the equation, leaving

Equation:

$$\text{fraction submerged} = \frac{\rho_{\text{obj}}}{\rho_{\text{fl}}}.$$

We can use this relationship to measure densities.



(a)



(b)

An unloaded ship (a) floats higher in the water than a loaded ship (b).

Example:

Calculating Average Density

Suppose a 60.0-kg woman floats in fresh water with 97.0% of her volume submerged when her lungs are full of air. What is her average density?

Strategy

We can find the woman's density by solving the equation

Equation:

$$\text{fraction submerged} = \frac{\rho_{\text{obj}}}{\rho_{\text{fl}}}$$

for the density of the object. This yields

Equation:

$$\rho_{\text{obj}} = \rho_{\text{person}} = (\text{fraction submerged}) \cdot \rho_{\text{fl}}.$$

We know both the fraction submerged and the density of water, so we can calculate the woman's density.

Solution

Entering the known values into the expression for her density, we obtain

Equation:

$$\rho_{\text{person}} = 0.970 \cdot \left(10^3 \frac{\text{kg}}{\text{m}^3} \right) = 970 \frac{\text{kg}}{\text{m}^3}.$$

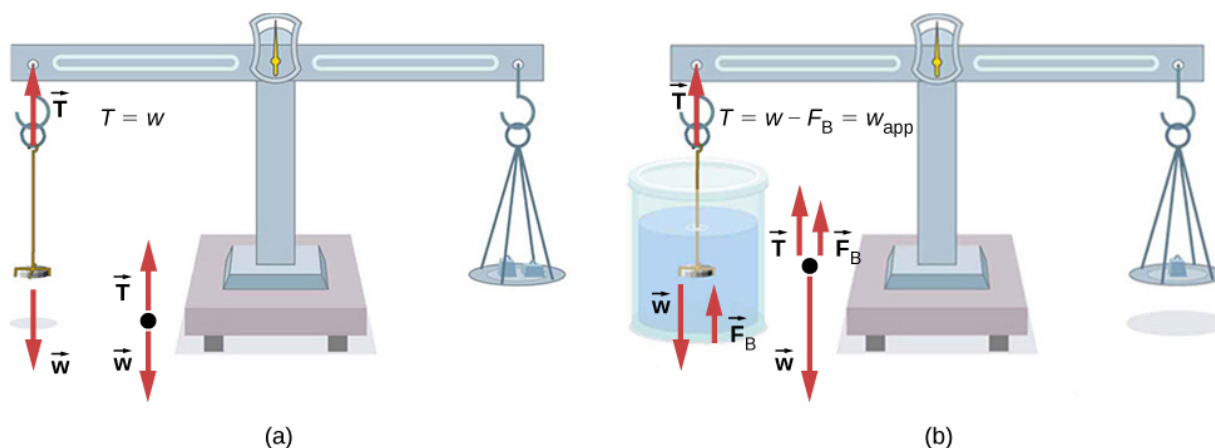
Significance

The woman's density is less than the fluid density. We expect this because she floats.

Numerous lower-density objects or substances float in higher-density fluids: oil on water, a hot-air balloon in the atmosphere, a bit of cork in wine, an iceberg in salt water, and hot wax in a “lava lamp,” to name a few. A less obvious example is mountain ranges floating on the higher-density crust and mantle beneath them. Even seemingly solid Earth has fluid characteristics.

Measuring Density

One of the most common techniques for determining density is shown in [\[link\]](#).



(a) A coin is weighed in air. (b) The apparent weight of the coin is determined while it is completely submerged in a fluid of known density. These two measurements are used to calculate the density of the coin.

An object, here a coin, is weighed in air and then weighed again while submerged in a liquid. The density of the coin, an indication of its authenticity, can be calculated if the fluid density is known. We can use this same technique to determine the density of the fluid if the density of the coin is known.

All of these calculations are based on Archimedes' principle, which states that the buoyant force on the object equals the weight of the fluid displaced. This, in turn, means that the object appears to weigh less when submerged; we call this measurement the object's apparent weight. The object suffers an apparent weight loss equal to the weight of the fluid displaced. Alternatively, on balances that measure mass, the object suffers an apparent mass loss equal to the mass of fluid displaced. That is, apparent weight loss equals weight of fluid displaced, or apparent mass loss equals mass of fluid displaced.

Summary

- Buoyant force is the net upward force on any object in any fluid. If the buoyant force is greater than the object's weight, the object will rise to

the surface and float. If the buoyant force is less than the object's weight, the object will sink. If the buoyant force equals the object's weight, the object can remain suspended at its present depth. The buoyant force is always present and acting on any object immersed either partially or entirely in a fluid.

- Archimedes' principle states that the buoyant force on an object equals the weight of the fluid it displaces.

Conceptual Questions

Exercise:

Problem:

More force is required to pull the plug in a full bathtub than when it is empty. Does this contradict Archimedes' principle? Explain your answer.

Solution:

Not at all. Pascal's principle says that the change in the pressure is exerted through the fluid. The reason that the full tub requires more force to pull the plug is because of the weight of the water above the plug.

Exercise:

Problem:

Do fluids exert buoyant forces in a "weightless" environment, such as in the space shuttle? Explain your answer.

Exercise:

Problem:

Will the same ship float higher in salt water than in freshwater? Explain your answer.

Solution:

The buoyant force is equal to the weight of the fluid displaced. The greater the density of the fluid, the less fluid that is needed to be displaced to have the weight of the object be supported and to float. Since the density of salt water is higher than that of fresh water, less salt water will be displaced, and the ship will float higher.

Exercise:

Problem:

Marbles dropped into a partially filled bathtub sink to the bottom. Part of their weight is supported by buoyant force, yet the downward force on the bottom of the tub increases by exactly the weight of the marbles. Explain why.

Problems

Exercise:

Problem:

What fraction of ice is submerged when it floats in freshwater, given the density of water at 0 °C is very close to 1000 kg/m³?

Exercise:

Problem:

If a person's body has a density of 995 kg/m³, what fraction of the body will be submerged when floating gently in (a) freshwater? (b) In salt water with a density of 1027 kg/m³?

Solution:

a. 99.5% submerged; b. 96.9% submerged

Exercise:

Problem:

A rock with a mass of 540 g in air is found to have an apparent mass of 342 g when submerged in water. (a) What mass of water is displaced? (b) What is the volume of the rock? (c) What is its average density? Is this consistent with the value for granite?

Exercise:**Problem:**

Archimedes' principle can be used to calculate the density of a fluid as well as that of a solid. Suppose a chunk of iron with a mass of 390.0 g in air is found to have an apparent mass of 350.5 g when completely submerged in an unknown liquid. (a) What mass of fluid does the iron displace? (b) What is the volume of iron, using its density as given in [\[link\]](#)? (c) Calculate the fluid's density and identify it.

Solution:

a. 39.5 g; b. 50 cm³; c. 0.79 g/cm³; ethyl alcohol

Exercise:**Problem:**

Calculate the buoyant force on a 2.00-L helium balloon. (b) Given the mass of the rubber in the balloon is 1.50 g, what is the net vertical force on the balloon if it is let go? Neglect the volume of the rubber.

Exercise:**Problem:**

What is the density of a woman who floats in fresh water with 4.00% of her volume above the surface? (This could be measured by placing her in a tank with marks on the side to measure how much water she displaces when floating and when held under water.) (b) What percent of her volume is above the surface when she floats in seawater?

Solution:

a. 960 kg/m^3 ; b. 6.34%; She floats higher in seawater.

Exercise:

Problem:

A man has a mass of 80 kg and a density of 955 kg/m^3 (excluding the air in his lungs). (a) Calculate his volume. (b) Find the buoyant force air exerts on him. (c) What is the ratio of the buoyant force to his weight?

Exercise:

Problem:

A simple compass can be made by placing a small bar magnet on a cork floating in water. (a) What fraction of a plain cork will be submerged when floating in water? (b) If the cork has a mass of 10.0 g and a 20.0-g magnet is placed on it, what fraction of the cork will be submerged? (c) Will the bar magnet and cork float in ethyl alcohol?

Solution:

a. 0.24; b. 0.72; c. Yes, the cork will float in ethyl alcohol.

Exercise:

Problem:

What percentage of an iron anchor's weight will be supported by buoyant force when submerged in salt water?

Exercise:

Problem:

Referring to [\[link\]](#), prove that the buoyant force on the cylinder is equal to the weight of the fluid displaced (Archimedes' principle). You may assume that the buoyant force is $F_2 - F_1$ and that the ends of the cylinder have equal areas A . Note that the volume of the cylinder (and that of the fluid it displaces) equals $(h_2 - h_1)A$.

Solution:

$$\begin{aligned}\text{net } F &= F_2 - F_1 = p_2 A - p_1 A = (p_2 - p_1) A = (h_2 \rho_{\text{fl}} g - h_1 \rho_{\text{fl}} g) A \\ &= (h_2 - h_1) \rho_{\text{fl}} g A, \text{ where } \rho_{\text{fl}} = \text{density of fluid.} \\ \text{net } F &= (h_2 - h_1) A \rho_{\text{fl}} g = V_{\text{fl}} \rho_{\text{fl}} g = m_{\text{fl}} g = w_{\text{fl}}\end{aligned}$$

Exercise:

Problem:

A 75.0-kg man floats in freshwater with 3.00% of his volume above water when his lungs are empty, and 5.00% of his volume above water when his lungs are full. Calculate the volume of air he inhales—called his lung capacity—in liters. (b) Does this lung volume seem reasonable?

Glossary

Archimedes' principle

buoyant force on an object equals the weight of the fluid it displaces

buoyant force

net upward force on any object in any fluid due to the pressure difference at different depths

Fluid Dynamics

By the end of this section, you will be able to:

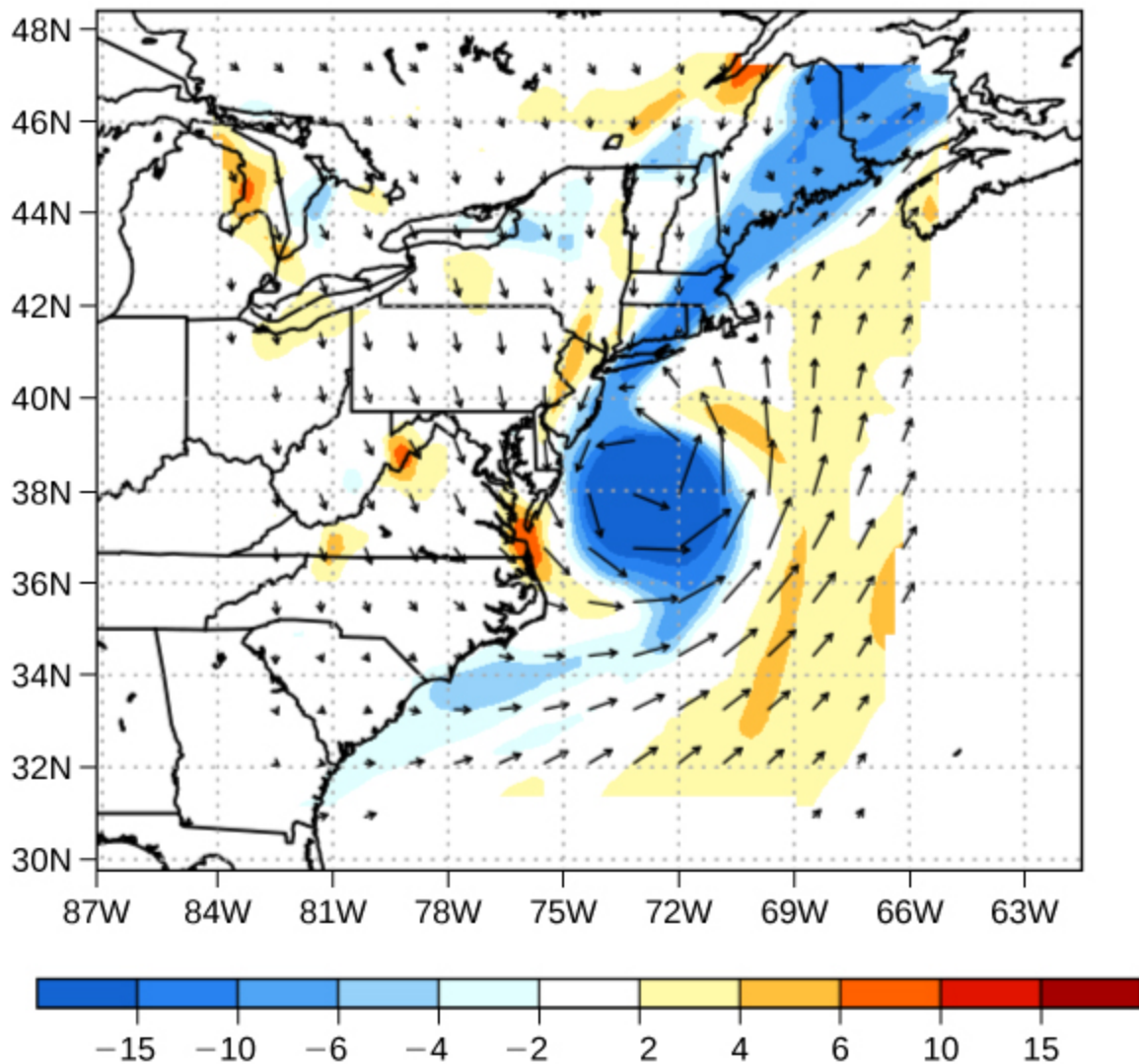
- Describe the characteristics of flow
- Calculate flow rate
- Describe the relationship between flow rate and velocity
- Explain the consequences of the equation of continuity to the conservation of mass

The first part of this chapter dealt with fluid statics, the study of fluids at rest. The rest of this chapter deals with fluid dynamics, the study of fluids in motion. Even the most basic forms of fluid motion can be quite complex. For this reason, we limit our investigation to **ideal fluids** in many of the examples. An ideal fluid is a fluid with negligible **viscosity**. Viscosity is a measure of the internal friction in a fluid; we examine it in more detail in [Viscosity and Turbulence](#). In a few examples, we examine an incompressible fluid—one for which an extremely large force is required to change the volume—since the density in an incompressible fluid is constant throughout.

Characteristics of Flow

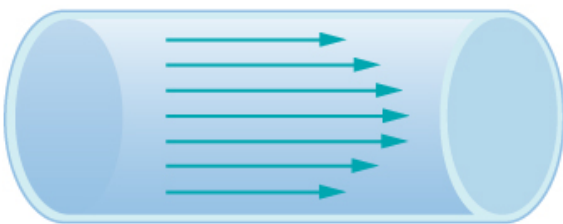
Velocity vectors are often used to illustrate fluid motion in applications like meteorology. For example, wind—the fluid motion of air in the atmosphere—can be represented by vectors indicating the speed and direction of the wind at any given point on a map. [\[link\]](#) shows velocity vectors describing the winds during Hurricane Arthur in 2014.

925hpa Rel.Vorticity, Winds & 2mTmp Forecast Valid 19Z04JUL2014

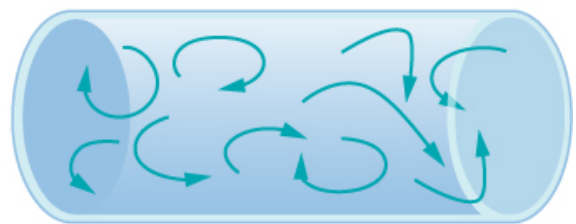


The velocity vectors show the flow of wind in Hurricane Arthur. Notice the circulation of the wind around the eye of the hurricane. Wind speeds are highest near the eye. The colors represent the relative vorticity, a measure of turning or spinning of the air. (credit: modification of work by Joseph Trout, Stockton University)

Another method for representing fluid motion is a *streamline*. A streamline represents the path of a small volume of fluid as it flows. The velocity is always tangential to the streamline. The diagrams in [\[link\]](#) use streamlines to illustrate two examples of fluids moving through a pipe. The first fluid exhibits a **laminar flow** (sometimes described as a steady flow), represented by smooth, parallel streamlines. Note that in the example shown in part (a), the velocity of the fluid is greatest in the center and decreases near the walls of the pipe due to the viscosity of the fluid and friction between the pipe walls and the fluid. This is a special case of laminar flow, where the friction between the pipe and the fluid is high, known as no slip boundary conditions. The second diagram represents **turbulent flow**, in which streamlines are irregular and change over time. In turbulent flow, the paths of the fluid flow are irregular as different parts of the fluid mix together or form small circular regions that resemble whirlpools. This can occur when the speed of the fluid reaches a certain critical speed.



(a) Laminar Flow



(b) Turbulent Flow

(a) Laminar flow can be thought of as layers of fluid moving in parallel, regular paths. (b) In turbulent flow, regions of fluid move in irregular, colliding paths, resulting in mixing and swirling.

Flow Rate and its Relation to Velocity

The volume of fluid passing by a given location through an area during a period of time is called **flow rate** Q , or more precisely, volume flow rate. In symbols, this is written as

Note:

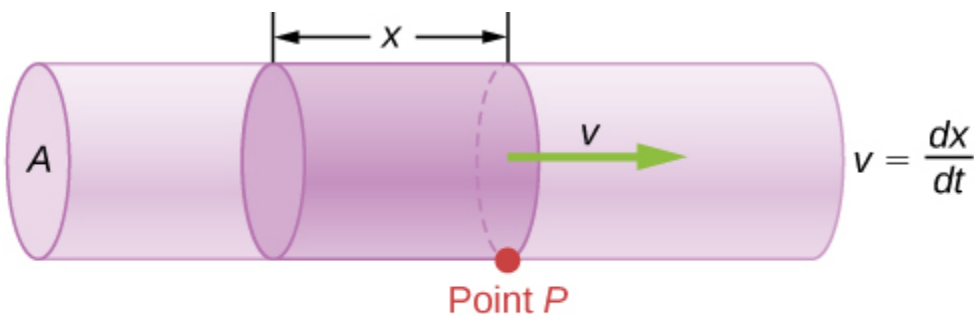
Equation:

$$Q = \frac{dV}{dt}$$

where V is the volume and t is the elapsed time. In [\[link\]](#), the volume of the cylinder is Ax , so the flow rate is

Equation:

$$Q = \frac{dV}{dt} = \frac{d}{dt}(Ax) = A \frac{dx}{dt} = Av.$$



$$Q = \frac{dV}{dt} = \frac{d}{dt}(Ax) = A \frac{dx}{dt} = Av$$

Flow rate is the volume of fluid flowing past a point through the area A per unit time. Here, the shaded cylinder of fluid flows past point P in a uniform pipe in time t .

The SI unit for flow rate is m^3/s , but several other units for Q are in common use, such as liters per minute (L/min). Note that a liter (L) is

1/1000 of a cubic meter or 1000 cubic centimeters (10^{-3} m^3 or 10^3 cm^3).

Flow rate and velocity are related, but quite different, physical quantities. To make the distinction clear, consider the flow rate of a river. The greater the velocity of the water, the greater the flow rate of the river. But flow rate also depends on the size and shape of the river. A rapid mountain stream carries far less water than the Amazon River in Brazil, for example. [\[link\]](#) illustrates the volume flow rate. The volume flow rate is $Q = \frac{dV}{dt} = Av$, where A is the cross-sectional area of the pipe and v is the magnitude of the velocity.

The precise relationship between flow rate Q and average speed v is
Equation:

$$Q = Av,$$

where A is the cross-sectional area and v is the average speed. The relationship tells us that flow rate is directly proportional to both the average speed of the fluid and the cross-sectional area of a river, pipe, or other conduit. The larger the conduit, the greater its cross-sectional area. [\[link\]](#) illustrates how this relationship is obtained. The shaded cylinder has a volume $V = Ad$, which flows past the point P in a time t . Dividing both sides of this relationship by t gives

Equation:

$$\frac{V}{t} = \frac{Ad}{t}.$$

We note that $Q = V/t$ and the average speed is $v = d/t$. Thus the equation becomes $Q = Av$.

[\[link\]](#) shows an incompressible fluid flowing along a pipe of decreasing radius. Because the fluid is incompressible, the same amount of fluid must flow past any point in the tube in a given time to ensure continuity of flow. The flow is continuous because there are no sources or sinks that add or remove mass, so the mass flowing into the pipe must be equal the mass

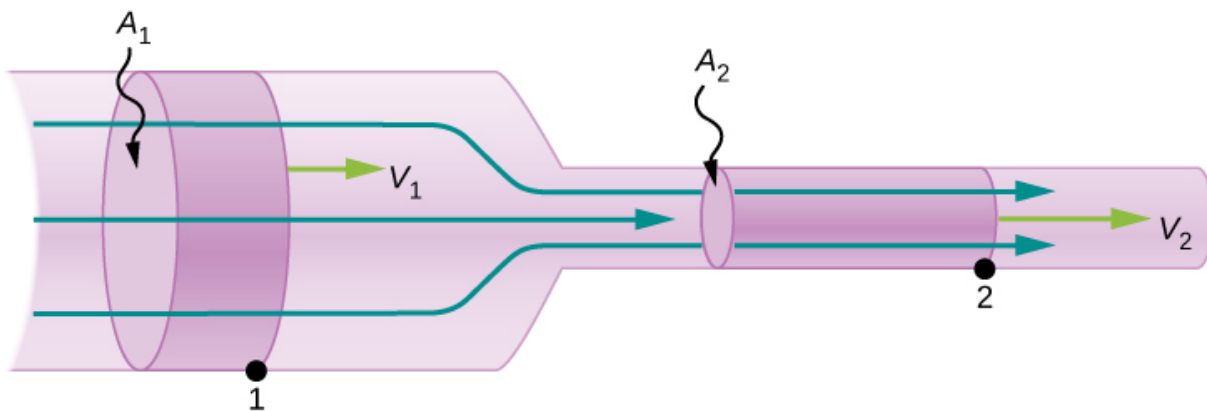
flowing out of the pipe. In this case, because the cross-sectional area of the pipe decreases, the velocity must necessarily increase. This logic can be extended to say that the flow rate must be the same at all points along the pipe. In particular, for arbitrary points 1 and 2,

Note:

Equation:

$$\begin{aligned}Q_1 &= Q_2, \\A_1 v_1 &= A_2 v_2.\end{aligned}$$

This is called the *equation of continuity* and is valid for any incompressible fluid (with constant density). The consequences of the equation of continuity can be observed when water flows from a hose into a narrow spray nozzle: It emerges with a large speed—that is the purpose of the nozzle. Conversely, when a river empties into one end of a reservoir, the water slows considerably, perhaps picking up speed again when it leaves the other end of the reservoir. In other words, speed increases when cross-sectional area decreases, and speed decreases when cross-sectional area increases.



When a tube narrows, the same volume occupies a greater length. For the same volume to pass points 1 and 2 in a given time, the speed must be greater at point 2. The process is exactly reversible. If the fluid flows in the opposite direction, its speed decreases when the tube widens. (Note that the relative volumes of the two cylinders and the corresponding velocity vector arrows are not drawn to scale.)

Since liquids are essentially incompressible, the equation of continuity is valid for all liquids. However, gases are compressible, so the equation must be applied with caution to gases if they are subjected to compression or expansion.

Example:**Calculating Fluid Speed through a Nozzle**

A nozzle with a diameter of 0.500 cm is attached to a garden hose with a radius of 0.900 cm. The flow rate through hose and nozzle is 0.500 L/s. Calculate the speed of the water (a) in the hose and (b) in the nozzle.

Strategy

We can use the relationship between flow rate and speed to find both speeds. We use the subscript 1 for the hose and 2 for the nozzle.

Solution

- a. We solve the flow rate equation for speed and use πr_1^2 for the cross-sectional area of the hose, obtaining

Equation:

$$v = \frac{Q}{A} = \frac{Q}{\pi r_1^2}.$$

Substituting values and using appropriate unit conversions yields

Equation:

$$v = \frac{(0.500 \text{ L/s})(10^{-3} \text{ m}^3/\text{L})}{3.14(9.00 \times 10^{-3} \text{ m})^2} = 1.96 \text{ m/s}.$$

- b. We could repeat this calculation to find the speed in the nozzle v_2 , but we use the equation of continuity to give a somewhat different insight. The equation states

Equation:

$$A_1 v_1 = A_2 v_2.$$

Solving for v_2 and substituting πr^2 for the cross-sectional area yields

Equation:

$$v_2 = \frac{A_1}{A_2} v_1 = \frac{\pi r_1^2}{\pi r_2^2} v_1 = \frac{r_1^2}{r_2^2} v_1.$$

Substituting known values,

Equation:

$$v_2 = \frac{(0.900 \text{ cm})^2}{(0.250 \text{ cm})^2} 1.96 \text{ m/s} = 25.5 \text{ m/s}.$$

Significance

A speed of 1.96 m/s is about right for water emerging from a hose with no nozzle. The nozzle produces a considerably faster stream merely by constricting the flow to a narrower tube.

The solution to the last part of the example shows that speed is inversely proportional to the square of the radius of the tube, making for large effects when radius varies. We can blow out a candle at quite a distance, for example, by pursing our lips, whereas blowing on a candle with our mouth wide open is quite ineffective.

Mass Conservation

The rate of flow of a fluid can also be described by the *mass flow rate* or mass rate of flow. This is the rate at which a mass of the fluid moves past a point. Refer once again to [\[link\]](#), but this time consider the mass in the shaded volume. The mass can be determined from the density and the volume:

Equation:

$$m = \rho V = \rho Ax.$$

The mass flow rate is then

Equation:

$$\frac{dm}{dt} = \frac{d}{dt}(\rho Ax) = \rho A \frac{dx}{dt} = \rho Av,$$

where ρ is the density, A is the cross-sectional area, and v is the magnitude of the velocity. The mass flow rate is an important quantity in fluid dynamics and can be used to solve many problems. Consider [\[link\]](#). The pipe in the figure starts at the inlet with a cross sectional area of A_1 and constricts to an outlet with a smaller cross sectional area of A_2 . The mass of fluid entering the pipe has to be equal to the mass of fluid leaving the pipe. For this reason the velocity at the outlet (v_2) is greater than the velocity of the inlet (v_1). Using the fact that the mass of fluid entering the pipe must be equal to the mass of fluid exiting the pipe, we can find a relationship between the velocity and the cross-sectional area by taking the rate of change of the mass in and the mass out:

Note:

Equation:

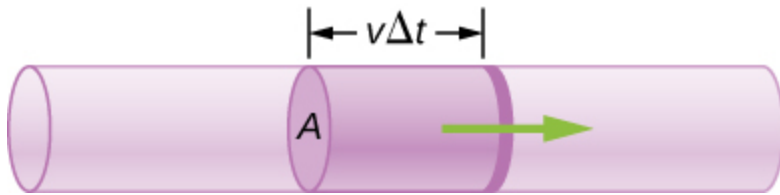
$$\begin{aligned} \left(\frac{dm}{dt}\right)_1 &= \left(\frac{dm}{dt}\right)_2 \\ \rho_1 A_1 v_1 &= \rho_2 A_2 v_2. \end{aligned}$$

[\[link\]](#) is also known as the continuity equation in general form. If the density of the fluid remains constant through the constriction—that is, the fluid is incompressible—then the density cancels from the continuity equation,

Equation:

$$A_1 v_1 = A_2 v_2.$$

The equation reduces to show that the volume flow rate into the pipe equals the volume flow rate out of the pipe.



Geometry for deriving the equation of continuity. The amount of liquid entering the cross-sectional (shaded) area must equal the amount of liquid leaving the cross-sectional area if the liquid is incompressible.

Summary

- Flow rate Q is defined as the volume V flowing past a point in time t , or $Q = \frac{dV}{dt}$ where V is volume and t is time. The SI unit of flow rate is m^3/s , but other rates can be used, such as L/min.
- Flow rate and velocity are related by $Q = Av$ where A is the cross-sectional area of the flow and v is its average velocity.
- The equation of continuity states that for an incompressible fluid, the mass flowing into a pipe must equal the mass flowing out of the pipe.

Conceptual Questions

Exercise:

Problem:

Many figures in the text show streamlines. Explain why fluid velocity is greatest where streamlines are closest together. (*Hint:* Consider the relationship between fluid velocity and the cross-sectional area through which the fluid flows.)

Solution:

Consider two different pipes connected to a single pipe of a smaller diameter, with fluid flowing from the two pipes into the smaller pipe. Since the fluid is forced through a smaller cross-sectional area, it must move faster as the flow lines become closer together. Likewise, if a pipe with a large radius feeds into a pipe with a small radius, the stream lines will become closer together and the fluid will move faster.

Problems

Exercise:

Problem:

What is the average flow rate in cm^3/s of gasoline to the engine of a car traveling at 100 km/h if it averages 10.0 km/L?

Solution:

$$2.77 \text{ cm}^3/\text{s}$$

Exercise:

Problem:

The heart of a resting adult pumps blood at a rate of 5.00 L/min. (a) Convert this to cm^3/s . (b) What is this rate in m^3/s ?

Exercise:**Problem:**

The Huka Falls on the Waikato River is one of New Zealand's most visited natural tourist attractions. On average, the river has a flow rate of about 300,000 L/s. At the gorge, the river narrows to 20-m wide and averages 20-m deep. (a) What is the average speed of the river in the gorge? (b) What is the average speed of the water in the river downstream of the falls when it widens to 60 m and its depth increases to an average of 40 m?

Solution:

a. 0.75 m/s; b. 0.13 m/s

Exercise:**Problem:**

(a) Estimate the time it would take to fill a private swimming pool with a capacity of 80,000 L using a garden hose delivering 60 L/min. (b) How long would it take if you could divert a moderate size river, flowing at $5000 \text{ m}^3/\text{s}$ into the pool?

Exercise:**Problem:**

What is the fluid speed in a fire hose with a 9.00-cm diameter carrying 80.0 L of water per second? (b) What is the flow rate in cubic meters per second? (c) Would your answers be different if salt water replaced the fresh water in the fire hose?

Solution:

a. 12.6 m/s; b. $0.0800 \text{ m}^3/\text{s}$; c. No, the flow rate and the velocity are independent of the density of the fluid.

Exercise:

Problem:

Water is moving at a velocity of 2.00 m/s through a hose with an internal diameter of 1.60 cm. (a) What is the flow rate in liters per second? (b) The fluid velocity in this hose's nozzle is 15.0 m/s. What is the nozzle's inside diameter?

Exercise:**Problem:**

Prove that the speed of an incompressible fluid through a constriction, such as in a Venturi tube, increases by a factor equal to the square of the factor by which the diameter decreases. (The converse applies for flow out of a constriction into a larger-diameter region.)

Solution:

If the fluid is incompressible, the flow rate through both sides will be equal:

$$Q = A_1 \bar{v}_1 = A_2 \bar{v}_2, \text{ or}$$
$$\pi \frac{d_1^2}{4} \bar{v}_1 = \pi \frac{d_2^2}{4} \bar{v}_2 \Rightarrow \bar{v}_2 = \bar{v}_1 (d_1^2 / d_2^2) = \bar{v}_1 (d_1 / d_2)^2$$

Exercise:**Problem:**

Water emerges straight down from a faucet with a 1.80-cm diameter at a speed of 0.500 m/s. (Because of the construction of the faucet, there is no variation in speed across the stream.) (a) What is the flow rate in cm^3/s ? (b) What is the diameter of the stream 0.200 m below the faucet? Neglect any effects due to surface tension.

Glossary

flow rate

abbreviated Q , it is the volume V that flows past a particular point during a time t , or $Q = dV/dt$

ideal fluid

fluid with negligible viscosity

laminar flow

type of fluid flow in which layers do not mix

turbulent flow

type of fluid flow in which layers mix together via eddies and swirls

viscosity

measure of the internal friction in a fluid

Bernoulli's Equation

By the end of this section, you will be able to:

- Explain the terms in Bernoulli's equation
- Explain how Bernoulli's equation is related to the conservation of energy
- Describe how to derive Bernoulli's principle from Bernoulli's equation
- Perform calculations using Bernoulli's principle
- Describe some applications of Bernoulli's principle

As we showed in [\[link\]](#), when a fluid flows into a narrower channel, its speed increases. That means its kinetic energy also increases. The increased kinetic energy comes from the net work done on the fluid to push it into the channel. Also, if the fluid changes vertical position, work is done on the fluid by the gravitational force.

A pressure difference occurs when the channel narrows. This pressure difference results in a net force on the fluid because the pressure times the area equals the force, and this net force does work. Recall the work-energy theorem,

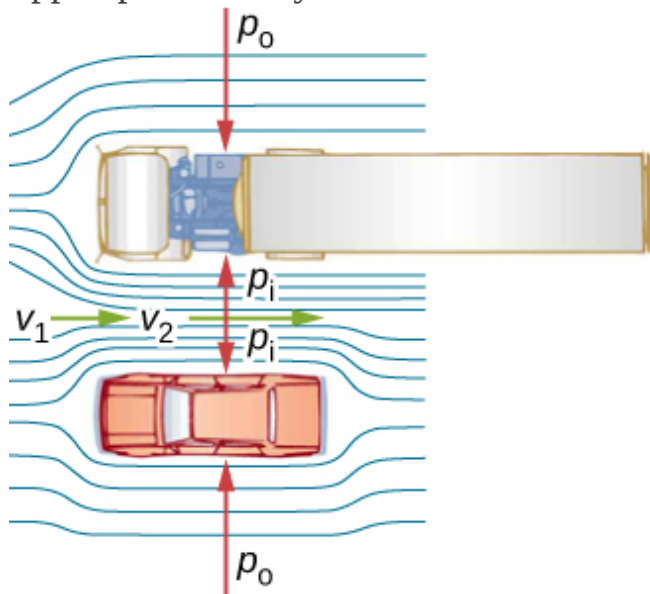
Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

The net work done increases the fluid's kinetic energy. As a result, the pressure drops in a rapidly moving fluid whether or not the fluid is confined to a tube.

There are many common examples of pressure dropping in rapidly moving fluids. For instance, shower curtains have a disagreeable habit of bulging into the shower stall when the shower is on. The reason is that the high-velocity stream of water and air creates a region of lower pressure inside the shower, whereas the pressure on the other side remains at the standard atmospheric pressure. This pressure difference results in a net force, pushing the curtain inward. Similarly, when a car passes a truck on the highway, the two vehicles seem to pull toward each other. The reason is the same: The

high velocity of the air between the car and the truck creates a region of lower pressure between the vehicles, and they are pushed together by greater pressure on the outside ([link](#)). This effect was observed as far back as the mid-1800s, when it was found that trains passing in opposite directions tipped precariously toward one another.



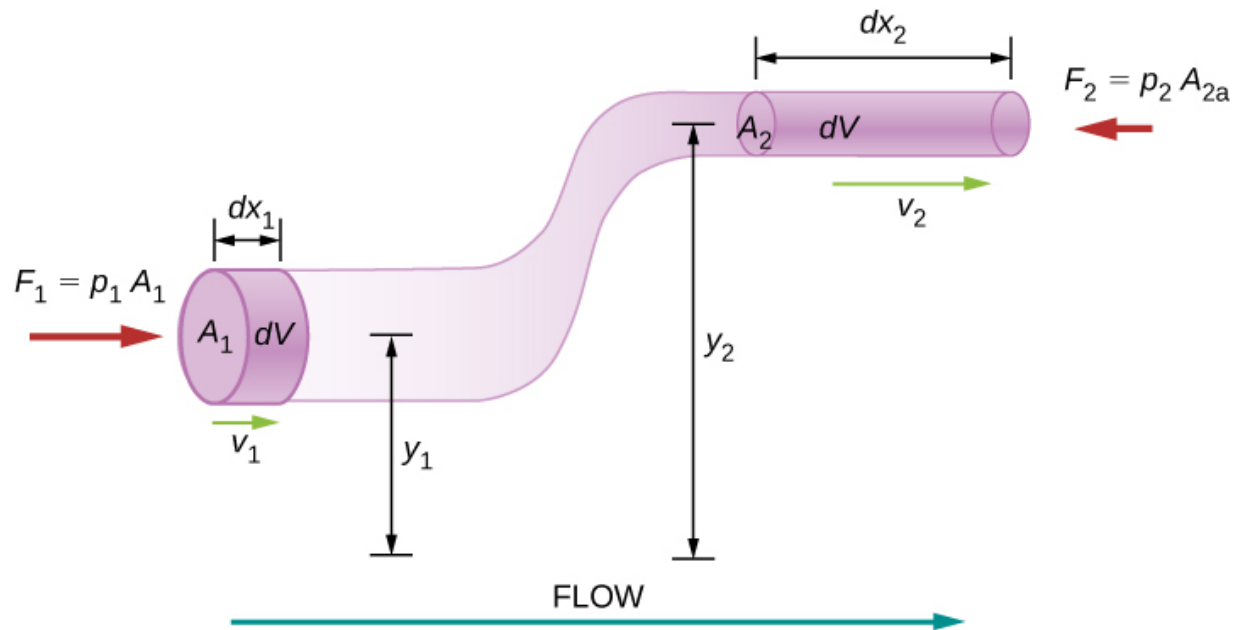
An overhead view of a car passing a truck on a highway. Air passing between the vehicles flows in a narrower channel and must increase its speed (v_2 is greater than v_1), causing the pressure between them to drop (p_i is less than p_o). Greater pressure on the outside pushes the car and truck together.

Energy Conservation and Bernoulli's Equation

The application of the principle of conservation of energy to frictionless laminar flow leads to a very useful relation between pressure and flow speed in a fluid. This relation is called **Bernoulli's equation**, named after Daniel

Bernoulli (1700–1782), who published his studies on fluid motion in his book *Hydrodynamica* (1738).

Consider an incompressible fluid flowing through a pipe that has a varying diameter and height, as shown in [\[link\]](#). Subscripts 1 and 2 in the figure denote two locations along the pipe and illustrate the relationships between the areas of the cross sections A , the speed of flow v , the height from ground y , and the pressure p at each point. We assume here that the density at the two points is the same—therefore, density is denoted by ρ without any subscripts—and since the fluid is incompressible, the shaded volumes must be equal.



The geometry used for the derivation of Bernoulli's equation.

We also assume that there are no viscous forces in the fluid, so the energy of any part of the fluid will be conserved. To derive Bernoulli's equation, we first calculate the work that was done on the fluid:

Equation:

$$dW = F_1 dx_1 - F_2 dx_2$$

Equation:

$$dW = p_1 A_1 dx_1 - p_2 A_2 dx_2 = p_1 dV - p_2 dV = (p_1 - p_2) dV.$$

The work done was due to the conservative force of gravity and the change in the kinetic energy of the fluid. The change in the kinetic energy of the fluid is equal to

Equation:

$$dK = \frac{1}{2} m_2 v_2^2 - \frac{1}{2} m_1 v_1^2 = \frac{1}{2} \rho dV (v_2^2 - v_1^2).$$

The change in potential energy is

Equation:

$$dU = mgy_2 - mgy_1 = \rho dV g (y_2 - y_1).$$

The energy equation then becomes

Equation:

$$\begin{aligned} dW &= dK + dU \\ (p_1 - p_2) dV &= \frac{1}{2} \rho dV (v_2^2 - v_1^2) + \rho dV g (y_2 - y_1) \\ (p_1 - p_2) &= \frac{1}{2} \rho (v_2^2 - v_1^2) + \rho g (y_2 - y_1). \end{aligned}$$

Rearranging the equation gives Bernoulli's equation:

Equation:

$$p_1 + \frac{1}{2} \rho v_1^2 + \rho g y_1 = p_2 + \frac{1}{2} \rho v_2^2 + \rho g y_2.$$

This relation states that the mechanical energy of any part of the fluid changes as a result of the work done by the fluid external to that part, due to

varying pressure along the way. Since the two points were chosen arbitrarily, we can write Bernoulli's equation more generally as a conservation principle along the flow.

Note:

Bernoulli's Equation

For an incompressible, frictionless fluid, the combination of pressure and the sum of kinetic and potential energy densities is constant not only over time, but also along a streamline:

Equation:

$$p + \frac{1}{2}\rho v^2 + \rho gy = \text{constant}$$

A special note must be made here of the fact that in a dynamic situation, the pressures at the same height in different parts of the fluid may be different if they have different speeds of flow.

Analyzing Bernoulli's Equation

According to Bernoulli's equation, if we follow a small volume of fluid along its path, various quantities in the sum may change, but the total remains constant. Bernoulli's equation is, in fact, just a convenient statement of conservation of energy for an incompressible fluid in the absence of friction.

The general form of Bernoulli's equation has three terms in it, and it is broadly applicable. To understand it better, let us consider some specific situations that simplify and illustrate its use and meaning.

Bernoulli's equation for static fluids

First consider the very simple situation where the fluid is static—that is, $v_1 = v_2 = 0$. Bernoulli's equation in that case is

Equation:

$$p_1 + \rho gh_1 = p_2 + \rho gh_2.$$

We can further simplify the equation by setting $h_2 = 0$. (Any height can be chosen for a reference height of zero, as is often done for other situations involving gravitational force, making all other heights relative.) In this case, we get

Equation:

$$p_2 = p_1 + \rho gh_1.$$

This equation tells us that, in static fluids, pressure increases with depth. As we go from point 1 to point 2 in the fluid, the depth increases by h_1 , and consequently, p_2 is greater than p_1 by an amount ρgh_1 . In the very simplest case, p_1 is zero at the top of the fluid, and we get the familiar relationship $p = \rho gh$. (Recall that $p = \rho gh$ and $\Delta U_g = -mgh$.) Thus, Bernoulli's equation confirms the fact that the pressure change due to the weight of a fluid is ρgh . Although we introduce Bernoulli's equation for fluid motion, it includes much of what we studied for static fluids earlier.

Bernoulli's principle

Suppose a fluid is moving but its depth is constant—that is, $h_1 = h_2$. Under this condition, Bernoulli's equation becomes

Equation:

$$p_1 + \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho v_2^2.$$

Situations in which fluid flows at a constant depth are so common that this equation is often also called **Bernoulli's principle**, which is simply

Bernoulli's equation for fluids at constant depth. (Note again that this applies to a small volume of fluid as we follow it along its path.) Bernoulli's principle reinforces the fact that pressure drops as speed increases in a moving fluid: If v_2 is greater than v_1 in the equation, then p_2 must be less than p_1 for the equality to hold.

Example:**Calculating Pressure**

In [\[link\]](#), we found that the speed of water in a hose increased from 1.96 m/s to 25.5 m/s going from the hose to the nozzle. Calculate the pressure in the hose, given that the absolute pressure in the nozzle is $1.01 \times 10^5 \text{ N/m}^2$ (atmospheric, as it must be) and assuming level, frictionless flow.

Strategy

Level flow means constant depth, so Bernoulli's principle applies. We use the subscript 1 for values in the hose and 2 for those in the nozzle. We are thus asked to find p_1 .

Solution

Solving Bernoulli's principle for p_1 yields

Equation:

$$p_1 = p_2 + \frac{1}{2}\rho v_2^2 - \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho(v_2^2 - v_1^2).$$

Substituting known values,

Equation:

$$\begin{aligned} p_1 &= 1.01 \times 10^5 \text{ N/m}^2 + \frac{1}{2}(10^3 \text{ kg/m}^3)[(25.5 \text{ m/s})^2 - (1.96 \text{ m/s})^2] \\ &= 4.24 \times 10^5 \text{ N/m}^2. \end{aligned}$$

Significance

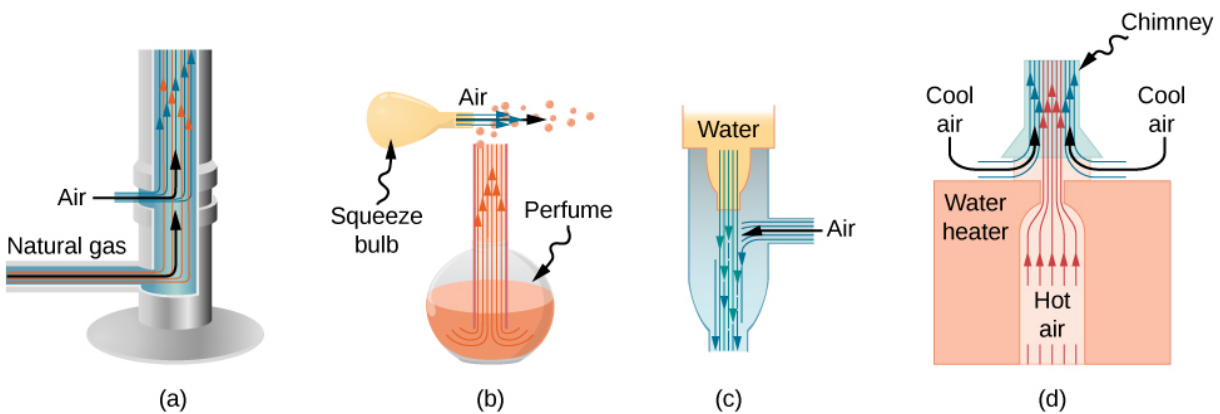
This absolute pressure in the hose is greater than in the nozzle, as expected, since v is greater in the nozzle. The pressure p_2 in the nozzle must be atmospheric, because the water emerges into the atmosphere without other changes in conditions.

Applications of Bernoulli's Principle

Many devices and situations occur in which fluid flows at a constant height and thus can be analyzed with Bernoulli's principle.

Entrainment

People have long put the Bernoulli principle to work by using reduced pressure in high-velocity fluids to move things about. With a higher pressure on the outside, the high-velocity fluid forces other fluids into the stream. This process is called *entrainment*. Entrainment devices have been in use since ancient times as pumps to raise water to small heights, as is necessary for draining swamps, fields, or other low-lying areas. Some other devices that use the concept of entrainment are shown in [\[link\]](#).



Entrainment devices use increased fluid speed to create low pressures, which then entrain one fluid into another. (a) A Bunsen burner uses an adjustable gas nozzle, entraining air for proper combustion. (b) An atomizer uses a squeeze bulb to create a jet of air that entrains drops of perfume. Paint sprayers and carburetors use very similar techniques to move their respective liquids. (c) A common aspirator uses a high-speed stream of water to create a region of lower pressure. Aspirators may be used as suction pumps in dental and surgical situations or for draining a flooded basement or producing a reduced pressure in a

vessel. (d) The chimney of a water heater is designed to entrain air into the pipe leading through the ceiling.

Velocity measurement

[\[link\]](#) shows two devices that apply Bernoulli's principle to measure fluid velocity. The manometer in part (a) is connected to two tubes that are small enough not to appreciably disturb the flow. The tube facing the oncoming fluid creates a dead spot having zero velocity ($v_1 = 0$) in front of it, while fluid passing the other tube has velocity v_2 . This means that Bernoulli's principle as stated in

Equation:

$$p_1 + \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho v_2^2$$

becomes

Equation:

$$p_1 = p_2 + \frac{1}{2}\rho v_2^2.$$

Thus pressure p_2 over the second opening is reduced by $\frac{1}{2}\rho v_2^2$, so the fluid in the manometer rises by h on the side connected to the second opening, where

Equation:

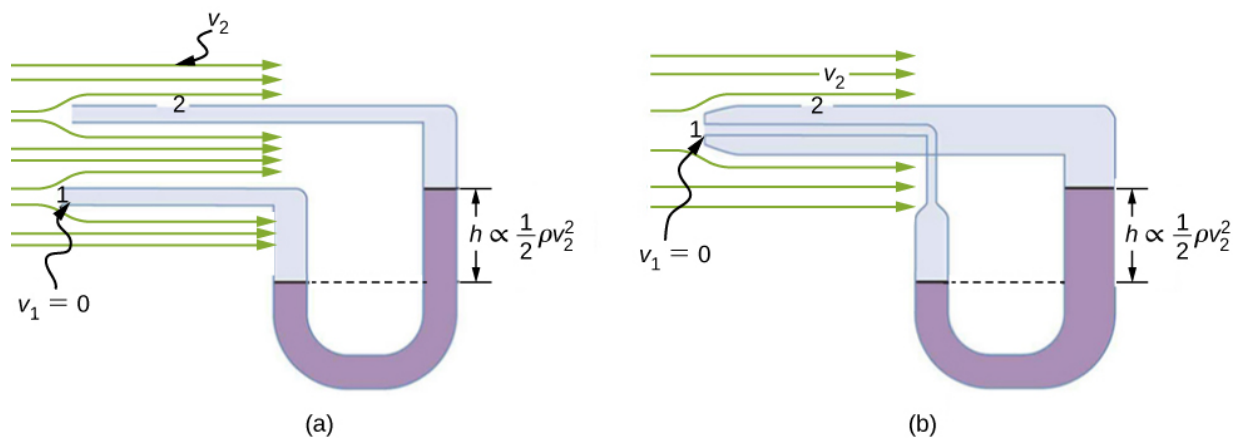
$$h \propto \frac{1}{2}\rho v_2^2.$$

(Recall that the symbol \propto means “proportional to.”) Solving for v_2 , we see that

Equation:

$$v_2 \propto \sqrt{h}.$$

Part (b) shows a version of this device that is in common use for measuring various fluid velocities; such devices are frequently used as air-speed indicators in aircraft.



Measurement of fluid speed based on Bernoulli's principle. (a) A manometer is connected to two tubes that are close together and small enough not to disturb the flow. Tube 1 is open at the end facing the flow. A dead spot having zero speed is created there. Tube 2 has an opening on the side, so the fluid has a speed v across the opening; thus, pressure there drops. The difference in pressure at the manometer is $\frac{1}{2} \rho v_2^2$, so h is proportional to $\frac{1}{2} \rho v_2^2$. (b) This type of velocity measuring device is a Prandtl tube, also known as a pitot tube.

A fire hose

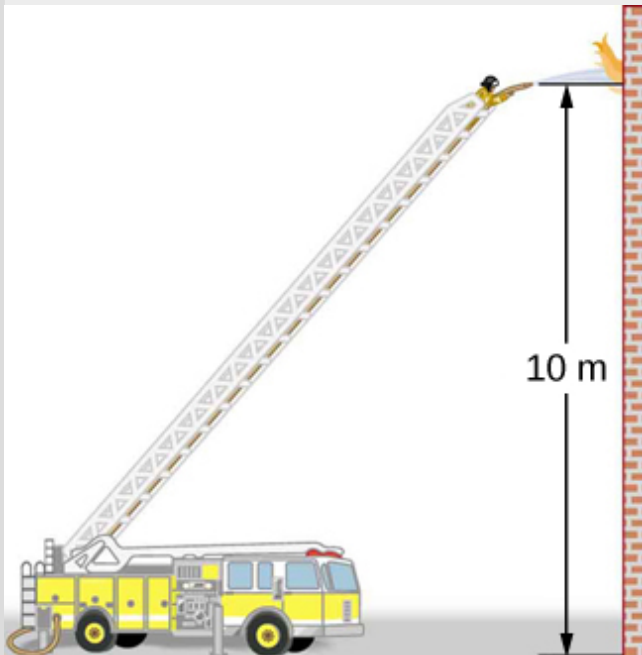
All preceding applications of Bernoulli's equation involved simplifying conditions, such as constant height or constant pressure. The next example is

a more general application of Bernoulli's equation in which pressure, velocity, and height all change.

Example:

Calculating Pressure: A Fire Hose Nozzle

Fire hoses used in major structural fires have an inside diameter of 6.40 cm ([\[link\]](#)). Suppose such a hose carries a flow of 40.0 L/s, starting at a gauge pressure of $1.62 \times 10^6 \text{ N/m}^2$. The hose rises up 10.0 m along a ladder to a nozzle having an inside diameter of 3.00 cm. What is the pressure in the nozzle?



Pressure in the nozzle of this fire hose is less than at ground level for two reasons: The water has to go uphill to get to the nozzle, and speed increases in the nozzle. In spite of its lowered pressure, the water can exert a large force on anything it strikes by virtue of its kinetic energy. Pressure in the water stream becomes equal to

atmospheric pressure once it emerges into the air.

Strategy

We must use Bernoulli's equation to solve for the pressure, since depth is not constant.

Solution

Bernoulli's equation is

Equation:

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho gh_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho gh_2$$

where subscripts 1 and 2 refer to the initial conditions at ground level and the final conditions inside the nozzle, respectively. We must first find the speeds v_1 and v_2 . Since $Q = A_1 v_1$, we get

Equation:

$$v_1 = \frac{Q}{A_1} = \frac{40.0 \times 10^{-3} \text{ m}^3/\text{s}}{\pi(3.20 \times 10^{-2} \text{ m})^2} = 12.4 \text{ m/s}.$$

Similarly, we find

Equation:

$$v_2 = 56.6 \text{ m/s}.$$

This rather large speed is helpful in reaching the fire. Now, taking h_1 to be zero, we solve Bernoulli's equation for p_2 :

Equation:

$$p_2 = p_1 + \frac{1}{2}\rho(v_1^2 - v_2^2) - \rho gh_2.$$

Substituting known values yields

Equation:

$$\begin{aligned}
 p_2 &= 1.62 \times 10^6 \text{ N/m}^2 + \frac{1}{2}(1000 \text{ kg/m}^3)[(12.4 \text{ m/s})^2 - (56.6 \text{ m/s})^2] \\
 &\quad - (1000 \text{ kg/m}^3)(9.80 \text{ m/s}^2)(10.0 \text{ m}) \\
 &= -1.82 \text{ kPa} \approx 0 \text{ kPa (when compared to air pressure)}.
 \end{aligned}$$

Significance

This value is a gauge pressure, since the initial pressure was given as a gauge pressure. Thus, the nozzle pressure equals atmospheric pressure because the water exits into the atmosphere without changes in its conditions.

Summary

- Bernoulli's equation states that the sum on each side of the following equation is constant, or the same at any two points in an incompressible frictionless fluid:

Equation:

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho gh_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho gh_2.$$

- Bernoulli's principle is Bernoulli's equation applied to situations in which the height of the fluid is constant. The terms involving depth (or height h) subtract out, yielding

Equation:

$$p_1 + \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho v_2^2.$$

- Bernoulli's principle has many applications, including entrainment and velocity measurement.

Conceptual Questions

Exercise:

Problem:

You can squirt water from a garden hose a considerably greater distance by partially covering the opening with your thumb. Explain how this works.

Exercise:**Problem:**

Water is shot nearly vertically upward in a decorative fountain and the stream is observed to broaden as it rises. Conversely, a stream of water falling straight down from a faucet narrows. Explain why.

Solution:

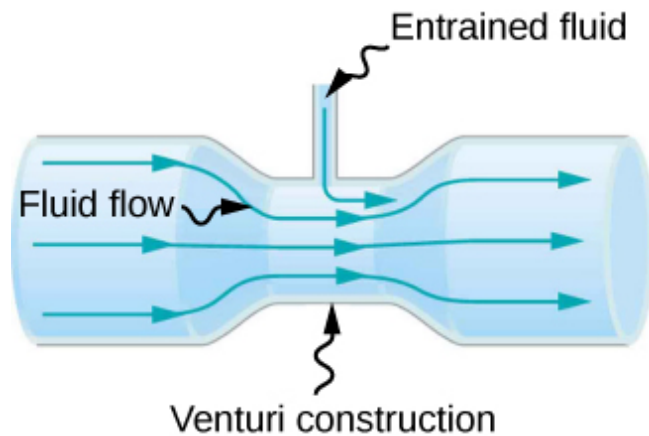
The mass of water that enters a cross-sectional area must equal the amount that leaves. From the continuity equation, we know that the density times the area times the velocity must remain constant. Since the density of the water does not change, the velocity times the cross-sectional area entering a region must equal the cross-sectional area times the velocity leaving the region. Since the velocity of the fountain stream decreases as it rises due to gravity, the area must increase. Since the velocity of the faucet stream speeds up as it falls, the area must decrease.

Exercise:**Problem:**

Look back to [\[link\]](#). Answer the following two questions. Why is p_o less than atmospheric? Why is p_o greater than p_i ?

Exercise:**Problem:**

A tube with a narrow segment designed to enhance entrainment is called a Venturi, such as shown below. Venturis are very commonly used in carburetors and aspirators. How does this structure bolster entrainment?



Solution:

When the tube narrows, the fluid is forced to speed up, thanks to the continuity equation and the work done on the fluid. Where the tube is narrow, the pressure decreases. This means that the entrained fluid will be pushed into the narrow area.

Exercise:**Problem:**

Some chimney pipes have a T-shape, with a crosspiece on top that helps draw up gases whenever there is even a slight breeze. Explain how this works in terms of Bernoulli's principle.

Exercise:**Problem:**

Is there a limit to the height to which an entrainment device can raise a fluid? Explain your answer.

Solution:

The work done by pressure can be used to increase the kinetic energy and to gain potential energy. As the height becomes larger, there is less energy left to give to kinetic energy. Eventually, there will be a maximum height that cannot be overcome.

Exercise:**Problem:**

Why is it preferable for airplanes to take off into the wind rather than with the wind?

Exercise:**Problem:**

Roofs are sometimes pushed off vertically during a tropical cyclone, and buildings sometimes explode outward when hit by a tornado. Use Bernoulli's principle to explain these phenomena.

Solution:

Because of the speed of the air outside the building, the pressure outside the house decreases. The greater pressure inside the building can essentially blow off the roof or cause the building to explode.

Exercise:**Problem:**

It is dangerous to stand close to railroad tracks when a rapidly moving commuter train passes. Explain why atmospheric pressure would push you toward the moving train.

Exercise:**Problem:**

Water pressure inside a hose nozzle can be less than atmospheric pressure due to the Bernoulli effect. Explain in terms of energy how the water can emerge from the nozzle against the opposing atmospheric pressure.

Solution:

The air inside the hose has kinetic energy due to its motion. The kinetic energy can be used to do work against the pressure difference.

Exercise:**Problem:**

David rolled down the window on his car while driving on the freeway. An empty plastic bag on the floor promptly flew out the window. Explain why.

Exercise:**Problem:**

Based on Bernoulli's equation, what are three forms of energy in a fluid? (Note that these forms are conservative, unlike heat transfer and other dissipative forms not included in Bernoulli's equation.)

Solution:

Potential energy due to position, kinetic energy due to velocity, and the work done by a pressure difference.

Exercise:**Problem:**

The old rubber boot shown below has two leaks. To what maximum height can the water squirt from Leak 1? How does the velocity of water emerging from Leak 2 differ from that of Leak 1? Explain your responses in terms of energy.

**Exercise:****Problem:**

Water pressure inside a hose nozzle can be less than atmospheric pressure due to the Bernoulli effect. Explain in terms of energy how the water can emerge from the nozzle against the opposing atmospheric pressure.

Solution:

The water has kinetic energy due to its motion. This energy can be converted into work against the difference in pressure.

Problems**Exercise:**

Problem: Verify that pressure has units of energy per unit volume.

Solution:

$$F = pA \Rightarrow p = \frac{F}{A},$$

$$[p] = \text{N/m}^2 = \text{N} \cdot \text{m/m}^3 = \text{J/m}^3 = \text{energy/volume}$$

Exercise:**Problem:**

Suppose you have a wind speed gauge like the pitot tube shown in [\[link\]](#). By what factor must wind speed increase to double the value of h in the manometer? Is this independent of the moving fluid and the fluid in the manometer?

Exercise:**Problem:**

If the pressure reading of your pitot tube is 15.0 mm Hg at a speed of 200 km/h, what will it be at 700 km/h at the same altitude?

Solution:

−135 mm Hg

Exercise:**Problem:**

Every few years, winds in Boulder, Colorado, attain sustained speeds of 45.0 m/s (about 100 mph) when the jet stream descends during early spring. Approximately what is the force due to the Bernoulli equation on a roof having an area of 220m²? Typical air density in Boulder is 1.14kg/m³, and the corresponding atmospheric pressure is $8.89 \times 10^4 \text{N/m}^2$. (Bernoulli's principle as stated in the text assumes laminar flow. Using the principle here produces only an approximate result, because there is significant turbulence.)

Exercise:

Problem:

What is the pressure drop due to the Bernoulli Effect as water goes into a 3.00-cm-diameter nozzle from a 9.00-cm-diameter fire hose while carrying a flow of 40.0 L/s? (b) To what maximum height above the nozzle can this water rise? (The actual height will be significantly smaller due to air resistance.)

Solution:

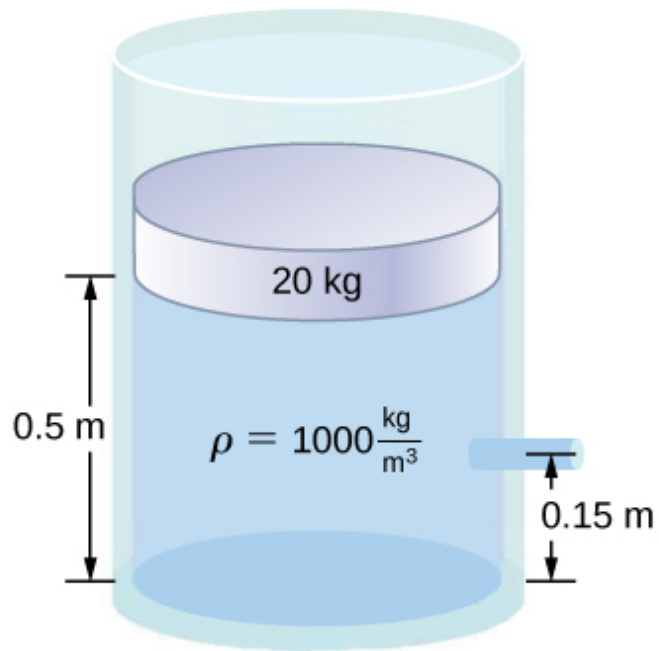
a. $1.58 \times 10^6 \text{ N/m}^2$; b. 163 m

Exercise:**Problem:**

(a) Using Bernoulli's equation, show that the measured fluid speed v for a pitot tube, like the one in [\[link\]](#)(b), is given by $v = \left(\frac{2\rho'gh}{\rho} \right)^{1/2}$, where h is the height of the manometer fluid, ρ' is the density of the manometer fluid, ρ is the density of the moving fluid, and g is the acceleration due to gravity. (Note that v is indeed proportional to the square root of h , as stated in the text.) (b) Calculate v for moving air if a mercury manometer's h is 0.200 m.

Exercise:**Problem:**

A container of water has a cross-sectional area of $A = 0.1 \text{ m}^2$. A piston sits on top of the water (see the following figure). There is a spout located 0.15 m from the bottom of the tank, open to the atmosphere, and a stream of water exits the spout. The cross sectional area of the spout is $A_s = 7.0 \times 10^{-4} \text{ m}^2$. (a) What is the velocity of the water as it leaves the spout? (b) If the opening of the spout is located 1.5 m above the ground, how far from the spout does the water hit the floor? Ignore all friction and dissipative forces.



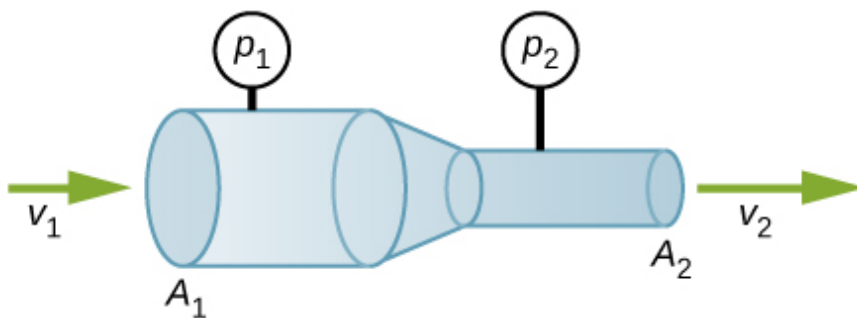
Solution:

- a. $v_2 = 3.28 \frac{\text{m}}{\text{s}}$;
- b. $t = 0.55 \text{ s}$
- $x = vt = 1.81 \text{ m}$

Exercise:

Problem:

A fluid of a constant density flows through a reduction in a pipe. Find an equation for the change in pressure, in terms of v_1 , A_1 , A_2 , and the density.



Glossary

Bernoulli's equation

equation resulting from applying conservation of energy to an incompressible frictionless fluid: $p + \frac{1}{2}\rho v^2 + \rho gh = \text{constant}$, throughout the fluid

Bernoulli's principle

Bernoulli's equation applied at constant depth:

Equation:

$$p_1 + \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho v_2^2$$

Viscosity and Turbulence

By the end of this section, you will be able to:

- Explain what viscosity is
- Calculate flow and resistance with Poiseuille's law
- Explain how pressure drops due to resistance
- Calculate the Reynolds number for an object moving through a fluid
- Use the Reynolds number for a system to determine whether it is laminar or turbulent
- Describe the conditions under which an object has a terminal speed

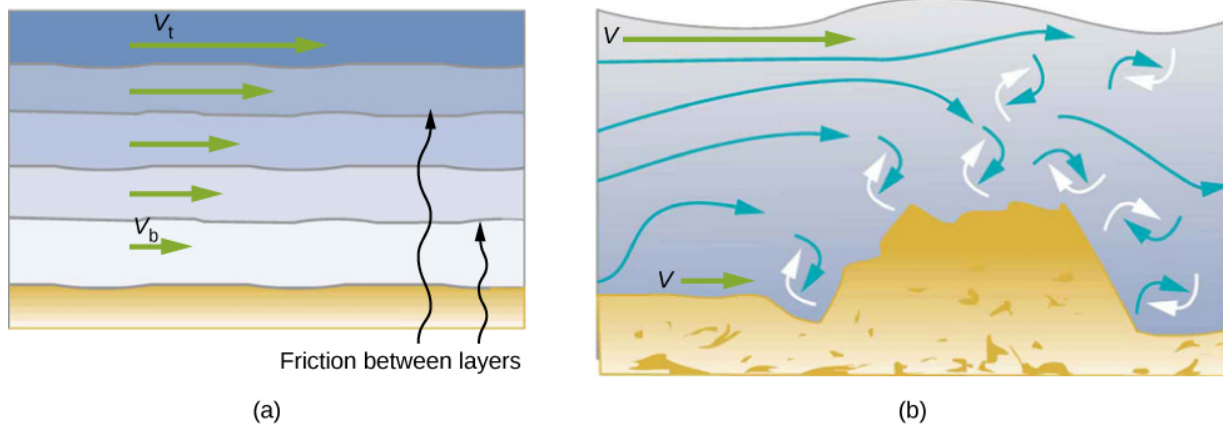
In [Applications of Newton's Laws](#), which introduced the concept of friction, we saw that an object sliding across the floor with an initial velocity and no applied force comes to rest due to the force of friction. Friction depends on the types of materials in contact and is proportional to the normal force. We also discussed drag and air resistance in that same chapter. We explained that at low speeds, the drag is proportional to the velocity, whereas at high speeds, drag is proportional to the velocity squared. In this section, we introduce the forces of friction that act on fluids in motion. For example, a fluid flowing through a pipe is subject to resistance, a type of friction, between the fluid and the walls. Friction also occurs between the different layers of fluid. These resistive forces affect the way the fluid flows through the pipe.

Viscosity and Laminar Flow

When you pour yourself a glass of juice, the liquid flows freely and quickly. But if you pour maple syrup on your pancakes, that liquid flows slowly and sticks to the pitcher. The difference is fluid friction, both within the fluid itself and between the fluid and its surroundings. We call this property of fluids *viscosity*. Juice has low viscosity, whereas syrup has high viscosity.

The precise definition of viscosity is based on laminar, or nonturbulent, flow. [\[link\]](#) shows schematically how laminar and turbulent flow differ. When flow is laminar, layers flow without mixing. When flow is turbulent,

the layers mix, and significant velocities occur in directions other than the overall direction of flow.



- (a) Laminar flow occurs in layers without mixing. Notice that viscosity causes drag between layers as well as with the fixed surface. The speed near the bottom of the flow (v_b) is less than speed near the top (v_t) because in this case, the surface of the containing vessel is at the bottom. (b) An obstruction in the vessel causes turbulent flow. Turbulent flow mixes the fluid. There is more interaction, greater heating, and more resistance than in laminar flow.

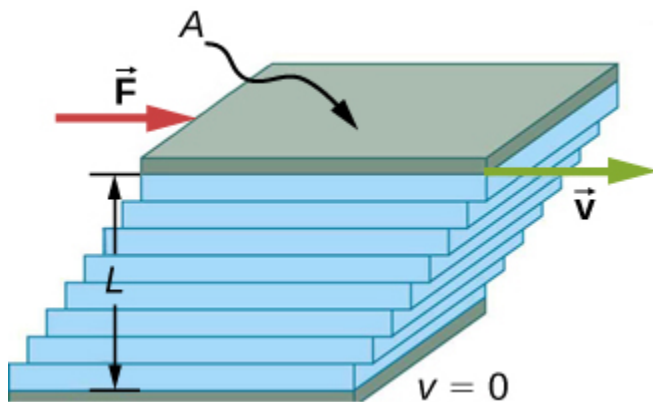
Turbulence is a fluid flow in which layers mix together via eddies and swirls. It has two main causes. First, any obstruction or sharp corner, such as in a faucet, creates turbulence by imparting velocities perpendicular to the flow. Second, high speeds cause turbulence. The drag between adjacent layers of fluid and between the fluid and its surroundings can form swirls and eddies if the speed is great enough. In [\[link\]](#), the speed of the accelerating smoke reaches the point that it begins to swirl due to the drag between the smoke and the surrounding air.



Smoke rises smoothly for a while and then begins to form swirls and eddies. The smooth flow is called laminar flow, whereas the swirls and eddies typify turbulent flow. Smoke rises more rapidly when flowing smoothly than after it becomes turbulent, suggesting that turbulence poses more resistance to flow. (credit: “Creativity103”/Flickr)

[\[link\]](#) shows how viscosity is measured for a fluid. The fluid to be measured is placed between two parallel plates. The bottom plate is held fixed, while

the top plate is moved to the right, dragging fluid with it. The layer (or lamina) of fluid in contact with either plate does not move relative to the plate, so the top layer moves at speed v while the bottom layer remains at rest. Each successive layer from the top down exerts a force on the one below it, trying to drag it along, producing a continuous variation in speed from v to 0 as shown. Care is taken to ensure that the flow is laminar, that is, the layers do not mix. The motion in the figure is like a continuous shearing motion. Fluids have zero shear strength, but the rate at which they are sheared is related to the same geometrical factors A and L as is shear deformation for solids. In the diagram, the fluid is initially at rest. The layer of fluid in contact with the moving plate is accelerated and starts to move due to the internal friction between moving plate and the fluid. The next layer is in contact with the moving layer; since there is internal friction between the two layers, it also accelerates, and so on through the depth of the fluid. There is also internal friction between the stationary plate and the lowest layer of fluid, next to the stationary plate. The force is required to keep the plate moving at a constant velocity due to the internal friction.



Measurement of viscosity for laminar flow of fluid between two plates of area A . The bottom plate is fixed. When the top plate is pushed to the right, it drags the fluid along with it.

A force F is required to keep the top plate in [\[link\]](#) moving at a constant velocity v , and experiments have shown that this force depends on four factors. First, F is directly proportional to v (until the speed is so high that turbulence occurs—then a much larger force is needed, and it has a more complicated dependence on v). Second, F is proportional to the area A of the plate. This relationship seems reasonable, since A is directly proportional to the amount of fluid being moved. Third, F is inversely proportional to the distance between the plates L . This relationship is also reasonable; L is like a lever arm, and the greater the lever arm, the less the force that is needed. Fourth, F is directly proportional to the coefficient of viscosity, η . The greater the viscosity, the greater the force required. These dependencies are combined into the equation

Equation:

$$F = \eta \frac{vA}{L}.$$

This equation gives us a working definition of fluid viscosity η . Solving for η gives

Note:

Equation:

$$\eta = \frac{FL}{vA}$$

which defines viscosity in terms of how it is measured.

The SI unit of viscosity is $\text{N} \cdot \text{m} / [(\text{m}/\text{s})\text{m}^2] = (\text{N}/\text{m}^2)\text{s}$ or $\text{Pa} \cdot \text{s}$.

[\[link\]](#) lists the coefficients of viscosity for various fluids. Viscosity varies from one fluid to another by several orders of magnitude. As you might

expect, the viscosities of gases are much less than those of liquids, and these viscosities often depend on temperature.

Fluid	Temperature (° C)	Viscosity $\eta \times 10^3$
Air	0	0.0171
	20	0.0181
	40	0.0190
	100	0.0218
Ammonia	20	0.00974
Carbon dioxide	20	0.0147
Helium	20	0.0196
Hydrogen	0	0.0090
Mercury	20	0.0450
Oxygen	20	0.0203
Steam	100	0.0130
Liquid water	0	1.792
	20	1.002

Fluid	Temperature (°C)	Viscosity $\eta \times 10^3$
	37	0.6947
	40	0.653
	100	0.282
Whole blood	20	3.015
	37	2.084
Blood plasma	20	1.810
	37	1.257
Ethyl alcohol	20	1.20
Methanol	20	0.584
Oil (heavy machine)	20	660
Oil (motor, SAE 10)	30	200
Oil (olive)	20	138
Glycerin	20	1500
Honey	20	2000–10000
Maple syrup	20	2000–3000
Milk	20	3.0
Oil (corn)	20	65

Coefficients of Viscosity of Various Fluids

Laminar Flow Confined to Tubes: Poiseuille's Law

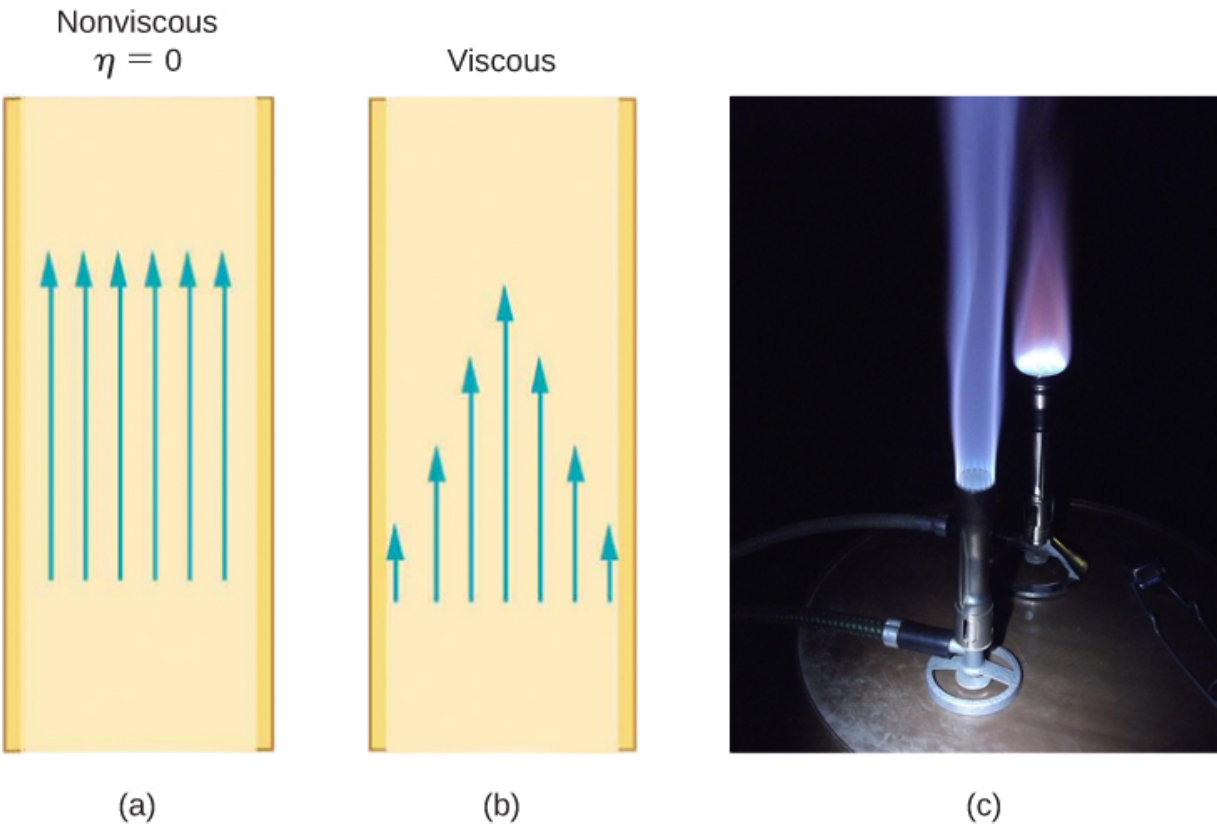
What causes flow? The answer, not surprisingly, is a pressure difference. In fact, there is a very simple relationship between horizontal flow and pressure. Flow rate Q is in the direction from high to low pressure. The greater the pressure differential between two points, the greater the flow rate. This relationship can be stated as

Equation:

$$Q = \frac{p_2 - p_1}{R}$$

where p_1 and p_2 are the pressures at two points, such as at either end of a tube, and R is the resistance to flow. The resistance R includes everything, except pressure, that affects flow rate. For example, R is greater for a long tube than for a short one. The greater the viscosity of a fluid, the greater the value of R . Turbulence greatly increases R , whereas increasing the diameter of a tube decreases R .

If viscosity is zero, the fluid is frictionless and the resistance to flow is also zero. Comparing frictionless flow in a tube to viscous flow, as in [\[link\]](#), we see that for a viscous fluid, speed is greatest at midstream because of drag at the boundaries. We can see the effect of viscosity in a Bunsen burner flame [part (c)], even though the viscosity of natural gas is small.



(a) If fluid flow in a tube has negligible resistance, the speed is the same all across the tube. (b) When a viscous fluid flows through a tube, its speed at the walls is zero, increasing steadily to its maximum at the center of the tube. (c) The shape of a Bunsen burner flame is due to the velocity profile across the tube. (credit c: modification of work by "jasonwoodhead23"/Flickr)

The resistance R to laminar flow of an incompressible fluid with viscosity η through a horizontal tube of uniform radius r and length l , is given by

Note:
Equation:

$$R = \frac{8\eta l}{\pi r^4}.$$

This equation is called **Poiseuille's law for resistance**, named after the French scientist J. L. Poiseuille (1799–1869), who derived it in an attempt to understand the flow of blood through the body.

Let us examine Poiseuille's expression for R to see if it makes good intuitive sense. We see that resistance is directly proportional to both fluid viscosity η and the length l of a tube. After all, both of these directly affect the amount of friction encountered—the greater either is, the greater the resistance and the smaller the flow. The radius r of a tube affects the resistance, which again makes sense, because the greater the radius, the greater the flow (all other factors remaining the same). But it is surprising that r is raised to the fourth power in Poiseuille's law. This exponent means that any change in the radius of a tube has a very large effect on resistance. For example, doubling the radius of a tube decreases resistance by a factor of $2^4 = 16$.

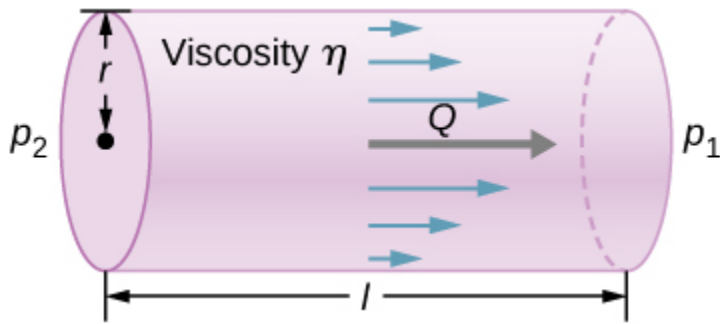
Taken together, $Q = \frac{p_2 - p_1}{R}$ and $R = \frac{8\eta l}{\pi r^4}$ give the following expression for flow rate:

Note:

Equation:

$$Q = \frac{(p_2 - p_1)\pi r^4}{8\eta l}.$$

This equation describes laminar flow through a tube. It is sometimes called Poiseuille's law for laminar flow, or simply **Poiseuille's law** ([\[link\]](#)).



Poiseuille's law applies to laminar flow of an incompressible fluid of viscosity η through a tube of length l and radius r . The direction of flow is from greater to lower pressure. Flow rate Q is directly proportional to the pressure difference $p_2 - p_1$, and inversely proportional to the length l of the tube and viscosity η of the fluid. Flow rate increases with radius by a factor of r^4 .

Example:

Using Flow Rate: Air Conditioning Systems

An air conditioning system is being designed to supply air at a gauge pressure of 0.054 Pa at a temperature of 20 °C. The air is sent through an insulated, round conduit with a diameter of 18.00 cm. The conduit is 20-meters long and is open to a room at atmospheric pressure 101.30 kPa. The room has a length of 12 meters, a width of 6 meters, and a height of 3 meters. (a) What is the volume flow rate through the pipe, assuming laminar flow? (b) Estimate the length of time to completely replace the air in the room. (c) The builders decide to save money by using a conduit with a diameter of 9.00 cm. What is the new flow rate?

Strategy

Assuming laminar flow, Poiseuille's law states that

Equation:

$$Q = \frac{(p_2 - p_1)\pi r^4}{8\eta l} = \frac{dV}{dt}.$$

We need to compare the artery radius before and after the flow rate reduction. Note that we are given the diameter of the conduit, so we must divide by two to get the radius.

Solution

- a. Assuming a constant pressure difference and using the viscosity $\eta = 0.0181 \text{ mPa} \cdot \text{s}$,

Equation:

$$Q = \frac{(0.054 \text{ Pa}) (3.14)(0.09 \text{ m})^4}{8 (0.0181 \times 10^{-3} \text{ Pa} \cdot \text{s}) (20 \text{ m})} = 3.84 \times 10^{-3} \frac{\text{m}^3}{\text{s}}.$$

- b. Assuming constant flow $Q = \frac{dV}{dt} \approx \frac{\Delta V}{\Delta t}$

Equation:

$$\Delta t = \frac{\Delta V}{Q} = \frac{(12 \text{ m}) (6 \text{ m}) (3 \text{ m})}{3.84 \times 10^{-3} \frac{\text{m}^3}{\text{s}}} = 5.63 \times 10^4 \text{ s} = 15.63 \text{ hr}.$$

- c. Using laminar flow, Poiseuille's law yields

Equation:

$$Q = \frac{(0.054 \text{ Pa}) (3.14)(0.045 \text{ m})^4}{8 (0.0181 \times 10^{-3} \text{ Pa} \cdot \text{s}) (20 \text{ m})} = 2.40 \times 10^{-4} \frac{\text{m}^3}{\text{s}}.$$

Thus, the radius of the conduit decreases by half reduces the flow rate to 6.25% of the original value.

Significance

In general, assuming laminar flow, decreasing the radius has a more dramatic effect than changing the length. If the length is increased and all other variables remain constant, the flow rate is decreased:

Equation:

$$\frac{Q_A}{Q_B} = \frac{\frac{(p_2-p_1)\pi r_A^4}{8\eta l_A}}{\frac{(p_2-p_1)\pi r_B^4}{8\eta l_B}} = \frac{l_B}{l_A}$$

$$Q_B = \frac{l_A}{l_B} Q_A.$$

Doubling the length cuts the flow rate to one-half the original flow rate. If the radius is decreased and all other variables remain constant, the volume flow rate decreases by a much larger factor.

Equation:

$$\frac{Q_A}{Q_B} = \frac{\frac{(p_2-p_1)\pi r_A^4}{8\eta l_A}}{\frac{(p_2-p_1)\pi r_B^4}{8\eta l_B}} = \left(\frac{r_A}{r_B}\right)^4$$

$$Q_B = \left(\frac{r_B}{r_A}\right)^4 Q_A$$

Cutting the radius in half decreases the flow rate to one-sixteenth the original flow rate.

Flow and Resistance as Causes of Pressure Drops

Water pressure in homes is sometimes lower than normal during times of heavy use, such as hot summer days. The drop in pressure occurs in the water main before it reaches individual homes. Let us consider flow through the water main as illustrated in [\[link\]](#). We can understand why the pressure p_1 to the home drops during times of heavy use by rearranging the equation for flow rate:

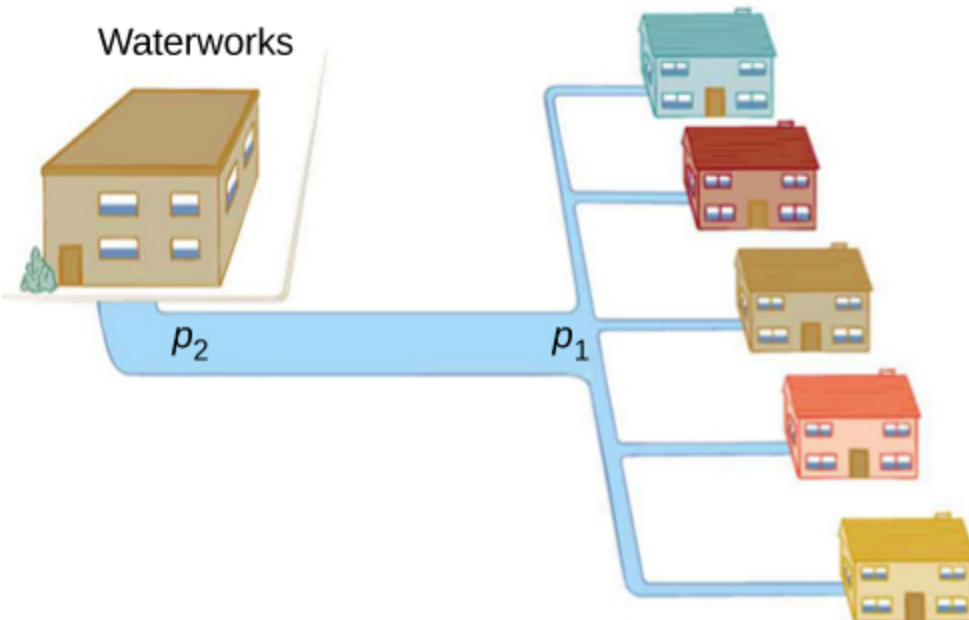
Equation:

$$Q = \frac{p_2-p_1}{R}$$

$$p_2-p_1 = RQ.$$

In this case, p_2 is the pressure at the water works and R is the resistance of the water main. During times of heavy use, the flow rate Q is large. This

means that $p_2 - p_1$ must also be large. Thus p_1 must decrease. It is correct to think of flow and resistance as causing the pressure to drop from p_2 to p_1 . The equation $p_2 - p_1 = RQ$ is valid for both laminar and turbulent flows.



During times of heavy use, there is a significant pressure drop in a water main, and p_1 supplied to users is significantly less than p_2 created at the water works.

If the flow is very small, then the pressure drop is negligible, and $p_2 \approx p_1$.

We can also use $p_2 - p_1 = RQ$ to analyze pressure drops occurring in more complex systems in which the tube radius is not the same everywhere. Resistance is much greater in narrow places, such as in an obstructed coronary artery. For a given flow rate Q , the pressure drop is greatest where the tube is most narrow. This is how water faucets control flow.

Additionally, R is greatly increased by turbulence, and a constriction that creates turbulence greatly reduces the pressure downstream. Plaque in an artery reduces pressure and hence flow, both by its resistance and by the turbulence it creates.

Measuring Turbulence

An indicator called the **Reynolds number** N_R can reveal whether flow is laminar or turbulent. For flow in a tube of uniform diameter, the Reynolds number is defined as

Equation:

$$N_R = \frac{2\rho v r}{\eta} \text{ (flow in tube)}$$

where ρ is the fluid density, v its speed, η its viscosity, and r the tube radius. The Reynolds number is a dimensionless quantity. Experiments have revealed that N_R is related to the onset of turbulence. For N_R below about 2000, flow is laminar. For N_R above about 3000, flow is turbulent.

For values of N_R between about 2000 and 3000, flow is unstable—that is, it can be laminar, but small obstructions and surface roughness can make it turbulent, and it may oscillate randomly between being laminar and turbulent. In fact, the flow of a fluid with a Reynolds number between 2000 and 3000 is a good example of chaotic behavior. A system is defined to be chaotic when its behavior is so sensitive to some factor that it is extremely difficult to predict. It is difficult, but not impossible, to predict whether flow is turbulent or not when a fluid's Reynolds number falls in this range due to extremely sensitive dependence on factors like roughness and obstructions on the nature of the flow. A tiny variation in one factor has an exaggerated (or nonlinear) effect on the flow.

Example:

Using Flow Rate: Turbulent Flow or Laminar Flow

In [\[link\]](#), we found the volume flow rate of an air conditioning system to be $Q = 3.84 \times 10^{-3} \text{ m}^3/\text{s}$. This calculation assumed laminar flow. (a) Was this a good assumption? (b) At what velocity would the flow become turbulent?

Strategy

To determine if the flow of air through the air conditioning system is laminar, we first need to find the velocity, which can be found by

Equation:

$$Q = Av = \pi r^2 v.$$

Then we can calculate the Reynold's number, using the equation below, and determine if it falls in the range for laminar flow

Equation:

$$R = \frac{2\rho v r}{\eta}.$$

Solution

a. Using the values given:

Equation:

$$\begin{aligned} v &= \frac{Q}{\pi r^2} = \frac{3.84 \times 10^{-3} \frac{\text{m}^3}{\text{s}}}{3.14(0.09 \text{ m})^2} = 0.15 \frac{\text{m}}{\text{s}} \\ R &= \frac{2\rho v r}{\eta} = \frac{2\left(1.23 \frac{\text{kg}}{\text{m}^3}\right)(0.15 \frac{\text{m}}{\text{s}})(0.09 \text{ m})}{0.0181 \times 10^{-3} \text{ Pa}\cdot\text{s}} = 1835. \end{aligned}$$

Since the Reynolds number is $1835 < 2000$, the flow is laminar and not turbulent. The assumption that the flow was laminar is valid.

b. To find the maximum speed of the air to keep the flow laminar, consider the Reynold's number.

Equation:

$$\begin{aligned} R &= \frac{2\rho v r}{\eta} \leq 2000 \\ v &= \frac{2000(0.0181 \times 10^{-3} \text{ Pa}\cdot\text{s})}{2\left(1.23 \frac{\text{kg}}{\text{m}^3}\right)(0.09 \text{ m})} = 0.16 \frac{\text{m}}{\text{s}}. \end{aligned}$$

Significance

When transferring a fluid from one point to another, it is desirable to limit turbulence. Turbulence results in wasted energy, as some of the energy

intended to move the fluid is dissipated when eddies are formed. In this case, the air conditioning system will become less efficient once the velocity exceeds 0.16 m/s, since this is the point at which turbulence will begin to occur.

Summary

- Laminar flow is characterized by smooth flow of the fluid in layers that do not mix.
- Turbulence is characterized by eddies and swirls that mix layers of fluid together.
- Fluid viscosity η is due to friction within a fluid.
- Flow is proportional to pressure difference and inversely proportional to resistance:

Equation:

$$Q = \frac{p - 2p_1}{R}.$$

- The pressure drop caused by flow and resistance is given by $p_2 - p_1 = RQ$.
- The Reynolds number N_R can reveal whether flow is laminar or turbulent. It is $N_R = \frac{2\rho vr}{\eta}$.
- For N_R below about 2000, flow is laminar. For N_R above about 3000, flow is turbulent. For values of N_R between 2000 and 3000, it may be either or both.

Key Equations

Density of a sample at constant

$$\rho = \frac{m}{V}$$

density	
Pressure	$p = \frac{F}{A}$
Pressure at a depth h in a fluid of constant density	$p = p_0 + \rho gh$
Change of pressure with height in a constant-density fluid	$\frac{dp}{dy} = -\rho g$
Absolute pressure	$p_{\text{abs}} = p_{\text{g}} + p_{\text{atm}}$
Pascal's principle	$\frac{F_1}{A_1} = \frac{F_2}{A_2}$
Volume flow rate	$Q = \frac{dV}{dt}$
Continuity equation (constant density)	$A_1 v_1 = A_2 v_2$
Continuity equation (general form)	$\rho_1 A_1 v_1 = \rho_2 A_2 v_2$
Bernoulli's equation	$p + \frac{1}{2} \rho v^2 + \rho gy = \text{constant}$
Viscosity	$\eta = \frac{FL}{vA}$
Poiseuille's law for resistance	$R = \frac{8\eta l}{\pi r^4}$
Poiseuille's law	$Q = \frac{(p_2 - p_1) \pi r^4}{8\eta l}$

Conceptual Questions

Exercise:

Problem:

Explain why the viscosity of a liquid decreases with temperature, that is, how might an increase in temperature reduce the effects of cohesive forces in a liquid? Also explain why the viscosity of a gas increases with temperature, that is, how does increased gas temperature create more collisions between atoms and molecules?

Exercise:**Problem:**

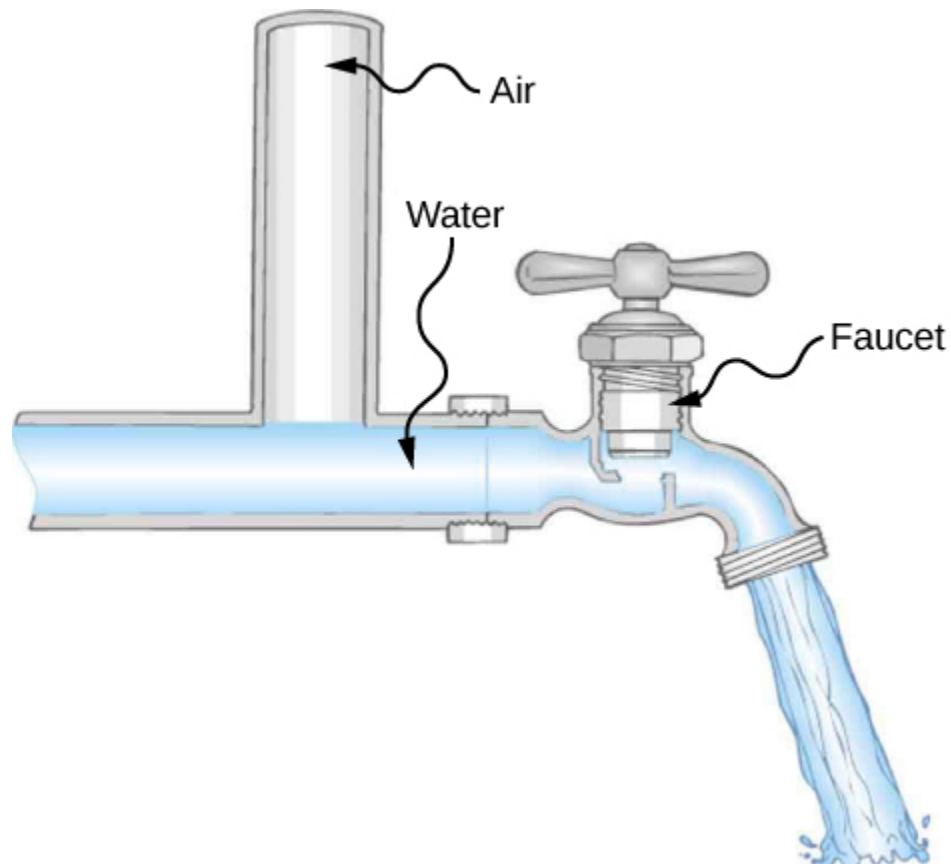
When paddling a canoe upstream, it is wisest to travel as near to the shore as possible. When canoeing downstream, it is generally better to stay near the middle. Explain why.

Solution:

The water in the center of the stream is moving faster than the water near the shore due to resistance between the water and the shore and between the layers of fluid. There is also probably more turbulence near the shore, which will also slow the water down. When paddling up stream, the water pushes against the canoe, so it is better to stay near the shore to minimize the force pushing against the canoe. When moving downstream, the water pushes the canoe, increasing its velocity, so it is better to stay in the middle of the stream to maximize this effect.

Exercise:**Problem:**

Plumbing usually includes air-filled tubes near water faucets (see the following figure). Explain why they are needed and how they work.



Exercise:

Problem:

Doppler ultrasound can be used to measure the speed of blood in the body. If there is a partial constriction of an artery, where would you expect blood speed to be greatest: at or after the constriction? What are the two distinct causes of higher resistance in the constriction?

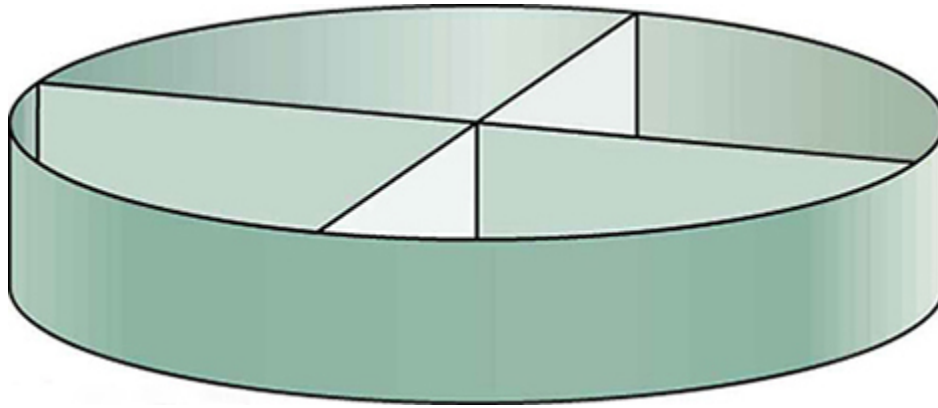
Solution:

You would expect the speed to be slower after the obstruction. Resistance is increased due to the reduction in size of the opening, and turbulence will be created because of the obstruction, both of which will cause the fluid to slow down.

Exercise:

Problem:

Sink drains often have a device such as that shown below to help speed the flow of water. How does this work?

**Problems****Exercise:****Problem:**

(a) Calculate the retarding force due to the viscosity of the air layer between a cart and a level air track given the following information: air temperature is 20°C , the cart is moving at 0.400 m/s , its surface area is $2.50 \times 10^{-2}\text{ m}^2$, and the thickness of the air layer is $6.00 \times 10^{-5}\text{ m}$. (b) What is the ratio of this force to the weight of the 0.300-kg cart?

Solution:

a. $3.02 \times 10^{-3}\text{ N}$; b. 1.03×10^{-3}

Exercise:

Problem:

The arterioles (small arteries) leading to an organ constrict in order to decrease flow to the organ. To shut down an organ, blood flow is reduced naturally to 1.00% of its original value. By what factor do the radii of the arterioles constrict?

Exercise:**Problem:**

A spherical particle falling at a terminal speed in a liquid must have the gravitational force balanced by the drag force and the buoyant force. The buoyant force is equal to the weight of the displaced fluid, while the drag force is assumed to be given by Stokes Law, $F_s = 6\pi r\eta v$. Show that the terminal speed is given by $v = \frac{2R^2g}{9\eta}(\rho_s - \rho_1)$, where R is the radius of the sphere, ρ_s is its density, and ρ_1 is the density of the fluid, and η the coefficient of viscosity.

Solution:

proof

Exercise:**Problem:**

Using the equation of the previous problem, find the viscosity of motor oil in which a steel ball of radius 0.8 mm falls with a terminal speed of 4.32 cm/s. The densities of the ball and the oil are 7.86 and 0.88 g/mL, respectively.

Exercise:

Problem:

A skydiver will reach a terminal velocity when the air drag equals his or her weight. For a skydiver with a large body, turbulence is a factor at high speeds. The drag force then is approximately proportional to the square of the velocity. Taking the drag force to be $F_D = \frac{1}{2}\rho A v^2$, and setting this equal to the skydiver's weight, find the terminal speed for a person falling "spread eagle."

Solution:

40 m/s

Exercise:**Problem:**

(a) Verify that a 19.0% decrease in laminar flow through a tube is caused by a 5.00% decrease in radius, assuming that all other factors remain constant. (b) What increase in flow is obtained from a 5.00% increase in radius, again assuming all other factors remain constant?

Exercise:**Problem:**

When physicians diagnose arterial blockages, they quote the reduction in flow rate. If the flow rate in an artery has been reduced to 10.0% of its normal value by a blood clot and the average pressure difference has increased by 20.0%, by what factor has the clot reduced the radius of the artery?

Solution:

$0.537r$; The radius is reduced to 53.7% of its normal value.

Exercise:

Problem:

An oil gusher shoots crude oil 25.0 m into the air through a pipe with a 0.100-m diameter. Neglecting air resistance but not the resistance of the pipe, and assuming laminar flow, calculate the pressure at the entrance of the 50.0-m-long vertical pipe. Take the density of the oil to be 900 kg/m^3 and its viscosity to be $1.00(\text{N/m}^2) \cdot \text{s}$ (or $1.00 \text{ Pa} \cdot \text{s}$). Note that you must take into account the pressure due to the 50.0-m column of oil in the pipe.

Exercise:**Problem:**

Concrete is pumped from a cement mixer to the place it is being laid, instead of being carried in wheelbarrows. The flow rate is 200 L/min through a 50.0-m-long, 8.00-cm-diameter hose, and the pressure at the pump is $8.00 \times 10^6 \text{ N/m}^2$. (a) Calculate the resistance of the hose. (b) What is the viscosity of the concrete, assuming the flow is laminar? (c) How much power is being supplied, assuming the point of use is at the same level as the pump? You may neglect the power supplied to increase the concrete's velocity.

Solution:

a. $2.40 \times 10^9 \text{ N} \cdot \text{s/m}^5$; b. $48.3 (\text{N/m}^2) \cdot \text{s}$; c. $2.67 \times 10^4 \text{ W}$

Exercise:**Problem:**

Verify that the flow of oil is laminar for an oil gusher that shoots crude oil 25.0 m into the air through a pipe with a 0.100-m diameter. The vertical pipe is 50 m long. Take the density of the oil to be 900 kg/m^3 and its viscosity to be $1.00(\text{N/m}^2) \cdot \text{s}$ (or $1.00 \text{ Pa} \cdot \text{s}$).

Exercise:

Problem:

Calculate the Reynolds numbers for the flow of water through (a) a nozzle with a radius of 0.250 cm and (b) a garden hose with a radius of 0.900 cm, when the nozzle is attached to the hose. The flow rate through hose and nozzle is 0.500 L/s. Can the flow in either possibly be laminar?

Solution:

- a. Nozzle: $v = 25.5 \frac{\text{m}}{\text{s}}$
 $N_R = 1.27 \times 10^5 > 2000 \Rightarrow$
Flow is not laminar.
- b. Hose: $v = 1.96 \frac{\text{m}}{\text{s}}$
 $N_R = 35,100 > 2000 \Rightarrow$
Flow is not laminar.

Exercise:**Problem:**

A fire hose has an inside diameter of 6.40 cm. Suppose such a hose carries a flow of 40.0 L/s starting at a gauge pressure of $1.62 \times 10^6 \text{ N/m}^2$. The hose goes 10.0 m up a ladder to a nozzle having an inside diameter of 3.00 cm. Calculate the Reynolds numbers for flow in the fire hose and nozzle to show that the flow in each must be turbulent.

Exercise:**Problem:**

At what flow rate might turbulence begin to develop in a water main with a 0.200-m diameter? Assume a 20 °C temperature.

Solution:

$$3.16 \times 10^{-4} \text{ m}^3/\text{s}$$

Additional Problems

Exercise:

Problem:

Before digital storage devices, such as the memory in your cell phone, music was stored on vinyl disks with grooves with varying depths cut into the disk. A phonograph used a needle, which moved over the grooves, measuring the depth of the grooves. The pressure exerted by a phonograph needle on a record is surprisingly large. If the equivalent of 1.00 g is supported by a needle, the tip of which is a circle with a 0.200-mm radius, what pressure is exerted on the record in Pa?

Exercise:

Problem:

Water towers store water above the level of consumers for times of heavy use, eliminating the need for high-speed pumps. How high above a user must the water level be to create a gauge pressure of $3.00 \times 10^5 \text{ N/m}^2$?

Solution:

30.6 m

Exercise:

Problem:

The aqueous humor in a person's eye is exerting a force of 0.300 N on the 1.10-cm^2 area of the cornea. What pressure is this in mm Hg?

Exercise:

Problem:

(a) Convert normal blood pressure readings of 120 over 80 mm Hg to newtons per meter squared using the relationship for pressure due to the weight of a fluid ($p = h\rho g$) rather than a conversion factor. (b) Explain why the blood pressure of an infant would likely be smaller than that of an adult. Specifically, consider the smaller height to which blood must be pumped.

Solution:

a. $p_{120} = 1.60 \times 10^4 \text{ N/m}^2$;

$$p_{80} = 1.07 \times 10^4 \text{ N/m}^2$$

b. Since an infant is only approximately 20 inches tall, while an adult is approximately 70 inches tall, the blood pressure for an infant would be expected to be smaller than that of an adult. The blood only feels a pressure of 20 inches rather than 70 inches, so the pressure should be smaller.

Exercise:**Problem:**

Pressure cookers have been around for more than 300 years, although their use has greatly declined in recent years (early models had a nasty habit of exploding). How much force must the latches holding the lid onto a pressure cooker be able to withstand if the circular lid is 25.0 cm in diameter and the gauge pressure inside is 300 atm? Neglect the weight of the lid.

Exercise:

Problem:

Bird bones have air pockets in them to reduce their weight—this also gives them an average density significantly less than that of the bones of other animals. Suppose an ornithologist weighs a bird bone in air and in water and finds its mass is 45.0 g and its apparent mass when submerged is 3.60 g (assume the bone is watertight). (a) What mass of water is displaced? (b) What is the volume of the bone? (c) What is its average density?

Solution:

a. 41.4 g; b. 41.4 cm³; c. 1.09 g/cm³. This is clearly not the density of the bone everywhere. The air pockets will have a density of approximately $1.29 \times 10^{-3} \text{ g/cm}^3$, while the bone will be substantially denser.

Exercise:**Problem:**

In an immersion measurement of a woman's density, she is found to have a mass of 62.0 kg in air and an apparent mass of 0.0850 kg when completely submerged with lungs empty. (a) What mass of water does she displace? (b) What is her volume? (c) Calculate her density. (d) If her lung capacity is 1.75 L, is she able to float without treading water with her lungs filled with air?

Exercise:**Problem:**

Some fish have a density slightly less than that of water and must exert a force (swim) to stay submerged. What force must an 85.0-kg grouper exert to stay submerged in salt water if its body density is 1015 kg/m³?

Solution:

12.3 N

Exercise:

Problem:

The human circulation system has approximately 1×10^9 capillary vessels. Each vessel has a diameter of about $8\mu\text{m}$. Assuming cardiac output is 5 L/min, determine the average velocity of blood flow through each capillary vessel.

Exercise:

Problem:

The flow rate of blood through a 2.00×10^{-6} m-radius capillary is $3.80 \times 10^9 \text{ cm}^3/\text{s}$. (a) What is the speed of the blood flow? (b) Assuming all the blood in the body passes through capillaries, how many of them must there be to carry a total flow of $90.0 \text{ cm}^3/\text{s}$?

Solution:

a. $3.02 \times 10^{-2} \text{ cm/s}$. (This small speed allows time for diffusion of materials to and from the blood.) b. 2.37×10^{10} capillaries. (This large number is an overestimate, but it is still reasonable.)

Exercise:

Problem:

The left ventricle of a resting adult's heart pumps blood at a flow rate of $83.0 \text{ cm}^3/\text{s}$, increasing its pressure by 110 mm Hg, its speed from zero to 30.0 cm/s , and its height by 5.00 cm . (All numbers are averaged over the entire heartbeat.) Calculate the total power output of the left ventricle. Note that most of the power is used to increase blood pressure.

Exercise:

Problem:

A sump pump (used to drain water from the basement of houses built below the water table) is draining a flooded basement at the rate of 0.750 L/s, with an output pressure of $3.00 \times 10^5 \text{ N/m}^2$. (a) The water enters a hose with a 3.00-cm inside diameter and rises 2.50 m above the pump. What is its pressure at this point? (b) The hose goes over the foundation wall, losing 0.500 m in height, and widens to 4.00 cm in diameter. What is the pressure now? You may neglect frictional losses in both parts of the problem.

Solution:

a. $2.76 \times 10^5 \text{ N/m}^2$; b. $P_2 = 2.81 \times 10^5 \text{ N/m}^2$

Exercise:**Problem:**

A glucose solution being administered with an IV has a flow rate of $4.00 \text{ cm}^3/\text{min}$. What will the new flow rate be if the glucose is replaced by whole blood having the same density but a viscosity 2.50 times that of the glucose? All other factors remain constant.

Exercise:**Problem:**

A small artery has a length of $1.1 \times 10^{-3} \text{ m}$ and a radius of $2.5 \times 10^{-5} \text{ m}$. If the pressure drop across the artery is 1.3 kPa, what is the flow rate through the artery? (Assume that the temperature is 37°C .)

Solution:

$8.7 \times 10^{-2} \text{ mm}^3/\text{s}$

Exercise:

Problem:

Angioplasty is a technique in which arteries partially blocked with plaque are dilated to increase blood flow. By what factor must the radius of an artery be increased in order to increase blood flow by a factor of 10?

Exercise:**Problem:**

Suppose a blood vessel's radius is decreased to 90.0% of its original value by plaque deposits and the body compensates by increasing the pressure difference along the vessel to keep the flow rate constant. By what factor must the pressure difference increase? (b) If turbulence is created by the obstruction, what additional effect would it have on the flow rate?

Solution:

a. 1.52; b. Turbulence would decrease the flow rate of the blood, which would require an even larger increase in the pressure difference, leading to higher blood pressure.

Challenge Problems**Exercise:****Problem:**

The pressure on the dam shown early in the problems section increases with depth. Therefore, there is a net torque on the dam. Find the net torque.

Exercise:

Problem:

The temperature of the atmosphere is not always constant and can increase or decrease with height. In a neutral atmosphere, where there is not a significant amount of vertical mixing, the temperature decreases at a rate of approximately 6.5 K per km. The magnitude of the decrease in temperature as height increases is known as the lapse rate (Γ). (The symbol is the upper case Greek letter gamma.) Assume that the surface pressure is $p_0 = 1.013 \times 10^5 \text{ Pa}$ where $T = 293 \text{ K}$ and the lapse rate is $(-\Gamma = 6.5 \frac{\text{K}}{\text{km}})$. Estimate the pressure 3.0 km above the surface of Earth.

Solution:

$$p = 0.99 \times 10^5 \text{ Pa}$$

Exercise:**Problem:**

A submarine is stranded on the bottom of the ocean with its hatch 25.0 m below the surface. Calculate the force needed to open the hatch from the inside, given it is circular and 0.450 m in diameter. Air pressure inside the submarine is 1.00 atm.

Exercise:**Problem:**

Logs sometimes float vertically in a lake because one end has become water-logged and denser than the other. What is the average density of a uniform-diameter log that floats with 20.0% of its length above water?

Solution:

$$800 \text{ kg/m}^3$$

Exercise:

Problem:

Scurrilous con artists have been known to represent gold-plated tungsten ingots as pure gold and sell them at prices much below gold value but high above the cost of tungsten. With what accuracy must you be able to measure the mass of such an ingot in and out of water to tell that it is almost pure tungsten rather than pure gold?

Exercise:**Problem:**

The inside volume of a house is equivalent to that of a rectangular solid 13.0 m wide by 20.0 m long by 2.75 m high. The house is heated by a forced air gas heater. The main uptake air duct of the heater is 0.300 m in diameter. What is the average speed of air in the duct if it carries a volume equal to that of the house's interior every 15 minutes?

Solution:

11.2 m/s

Exercise:**Problem:**

A garden hose with a diameter of 2.0 cm is used to fill a bucket, which has a volume of 0.10 cubic meters. It takes 1.2 minutes to fill. An adjustable nozzle is attached to the hose to decrease the diameter of the opening, which increases the speed of the water. The hose is held level to the ground at a height of 1.0 meters and the diameter is decreased until a flower bed 3.0 meters away is reached. (a) What is the volume flow rate of the water through the nozzle when the diameter is 2.0 cm? (b) What is the speed of the water coming out of the hose? (c) What does the speed of the water coming out of the hose need to be to reach the flower bed 3.0 meters away? (d) What is the diameter of the nozzle needed to reach the flower bed?

Exercise:

Problem:

A frequently quoted rule of thumb in aircraft design is that wings should produce about 1000 N of lift per square meter of wing. (The fact that a wing has a top and bottom surface does not double its area.)

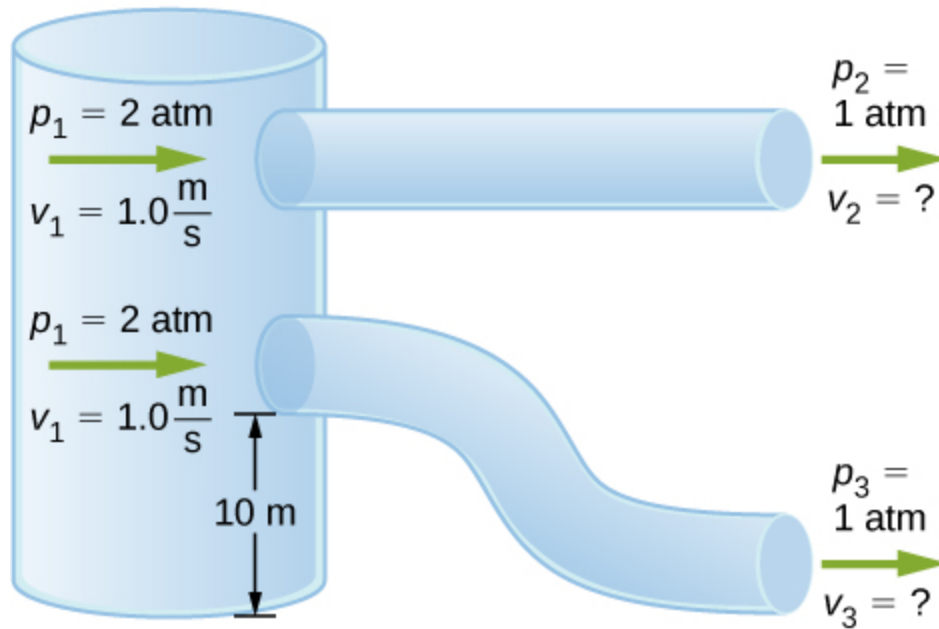
(a) At takeoff, an aircraft travels at 60.0 m/s, so that the air speed relative to the bottom of the wing is 60.0 m/s. Given the sea level density of air as 1.29 kg/m^3 , how fast must it move over the upper surface to create the ideal lift? (b) How fast must air move over the upper surface at a cruising speed of 245 m/s and at an altitude where air density is one-fourth that at sea level? (Note that this is not all of the aircraft's lift—some comes from the body of the plane, some from engine thrust, and so on. Furthermore, Bernoulli's principle gives an approximate answer because flow over the wing creates turbulence.)

Solution:

a. 71.8 m/s; b. 257 m/s

Exercise:**Problem:**

Two pipes of equal and constant diameter leave a water pumping station and dump water out of an open end that is open to the atmosphere (see the following figure). The water enters at a pressure of two atmospheres and a speed of ($v_1 = 1.0 \text{ m/s}$). One pipe drops a height of 10 m. What is the velocity of the water as the water leaves each pipe?



Exercise:

Problem:

Fluid originally flows through a tube at a rate of $100 \text{ cm}^3/\text{s}$. To illustrate the sensitivity of flow rate to various factors, calculate the new flow rate for the following changes with all other factors remaining the same as in the original conditions. (a) Pressure difference increases by a factor of 1.50. (b) A new fluid with 3.00 times greater viscosity is substituted. (c) The tube is replaced by one having 4.00 times the length. (d) Another tube is used with a radius 0.100 times the original. (e) Yet another tube is substituted with a radius 0.100 times the original and half the length, and the pressure difference is increased by a factor of 1.50.

Solution:

a. $150 \text{ cm}^3/\text{s}$; b. $33.3 \text{ cm}^3/\text{s}$; c. $25.0 \text{ cm}^3/\text{s}$; d. $0.0100 \text{ cm}^3/\text{s}$; e. $0.0300 \text{ cm}^3/\text{s}$

Exercise:

Problem:

During a marathon race, a runner's blood flow increases to 10.0 times her resting rate. Her blood's viscosity has dropped to 95.0% of its normal value, and the blood pressure difference across the circulatory system has increased by 50.0%. By what factor has the average radii of her blood vessels increased?

Exercise:**Problem:**

Water supplied to a house by a water main has a pressure of $3.00 \times 10^5 \text{ N/m}^2$ early on a summer day when neighborhood use is low. This pressure produces a flow of 20.0 L/min through a garden hose. Later in the day, pressure at the exit of the water main and entrance to the house drops, and a flow of only 8.00 L/min is obtained through the same hose. (a) What pressure is now being supplied to the house, assuming resistance is constant? (b) By what factor did the flow rate in the water main increase in order to cause this decrease in delivered pressure? The pressure at the entrance of the water main is $5.00 \times 10^5 \text{ N/m}^2$, and the original flow rate was 200 L/min. (c) How many more users are there, assuming each would consume 20.0 L/min in the morning?

Solution:

- a. $1.20 \times 10^5 \text{ N/m}^2$; b. The flow rate in the main increases by 90%.
- c. There are approximately 38 more users in the afternoon.

Exercise:

Problem:

Gasoline is piped underground from refineries to major users. The flow rate is $3.00 \times 10^{-2} \text{ m}^3/\text{s}$ (about 500 gal/min), the viscosity of gasoline is $1.00 \times 10^{-3} \text{ (N/m}^2) \cdot \text{s}$, and its density is 680 kg/m^3 .

(a) What minimum diameter must the pipe have if the Reynolds number is to be less than 2000? (b) What pressure difference must be maintained along each kilometer of the pipe to maintain this flow rate?

Glossary

Poiseuille's law

rate of laminar flow of an incompressible fluid in a tube:

$$Q = \frac{(p_2 - p_1)\pi r^4}{8\eta l}.$$

Poiseuille's law for resistance

resistance to laminar flow of an incompressible fluid in a tube:

$$R = \frac{8\eta l}{\pi r^4}$$

Reynolds number

dimensionless parameter that can reveal whether a particular flow is laminar or turbulent

turbulence

fluid flow in which layers mix together via eddies and swirls

Introduction

class="introduction"

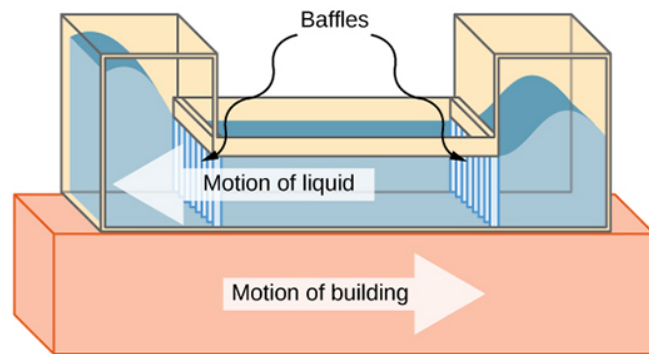
(a) The Comcast Building in Philadelphia, Pennsylvania, looming high above the skyline, is approximately 305 meters (1000 feet) tall. At this height, the top floors can oscillate back and forth due to seismic activity and fluctuating winds. (b)

Shown above is a schematic drawing of a tuned, liquid-column mass damper, installed at the top of the Comcast, consisting of a 300,000-gallon reservoir of water to

reduce
oscillations.



(a)



(b)

We begin the study of oscillations with simple systems of pendulums and springs. Although these systems may seem quite basic, the concepts involved have many real-life applications. For example, the Comcast Building in Philadelphia, Pennsylvania, stands approximately 305 meters (1000 feet) tall. As buildings are built taller, they can act as inverted, physical pendulums, with the top floors oscillating due to seismic activity and fluctuating winds. In the Comcast Building, a tuned-mass damper is used to reduce the oscillations. Installed at the top of the building is a tuned, liquid-column mass damper, consisting of a 300,000-gallon reservoir of water. This U-shaped tank allows the water to oscillate freely at a frequency that matches the natural frequency of the building. Damping is provided by tuning the turbulence levels in the moving water using baffles.

Simple Harmonic Motion

By the end of this section, you will be able to:

- Define the terms period and frequency
- List the characteristics of simple harmonic motion
- Explain the concept of phase shift
- Write the equations of motion for the system of a mass and spring undergoing simple harmonic motion
- Describe the motion of a mass oscillating on a vertical spring

When you pluck a guitar string, the resulting sound has a steady tone and lasts a long time ([link](#)). The string vibrates around an equilibrium position, and one oscillation is completed when the string starts from the initial position, travels to one of the extreme positions, then to the other extreme position, and returns to its initial position. We define **periodic motion** to be any motion that repeats itself at regular time intervals, such as exhibited by the guitar string or by a child swinging on a swing. In this section, we study the basic characteristics of oscillations and their mathematical description.



When a guitar string is plucked, the string oscillates up and down in periodic motion. The vibrating string causes the surrounding air molecules to oscillate, producing sound waves. (credit: Yutaka Tsutano)

Period and Frequency in Oscillations

In the absence of friction, the time to complete one oscillation remains constant and is called the **period (T)**. Its units are usually seconds, but may be any convenient unit of time. The word ‘period’ refers to the time for some event whether repetitive or not, but in this chapter, we shall deal primarily in periodic motion, which is by definition repetitive.

A concept closely related to period is the frequency of an event. **Frequency (f)** is defined to be the number of events per unit time. For periodic motion, frequency is the number of oscillations per unit time. The relationship between frequency and period is

Note:

Equation:

$$f = \frac{1}{T}.$$

The SI unit for frequency is the *hertz* (Hz) and is defined as one *cycle per second*:

Equation:

$$1 \text{ Hz} = 1 \frac{\text{cycle}}{\text{s}} \quad \text{or} \quad 1 \text{ Hz} = \frac{1}{\text{s}} = 1 \text{ s}^{-1}.$$

A cycle is one complete **oscillation**.

Example:

Determining the Frequency of Medical Ultrasound

Ultrasound machines are used by medical professionals to make images for examining internal organs of the body. An ultrasound machine emits high-frequency sound waves, which reflect off the organs, and a computer receives the waves, using them to create a picture. We can use the formulas presented in this module to determine the frequency, based on what we know about oscillations. Consider a medical imaging device that produces ultrasound by oscillating with a period of $0.400 \mu\text{s}$. What is the frequency of this oscillation?

Strategy

The period (T) is given and we are asked to find frequency (f).

Solution

Substitute $0.400\ \mu\text{s}$ for T in $f = \frac{1}{T}$:

Equation:

$$f = \frac{1}{T} = \frac{1}{0.400 \times 10^{-6}\ \text{s}}.$$

Solve to find

Equation:

$$f = 2.50 \times 10^6\ \text{Hz}.$$

Significance

This frequency of sound is much higher than the highest frequency that humans can hear (the range of human hearing is 20 Hz to 20,000 Hz); therefore, it is called ultrasound. Appropriate oscillations at this frequency generate ultrasound used for noninvasive medical diagnoses, such as observations of a fetus in the womb.

Characteristics of Simple Harmonic Motion

A very common type of periodic motion is called **simple harmonic motion (SHM)**. A system that oscillates with SHM is called a **simple harmonic oscillator**.

Note:

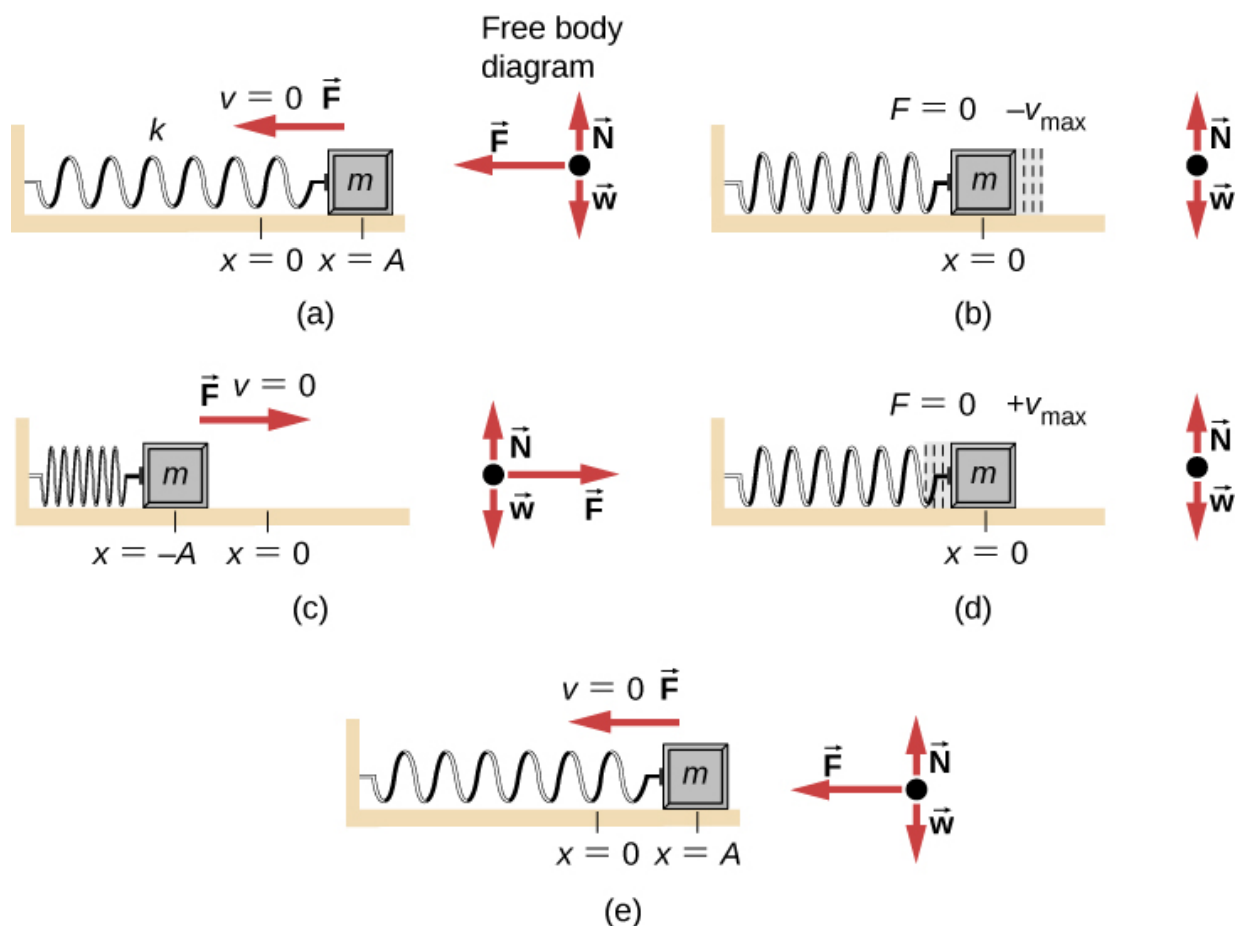
Simple Harmonic Motion

In simple harmonic motion, the acceleration of the system, and therefore the net force, is proportional to the displacement and acts in the opposite direction of the displacement.

A good example of SHM is an object with mass m attached to a spring on a frictionless surface, as shown in [\[link\]](#). The object oscillates around the equilibrium position, and the net force on the object is equal to the force provided by the spring. This force obeys Hooke's law $F_s = -kx$, as discussed in a previous chapter.

If the net force can be described by Hooke's law and there is no *damping* (slowing down due to friction or other nonconservative forces), then a simple harmonic oscillator oscillates with equal displacement on either side of the equilibrium position, as shown for an object on a spring in [\[link\]](#). The maximum displacement from

equilibrium is called the **amplitude (A)**. The units for amplitude and displacement are the same but depend on the type of oscillation. For the object on the spring, the units of amplitude and displacement are meters.



An object attached to a spring sliding on a frictionless surface is an uncomplicated simple harmonic oscillator. In the above set of figures, a mass is attached to a spring and placed on a frictionless table. The other end of the spring is attached to the wall. The position of the mass, when the spring is neither stretched nor compressed, is marked as $x = 0$ and is the equilibrium position. (a) The mass is displaced to a position $x = A$ and released from rest. (b) The mass accelerates as it moves in the negative x -direction, reaching a maximum negative velocity at $x = 0$. (c) The mass continues to move in the negative x -direction, slowing until it comes to a stop at $x = -A$. (d) The mass now begins to accelerate in the positive x -direction, reaching a positive maximum velocity at $x = 0$. (e) The mass then continues to move in the positive direction until it stops at $x = A$. The mass continues in SHM that has an amplitude A and a period T . The object's maximum speed occurs as it passes through equilibrium. The stiffer

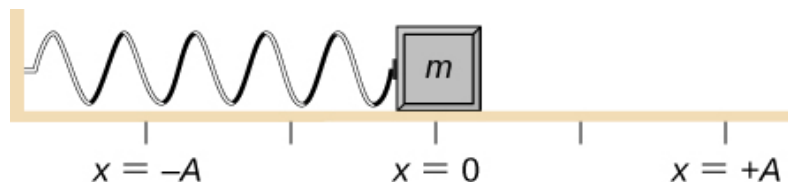
the spring is, the smaller the period T . The greater the mass of the object is, the greater the period T .

What is so significant about SHM? For one thing, the period T and frequency f of a simple harmonic oscillator are independent of amplitude. The string of a guitar, for example, oscillates with the same frequency whether plucked gently or hard.

Two important factors do affect the period of a simple harmonic oscillator. The period is related to how stiff the system is. A very stiff object has a large **force constant (k)**, which causes the system to have a smaller period. For example, you can adjust a diving board's stiffness—the stiffer it is, the faster it vibrates, and the shorter its period. Period also depends on the mass of the oscillating system. The more massive the system is, the longer the period. For example, a heavy person on a diving board bounces up and down more slowly than a light one. In fact, the mass m and the force constant k are the *only* factors that affect the period and frequency of SHM. To derive an equation for the period and the frequency, we must first define and analyze the equations of motion. Note that the force constant is sometimes referred to as the *spring constant*.

Equations of SHM

Consider a block attached to a spring on a frictionless table ([link](#)). The **equilibrium position** (the position where the spring is neither stretched nor compressed) is marked as $x = 0$. At the equilibrium position, the net force is zero.



A block is attached to a spring and placed on a frictionless table. The equilibrium position, where the spring is neither extended nor compressed, is marked as $x = 0$.

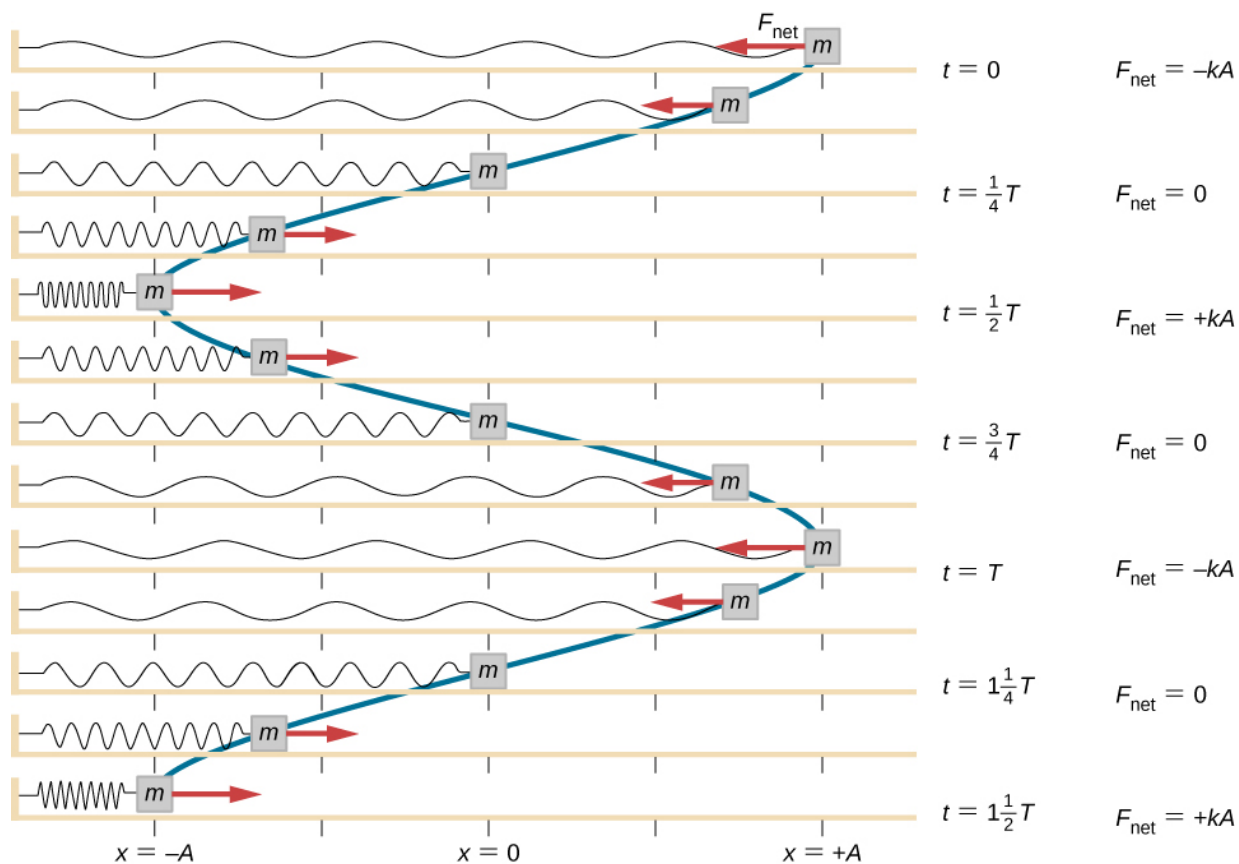
Work is done on the block to pull it out to a position of $x = +A$, and it is then released from rest. The maximum x -position (A) is called the amplitude of the motion. The block begins to oscillate in SHM between $x = +A$ and $x = -A$, where A is the amplitude of the motion and T is the period of the oscillation. The period is the time for one oscillation. [\[link\]](#) shows the motion of the block as it completes one and a half oscillations after release. [\[link\]](#) shows a plot of the position of the block versus time. When the position is plotted versus time, it is clear that the data can be modeled by a cosine function with an amplitude A and a period T . The cosine function $\cos\theta$ repeats every multiple of 2π , whereas the motion of the block repeats every period T . However, the function $\cos\left(\frac{2\pi}{T}t\right)$ repeats every integer multiple of the period. The maximum of the cosine function is one, so it is necessary to multiply the cosine function by the amplitude A .

Note:

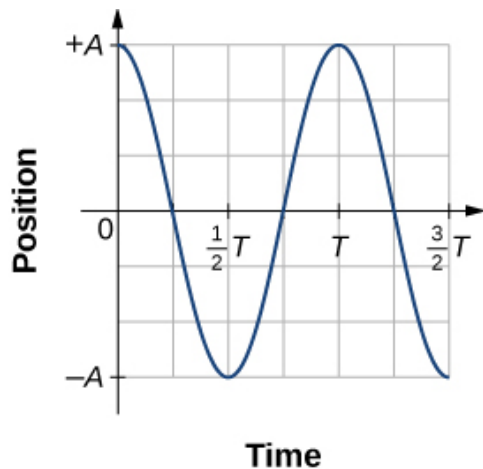
Equation:

$$x(t) = A \cos\left(\frac{2\pi}{T}t\right) = A \cos(\omega t).$$

Recall from the chapter on rotation that the angular frequency equals $\omega = \frac{d\theta}{dt}$. In this case, the period is constant, so the angular frequency is defined as 2π divided by the period, $\omega = \frac{2\pi}{T}$.

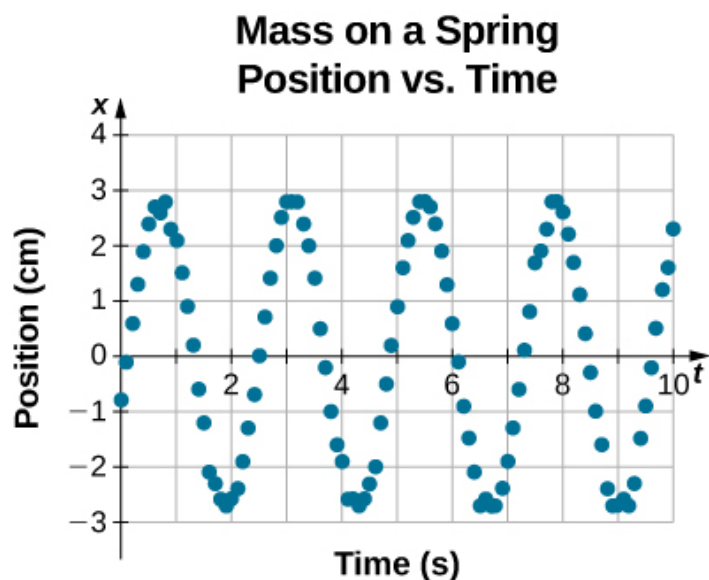


A block is attached to one end of a spring and placed on a frictionless table. The other end of the spring is anchored to the wall. The equilibrium position, where the net force equals zero, is marked as $x = 0$ m. Work is done on the block, pulling it out to $x = +A$, and the block is released from rest. The block oscillates between $x = +A$ and $x = -A$. The force is also shown as a vector.



A graph of the position of the block shown in [\[link\]](#) as a function of time. The position can be modeled as a periodic function, such as a cosine or sine function.

The equation for the position as a function of time $x(t) = A \cos(\omega t)$ is good for modeling data, where the position of the block at the initial time $t = 0.00$ s is at the amplitude A and the initial velocity is zero. Often when taking experimental data, the position of the mass at the initial time $t = 0.00$ s is not equal to the amplitude and the initial velocity is not zero. Consider 10 seconds of data collected by a student in lab, shown in [\[link\]](#).



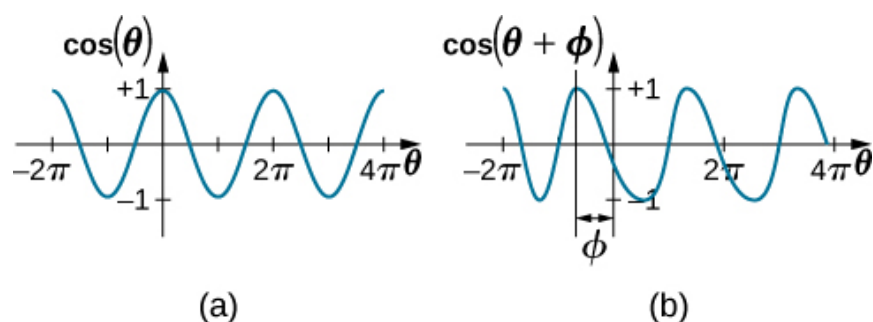
Data collected by a student in lab indicate the position of a block attached to a spring, measured with a sonic range finder. The data are collected starting at time $t = 0.00\text{s}$, but the initial position is near position $x \approx -0.80\text{ cm} \neq 3.00\text{ cm}$, so the initial position does not equal the amplitude $x_0 = +A$. The velocity is the time derivative of the position, which is the slope at a point on the graph of position versus time. The velocity is not $v = 0.00\text{ m/s}$ at time $t = 0.00\text{ s}$, as evident by the slope of the graph of position versus time, which is not zero at the initial time.

The data in [\[link\]](#) can still be modeled with a periodic function, like a cosine function, but the function is shifted to the right. This shift is known as a **phase shift** and is usually represented by the Greek letter phi (ϕ). The equation of the position as a function of time for a block on a spring becomes

Equation:

$$x(t) = A \cos(\omega t + \phi).$$

This is the generalized equation for SHM where t is the time measured in seconds, ω is the angular frequency with units of inverse seconds, A is the amplitude measured in meters or centimeters, and ϕ is the phase shift measured in radians ([link](#)). It should be noted that because sine and cosine functions differ only by a phase shift, this motion could be modeled using either the cosine or sine function.



(a) A cosine function. (b) A cosine function shifted to the left by an angle ϕ . The angle ϕ is known as the phase shift of the function.

The velocity of the mass on a spring, oscillating in SHM, can be found by taking the derivative of the position equation:

Equation:

$$v(t) = \frac{dx}{dt} = \frac{d}{dt}(A \cos(\omega t + \phi)) = -A\omega \sin(\omega t + \phi) = -v_{\max} \sin(\omega t + \phi).$$

Because the sine function oscillates between -1 and $+1$, the maximum velocity is the amplitude times the angular frequency, $v_{\max} = A\omega$. The maximum velocity occurs at the equilibrium position ($x = 0$) when the mass is moving toward $x = +A$. The maximum velocity in the negative direction is attained at the equilibrium position ($x = 0$) when the mass is moving toward $x = -A$ and is equal to $-v_{\max}$.

The acceleration of the mass on the spring can be found by taking the time derivative of the velocity:

Equation:

$$a(t) = \frac{dv}{dt} = \frac{d}{dt}(-A\omega \sin(\omega t + \phi)) = -A\omega^2 \cos(\omega t + \phi) = -a_{\max} \cos(\omega t + \phi).$$

The maximum acceleration is $a_{\max} = A\omega^2$. The maximum acceleration occurs at the position ($x = -A$), and the acceleration at the position ($x = -A$) and is equal to $-a_{\max}$.

Summary of Equations of Motion for SHM

In summary, the oscillatory motion of a block on a spring can be modeled with the following equations of motion:

Note:

Equation:

$$x(t) = A\cos(\omega t + \phi)$$

Equation:

$$v(t) = -v_{\max}\sin(\omega t + \phi)$$

Equation:

$$a(t) = -a_{\max}\cos(\omega t + \phi)$$

Equation:

$$x_{\max} = A$$

Equation:

$$v_{\max} = A\omega$$

Equation:

$$a_{\max} = A\omega^2.$$

Here, A is the amplitude of the motion, T is the period, ϕ is the phase shift, and $\omega = \frac{2\pi}{T} = 2\pi f$ is the angular frequency of the motion of the block.

Example:**Determining the Equations of Motion for a Block and a Spring**

A 2.00-kg block is placed on a frictionless surface. A spring with a force constant of $k = 32.00 \text{ N/m}$ is attached to the block, and the opposite end of the spring is attached to the wall. The spring can be compressed or extended. The equilibrium position is marked as $x = 0.00 \text{ m}$.

Work is done on the block, pulling it out to $x = +0.02 \text{ m}$. The block is released from rest and oscillates between $x = +0.02 \text{ m}$ and $x = -0.02 \text{ m}$. The period of the motion is 1.57 s. Determine the equations of motion.

Strategy

We first find the angular frequency. The phase shift is zero, $\phi = 0.00 \text{ rad}$, because the block is released from rest at $x = A = +0.02 \text{ m}$. Once the angular frequency is found, we can determine the maximum velocity and maximum acceleration.

Solution

The angular frequency can be found and used to find the maximum velocity and maximum acceleration:

Equation:

$$\begin{aligned}\omega &= \frac{2\pi}{1.57 \text{ s}} = 4.00 \text{ s}^{-1}; \\ v_{\text{max}} &= A\omega = 0.02 \text{ m} (4.00 \text{ s}^{-1}) = 0.08 \text{ m/s}; \\ a_{\text{max}} &= A\omega^2 = 0.02 \text{ m} (4.00 \text{ s}^{-1})^2 = 0.32 \text{ m/s}^2.\end{aligned}$$

All that is left is to fill in the equations of motion:

Equation:

$$\begin{aligned}x(t) &= A \cos(\omega t + \phi) = (0.02 \text{ m}) \cos(4.00 \text{ s}^{-1}t); \\ v(t) &= -v_{\text{max}} \sin(\omega t + \phi) = (-0.08 \text{ m/s}) \sin(4.00 \text{ s}^{-1}t); \\ a(t) &= -a_{\text{max}} \cos(\omega t + \phi) = (-0.32 \text{ m/s}^2) \cos(4.00 \text{ s}^{-1}t).\end{aligned}$$

Significance

The position, velocity, and acceleration can be found for any time. It is important to remember that when using these equations, your calculator must be in radians mode.

The Period and Frequency of a Mass on a Spring

One interesting characteristic of the SHM of an object attached to a spring is that the angular frequency, and therefore the period and frequency of the motion, depend on only the mass and the force constant, and not on other factors such as the amplitude of

the motion. We can use the equations of motion and Newton's second law ($\vec{\mathbf{F}}_{\text{net}} = m\vec{\mathbf{a}}$) to find equations for the angular frequency, frequency, and period.

Consider the block on a spring on a frictionless surface. There are three forces on the mass: the weight, the normal force, and the force due to the spring. The only two forces that act perpendicular to the surface are the weight and the normal force, which have equal magnitudes and opposite directions, and thus sum to zero. The only force that acts parallel to the surface is the force due to the spring, so the net force must be equal to the force of the spring:

Equation:

$$\begin{aligned}F_x &= -kx; \\ma &= -kx; \\m\frac{d^2x}{dt^2} &= -kx; \\\frac{d^2x}{dt^2} &= -\frac{k}{m}x.\end{aligned}$$

Substituting the equations of motion for x and a gives us

Equation:

$$-A\omega^2\cos(\omega t + \phi) = -\frac{k}{m}A\cos(\omega t + \phi).$$

Cancelling out like terms and solving for the angular frequency yields

Note:

Equation:

$$\omega = \sqrt{\frac{k}{m}}.$$

The angular frequency depends only on the force constant and the mass, and not the amplitude. The angular frequency is defined as $\omega = 2\pi/T$, which yields an equation for the period of the motion:

Note:

Equation:

$$T = 2\pi\sqrt{\frac{m}{k}}.$$

The period also depends only on the mass and the force constant. The greater the mass, the longer the period. The stiffer the spring, the shorter the period. The frequency is

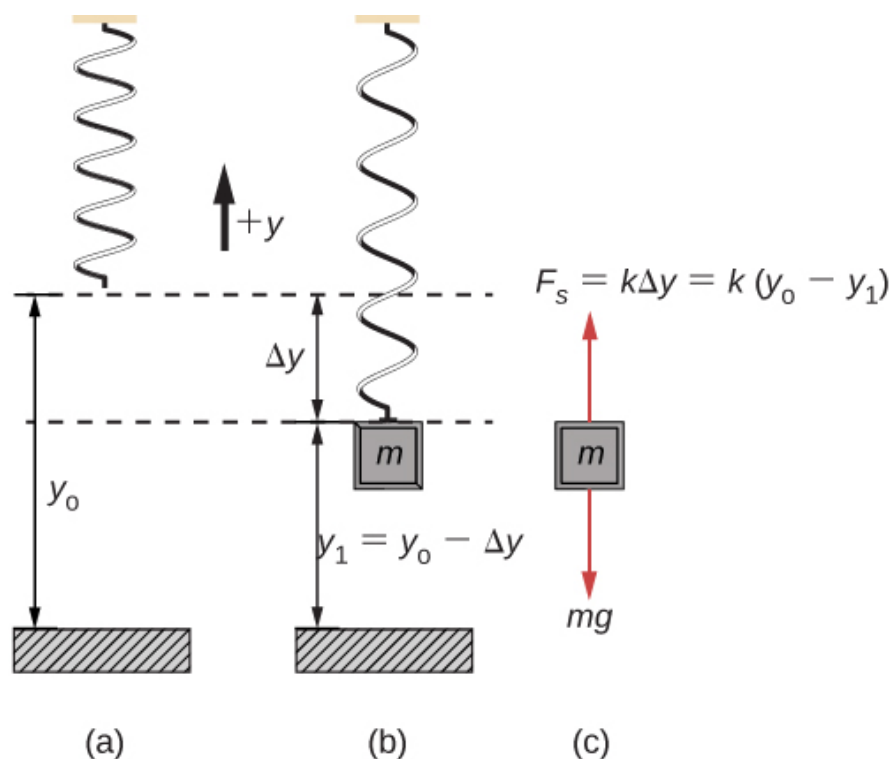
Note:

Equation:

$$f = \frac{1}{T} = \frac{1}{2\pi}\sqrt{\frac{k}{m}}.$$

Vertical Motion and a Horizontal Spring

When a spring is hung vertically and a block is attached and set in motion, the block oscillates in SHM. In this case, there is no normal force, and the net effect of the force of gravity is to change the equilibrium position. Consider [\[link\]](#). Two forces act on the block: the weight and the force of the spring. The weight is constant and the force of the spring changes as the length of the spring changes.



A spring is hung from the ceiling. When a block is attached, the block is at the equilibrium position where the weight of the block is equal to the force of the spring.

(a) The spring is hung from the ceiling and the equilibrium position is marked as y_0 . (b) A mass is attached to the spring and a new equilibrium position is reached ($y_1 = y_0 - \Delta y$) when the force provided by the spring equals the weight of the mass. (c) The free-body diagram of the mass shows the two forces acting on the mass: the weight and the force of the spring.

When the block reaches the equilibrium position, as seen in [\[link\]](#), the force of the spring equals the weight of the block, $F_{\text{net}} = F_s - mg = 0$, where

Equation:

$$-k(-\Delta y) = mg.$$

From the figure, the change in the position is $\Delta y = y_0 - y_1$ and since $-k(-\Delta y) = mg$, we have

Equation:

$$k(y_0 - y_1) - mg = 0.$$

If the block is displaced and released, it will oscillate around the new equilibrium position. As shown in [\[link\]](#), if the position of the block is recorded as a function of time, the recording is a periodic function.

If the block is displaced to a position y , the net force becomes

$F_{\text{net}} = k(y - y_0) - mg = 0$. But we found that at the equilibrium position, $mg = k\Delta y = ky_0 - ky_1$. Substituting for the weight in the equation yields

Equation:

$$F_{\text{net}} = ky - ky_0 - (ky_0 - ky_1) = -k(y - y_1).$$

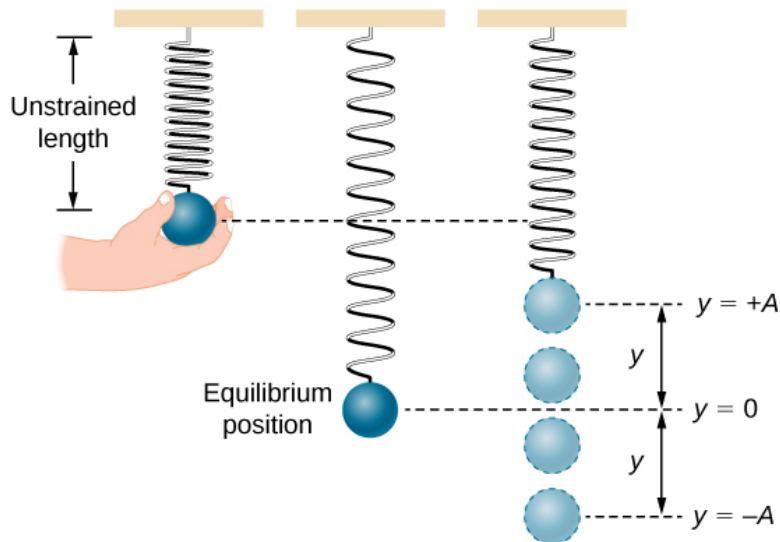
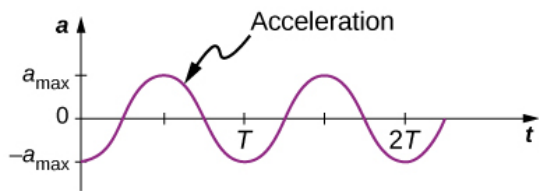
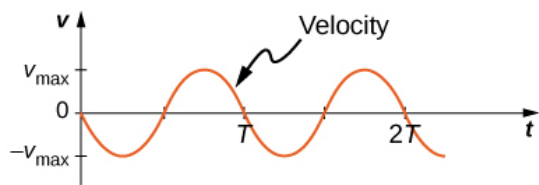
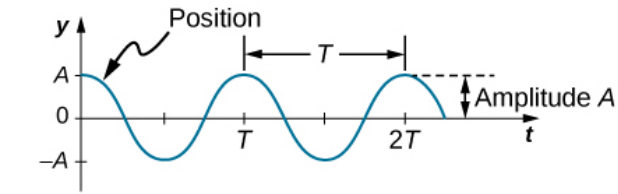
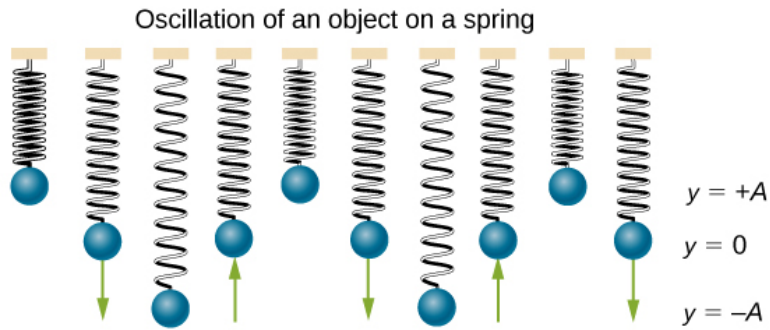
Recall that y_1 is just the equilibrium position and any position can be set to be the point $y = 0.00\text{m}$. So let's set y_1 to $y = 0.00\text{ m}$. The net force then becomes

Equation:

$$F_{\text{net}} = -ky;$$

$$m \frac{d^2 y}{dt^2} = -ky.$$

This is just what we found previously for a horizontally sliding mass on a spring. The constant force of gravity only served to shift the equilibrium location of the mass. Therefore, the solution should be the same form as for a block on a horizontal spring, $y(t) = A \cos(\omega t + \phi)$. The equations for the velocity and the acceleration also have the same form as for the horizontal case. Note that the inclusion of the phase shift means that the motion can actually be modeled using either a cosine or a sine function, since these two functions only differ by a phase shift.



Graphs of $y(t)$, $v(t)$, and $a(t)$ versus t for the motion of an object on a vertical spring. The net force on the object can be described by Hooke's law, so the object undergoes SHM. Note that the

initial position has the vertical displacement at its maximum value A ; v is initially zero and then negative as the object moves down; the initial acceleration is negative, back toward the equilibrium position and becomes zero at that point.

Summary

- Periodic motion is a repeating oscillation. The time for one oscillation is the period T and the number of oscillations per unit time is the frequency f . These quantities are related by $f = \frac{1}{T}$.
- Simple harmonic motion (SHM) is oscillatory motion for a system where the restoring force is proportional to the displacement and acts in the direction opposite to the displacement.
- Maximum displacement is the amplitude A . The angular frequency ω , period T , and frequency f of a simple harmonic oscillator are given by $\omega = \sqrt{\frac{k}{m}}$, $T = 2\pi\sqrt{\frac{m}{k}}$, and $f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$, where m is the mass of the system and k is the force constant.
- Displacement as a function of time in SHM is given by $x(t) = A \cos\left(\frac{2\pi}{T}t + \phi\right) = A \cos(\omega t + \phi)$.
- The velocity is given by $v(t) = -A\omega \sin(\omega t + \phi) = -v_{\max} \sin(\omega t + \phi)$, where $v_{\max} = A\omega = A\sqrt{\frac{k}{m}}$.
- The acceleration is $a(t) = -A\omega^2 \cos(\omega t + \phi) = -a_{\max} \cos(\omega t + \phi)$, where $a_{\max} = A\omega^2 = A\frac{k}{m}$.

Conceptual Questions

Exercise:

Problem: What conditions must be met to produce SHM?

Solution:

The restoring force must be proportional to the displacement and act opposite to the direction of motion with no drag forces or friction. The frequency of

oscillation does not depend on the amplitude.

Exercise:

Problem:

- (a) If frequency is not constant for some oscillation, can the oscillation be SHM?
- (b) Can you think of any examples of harmonic motion where the frequency may depend on the amplitude?

Exercise:

Problem:

Give an example of a simple harmonic oscillator, specifically noting how its frequency is independent of amplitude.

Solution:

Examples: Mass attached to a spring on a frictionless table, a mass hanging from a string, a simple pendulum with a small amplitude of motion. All of these examples have frequencies of oscillation that are independent of amplitude.

Exercise:

Problem:

Explain why you expect an object made of a stiff material to vibrate at a higher frequency than a similar object made of a more pliable material.

Exercise:

Problem:

As you pass a freight truck with a trailer on a highway, you notice that its trailer is bouncing up and down slowly. Is it more likely that the trailer is heavily loaded or nearly empty? Explain your answer.

Solution:

Since the frequency is proportional to the square root of the force constant and inversely proportional to the square root of the mass, it is likely that the truck is heavily loaded, since the force constant would be the same whether the truck is empty or heavily loaded.

Exercise:

Problem:

Some people modify cars to be much closer to the ground than when manufactured. Should they install stiffer springs? Explain your answer.

Problems**Exercise:****Problem:**

Prove that using $x(t) = A \sin(\omega t + \phi)$ will produce the same results for the period for the oscillations of a mass and a spring. Why do you think the cosine function was chosen?

Solution:

Proof

Exercise:

Problem: What is the period of 60.0 Hz of electrical power?

Exercise:**Problem:**

If your heart rate is 150 beats per minute during strenuous exercise, what is the time per beat in units of seconds?

Solution:

0.400 s/beat

Exercise:**Problem:**

Find the frequency of a tuning fork that takes 2.50×10^{-3} s to complete one oscillation.

Exercise:

Problem:

A stroboscope is set to flash every 8.00×10^{-5} s. What is the frequency of the flashes?

Solution:

12,500 Hz

Exercise:**Problem:**

A tire has a tread pattern with a crevice every 2.00 cm. Each crevice makes a single vibration as the tire moves. What is the frequency of these vibrations if the car moves at 30.0 m/s?

Exercise:**Problem:**

Each piston of an engine makes a sharp sound every other revolution of the engine. (a) How fast is a race car going if its eight-cylinder engine emits a sound of frequency 750 Hz, given that the engine makes 2000 revolutions per kilometer? (b) At how many revolutions per minute is the engine rotating?

Solution:

a. 340 km/hr; b. 11.3×10^3 rev/min

Exercise:**Problem:**

A type of cuckoo clock keeps time by having a mass bouncing on a spring, usually something cute like a cherub in a chair. What force constant is needed to produce a period of 0.500 s for a 0.0150-kg mass?

Exercise:**Problem:**

A mass m_0 is attached to a spring and hung vertically. The mass is raised a short distance in the vertical direction and released. The mass oscillates with a frequency f_0 . If the mass is replaced with a mass nine times as large, and the experiment was repeated, what would be the frequency of the oscillations in terms of f_0 ?

Solution:

$$f = \frac{1}{3} f_0$$

Exercise:**Problem:**

A 0.500-kg mass suspended from a spring oscillates with a period of 1.50 s. How much mass must be added to the object to change the period to 2.00 s?

Exercise:**Problem:**

How much leeway (both percentage and mass) would you have in the selection of the mass of the object in the previous problem if you did not wish the new period to be greater than 2.01 s or less than 1.99 s?

Solution:

0.009 kg; 2%

Glossary

amplitude (A)

maximum displacement from the equilibrium position of an object oscillating around the equilibrium position

equilibrium position

position where the spring is neither stretched nor compressed

force constant (k)

characteristic of a spring which is defined as the ratio of the force applied to the spring to the displacement caused by the force

frequency (f)

number of events per unit of time

oscillation

single fluctuation of a quantity, or repeated and regular fluctuations of a quantity, between two extreme values around an equilibrium or average value

periodic motion

motion that repeats itself at regular time intervals

period (T)

time taken to complete one oscillation

phase shift

angle, in radians, that is used in a cosine or sine function to shift the function left or right, used to match up the function with the initial conditions of data

simple harmonic motion (SHM)

oscillatory motion in a system where the restoring force is proportional to the displacement, which acts in the direction opposite to the displacement

simple harmonic oscillator

a device that oscillates in SHM where the restoring force is proportional to the displacement and acts in the direction opposite to the displacement

Energy in Simple Harmonic Motion

By the end of this section, you will be able to:

- Describe the energy conservation of the system of a mass and a spring
- Explain the concepts of stable and unstable equilibrium points

To produce a deformation in an object, we must do work. That is, whether you pluck a guitar string or compress a car's shock absorber, a force must be exerted through a distance. If the only result is deformation, and no work goes into thermal, sound, or kinetic energy, then all the work is initially stored in the deformed object as some form of potential energy.

Consider the example of a block attached to a spring on a frictionless table, oscillating in SHM. The force of the spring is a conservative force (which you studied in the chapter on potential energy and conservation of energy), and we can define a potential energy for it. This potential energy is the energy stored in the spring when the spring is extended or compressed. In this case, the block oscillates in one dimension with the force of the spring acting parallel to the motion:

Equation:

$$W = \int_{x_i}^{x_f} F_x dx = \int_{x_i}^{x_f} -kx dx = \left[-\frac{1}{2} kx^2 \right]_{x_i}^{x_f} = - \left[\frac{1}{2} kx_f^2 - \frac{1}{2} kx_i^2 \right] = -[U_f - U_i] = -\Delta U.$$

When considering the energy stored in a spring, the equilibrium position, marked as $x_i = 0.00$ m, is the position at which the energy stored in the spring is equal to zero. When the spring is stretched or compressed a distance x , the potential energy stored in the spring is

Equation:

$$U = \frac{1}{2} kx^2.$$

Energy and the Simple Harmonic Oscillator

To study the energy of a simple harmonic oscillator, we need to consider all the forms of energy. Consider the example of a block attached to a spring, placed on a frictionless surface, oscillating in SHM. The potential energy stored in the deformation of the spring is

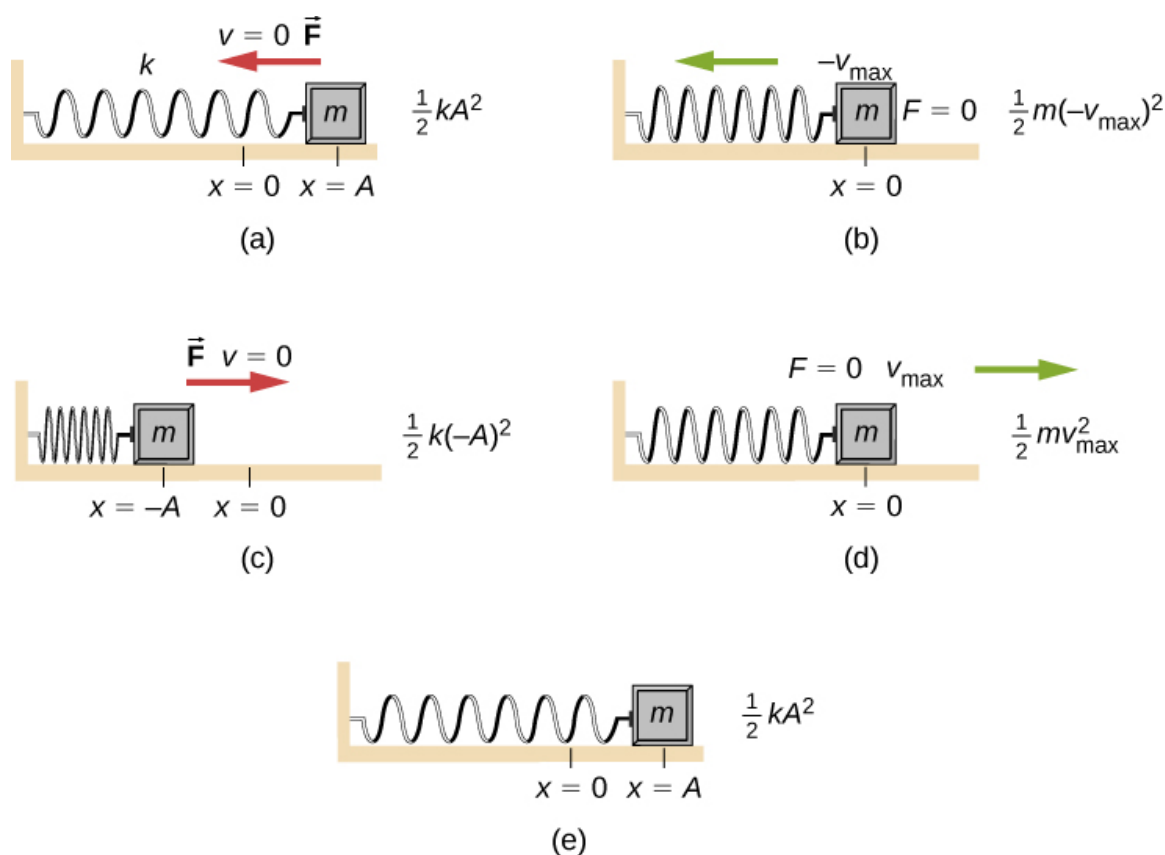
Equation:

$$U = \frac{1}{2} kx^2.$$

In a simple harmonic oscillator, the energy oscillates between kinetic energy of the mass $K = \frac{1}{2} mv^2$ and potential energy $U = \frac{1}{2} kx^2$ stored in the spring. In the SHM of the mass and spring system, there are no dissipative forces, so the total energy is the sum of the potential energy and kinetic energy. In this section, we consider the conservation of energy of the system.

The concepts examined are valid for all simple harmonic oscillators, including those where the gravitational force plays a role.

Consider [\[link\]](#), which shows an oscillating block attached to a spring. In the case of undamped SHM, the energy oscillates back and forth between kinetic and potential, going completely from one form of energy to the other as the system oscillates. So for the simple example of an object on a frictionless surface attached to a spring, the motion starts with all of the energy stored in the spring as **elastic potential energy**. As the object starts to move, the elastic potential energy is converted into kinetic energy, becoming entirely kinetic energy at the equilibrium position. The energy is then converted back into elastic potential energy by the spring as it is stretched or compressed. The velocity becomes zero when the kinetic energy is completely converted, and this cycle then repeats. Understanding the conservation of energy in these cycles will provide extra insight here and in later applications of SHM, such as alternating circuits.



The transformation of energy in SHM for an object attached to a spring on a frictionless surface. (a) When the mass is at the position $x = +A$, all the energy is stored as potential energy in the spring $U = \frac{1}{2}kA^2$. The kinetic energy is equal to zero because the velocity of the mass is zero. (b) As the mass moves toward $x = -A$, the mass crosses the position $x = 0$. At this point, the spring is neither extended nor compressed, so the potential energy stored in the spring is zero. At $x = 0$, the total energy is all kinetic energy where $K = \frac{1}{2}m(-v_{\max})^2$. (c) The mass continues to

move until it reaches $x = -A$ where the mass stops and starts moving toward $x = +A$. At the position $x = -A$, the total energy is stored as potential energy in the compressed $U = \frac{1}{2}k(-A)^2$ and the kinetic energy is zero. (d) As the mass passes through the position $x = 0$, the kinetic energy is $K = \frac{1}{2}mv_{\max}^2$ and the potential energy stored in the spring is zero. (e) The mass returns to the position $x = +A$, where $K = 0$ and $U = \frac{1}{2}kA^2$.

Consider [\[link\]](#), which shows the energy at specific points on the periodic motion. While staying constant, the energy oscillates between the kinetic energy of the block and the potential energy stored in the spring:

Equation:

$$E_{\text{Total}} = U + K = \frac{1}{2}kx^2 + \frac{1}{2}mv^2.$$

The motion of the block on a spring in SHM is defined by the position $x(t) = A\cos(\omega t + \phi)$ with a velocity of $v(t) = -A\omega\sin(\omega t + \phi)$. Using these equations, the trigonometric identity $\cos^2\theta + \sin^2\theta = 1$ and $\omega = \sqrt{\frac{k}{m}}$, we can find the total energy of the system:

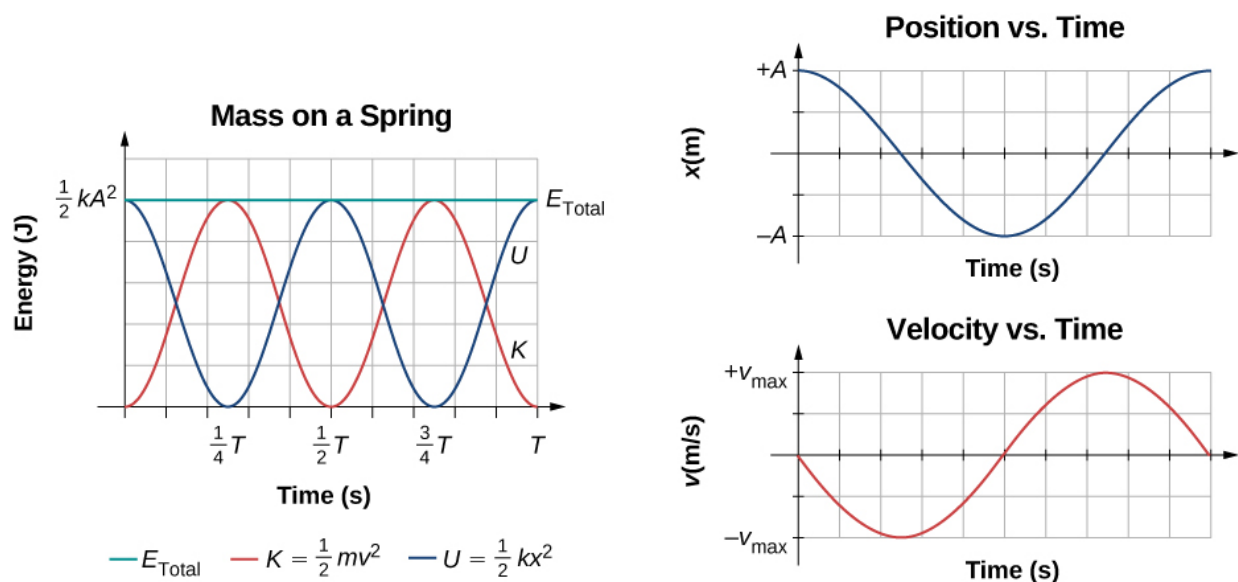
Equation:

$$\begin{aligned} E_{\text{Total}} &= \frac{1}{2}kA^2\cos^2(\omega t + \phi) + \frac{1}{2}mA^2\omega^2\sin^2(\omega t + \phi) \\ &= \frac{1}{2}kA^2\cos^2(\omega t + \phi) + \frac{1}{2}mA^2\left(\frac{k}{m}\right)\sin^2(\omega t + \phi) \\ &= \frac{1}{2}kA^2\cos^2(\omega t + \phi) + \frac{1}{2}kA^2\sin^2(\omega t + \phi) \\ &= \frac{1}{2}kA^2(\cos^2(\omega t + \phi) + \sin^2(\omega t + \phi)) \\ &= \frac{1}{2}kA^2. \end{aligned}$$

The total energy of the system of a block and a spring is equal to the sum of the potential energy stored in the spring plus the kinetic energy of the block and is proportional to the square of the amplitude $E_{\text{Total}} = (1/2)kA^2$. The total energy of the system is constant.

A closer look at the energy of the system shows that the kinetic energy oscillates like a sine-squared function, while the potential energy oscillates like a cosine-squared function. However, the total energy for the system is constant and is proportional to the amplitude squared. [\[link\]](#) shows a plot of the potential, kinetic, and total energies of the block and spring system as a function of time. Also plotted are the position and velocity as a function of time. Before time $t = 0.0$ s, the block is attached to the spring and placed at the equilibrium position. Work is done on the block by applying an external force, pulling it out to a position of $x = +A$. The system now has potential energy stored in the spring. At time $t = 0.00$ s, the position of the block is equal to the amplitude, the potential energy stored in the spring is equal to $U = \frac{1}{2}kA^2$, and the force on the block is maximum and points in the negative x-direction ($F_s = -kA$). The velocity

and kinetic energy of the block are zero at time $t = 0.00$ s. At time $t = 0.00$ s, the block is released from rest.

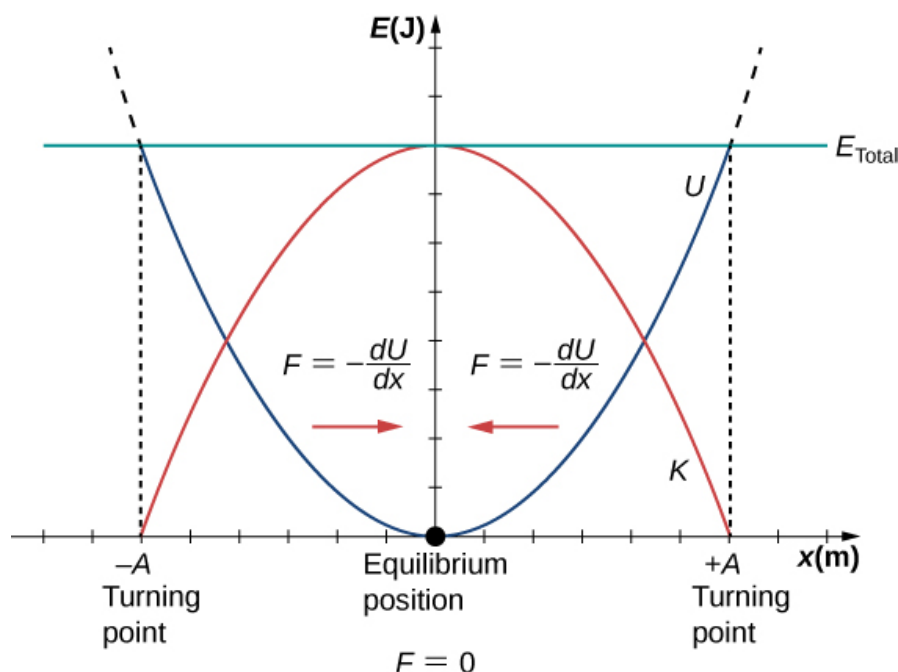


Graph of the kinetic energy, potential energy, and total energy of a block oscillating on a spring in SHM. Also shown are the graphs of position versus time and velocity versus time.

The total energy remains constant, but the energy oscillates between kinetic energy and potential energy. When the kinetic energy is maximum, the potential energy is zero. This occurs when the velocity is maximum and the mass is at the equilibrium position. The potential energy is maximum when the speed is zero. The total energy is the sum of the kinetic energy plus the potential energy and it is constant.

Oscillations About an Equilibrium Position

We have just considered the energy of SHM as a function of time. Another interesting view of the simple harmonic oscillator is to consider the energy as a function of position. [\[link\]](#) shows a graph of the energy versus position of a system undergoing SHM.



A graph of the kinetic energy (red), potential energy (blue), and total energy (green) of a simple harmonic oscillator. The force is equal to $F = -\frac{dU}{dx}$. The equilibrium position is shown as a black dot and is the point where the force is equal to zero. The force is positive when $x < 0$, negative when $x > 0$, and equal to zero when $x = 0$.

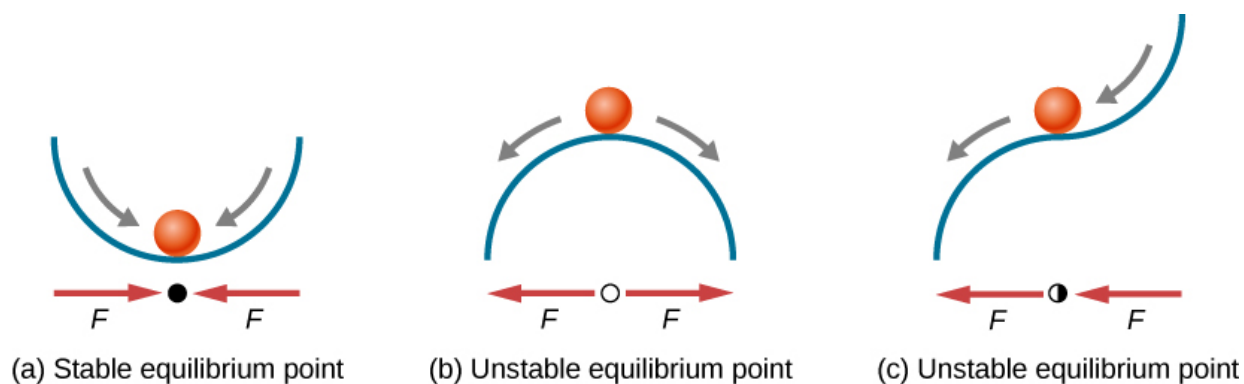
The potential energy curve in [\[link\]](#) resembles a bowl. When a marble is placed in a bowl, it settles to the equilibrium position at the lowest point of the bowl ($x = 0$). This happens because a **restoring force** points toward the equilibrium point. This equilibrium point is sometimes referred to as a *fixed point*. When the marble is disturbed to a different position ($x = +A$), the marble oscillates around the equilibrium position. Looking back at the graph of potential energy, the force can be found by looking at the slope of the potential energy graph ($F = -\frac{dU}{dx}$). Since the force on either side of the fixed point points back toward the equilibrium point, the equilibrium point is called a **stable equilibrium point**. The points $x = A$ and $x = -A$ are called the turning points. (See [Potential Energy and Conservation of Energy](#).)

Stability is an important concept. If an equilibrium point is stable, a slight disturbance of an object that is initially at the stable equilibrium point will cause the object to oscillate around that point. The stable equilibrium point occurs because the force on either side is directed toward it. For an unstable equilibrium point, if the object is disturbed slightly, it does not return to the equilibrium point.

Consider the marble in the bowl example. If the bowl is right-side up, the marble, if disturbed slightly, will oscillate around the stable equilibrium point. If the bowl is turned upside down, the marble can be balanced on the top, at the equilibrium point where the net force is zero. However,

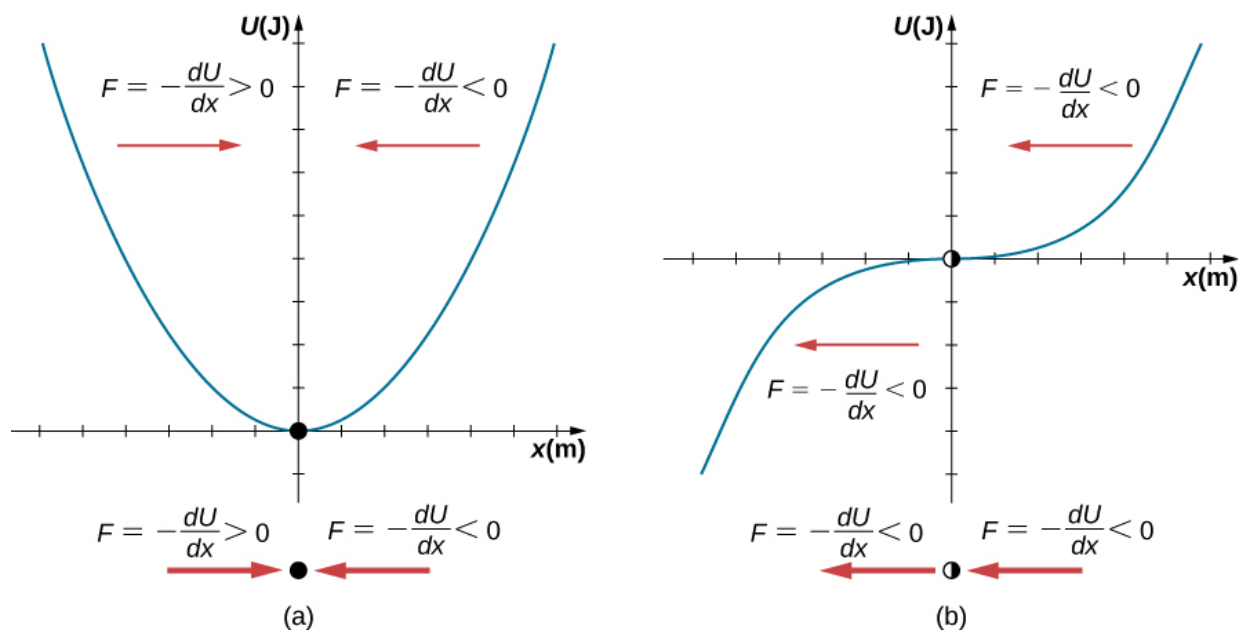
if the marble is disturbed slightly, it will not return to the equilibrium point, but will instead roll off the bowl. The reason is that the force on either side of the equilibrium point is directed away from that point. This point is an unstable equilibrium point.

[\[link\]](#) shows three conditions. The first is a stable equilibrium point (a), the second is an unstable equilibrium point (b), and the last is also an unstable equilibrium point (c), because the force on only one side points toward the equilibrium point.



Examples of equilibrium points. (a) Stable equilibrium point; (b) unstable equilibrium point; (c) unstable equilibrium point (sometimes referred to as a half-stable equilibrium point).

The process of determining whether an equilibrium point is stable or unstable can be formalized. Consider the potential energy curves shown in [\[link\]](#). The force can be found by analyzing the slope of the graph. The force is $F = -\frac{dU}{dx}$. In (a), the fixed point is at $x = 0.00$ m. When $x < 0.00$ m, the force is positive. When $x > 0.00$ m, the force is negative. This is a stable point. In (b), the fixed point is at $x = 0.00$ m. When $x < 0.00$ m, the force is negative. When $x > 0.00$ m, the force is also negative. This is an unstable point.



Two examples of a potential energy function. The force at a position is equal to the negative of the slope of the graph at that position. (a) A potential energy function with a stable equilibrium point. (b) A potential energy function with an unstable equilibrium point. This point is sometimes called half-stable because the force on one side points toward the fixed point.

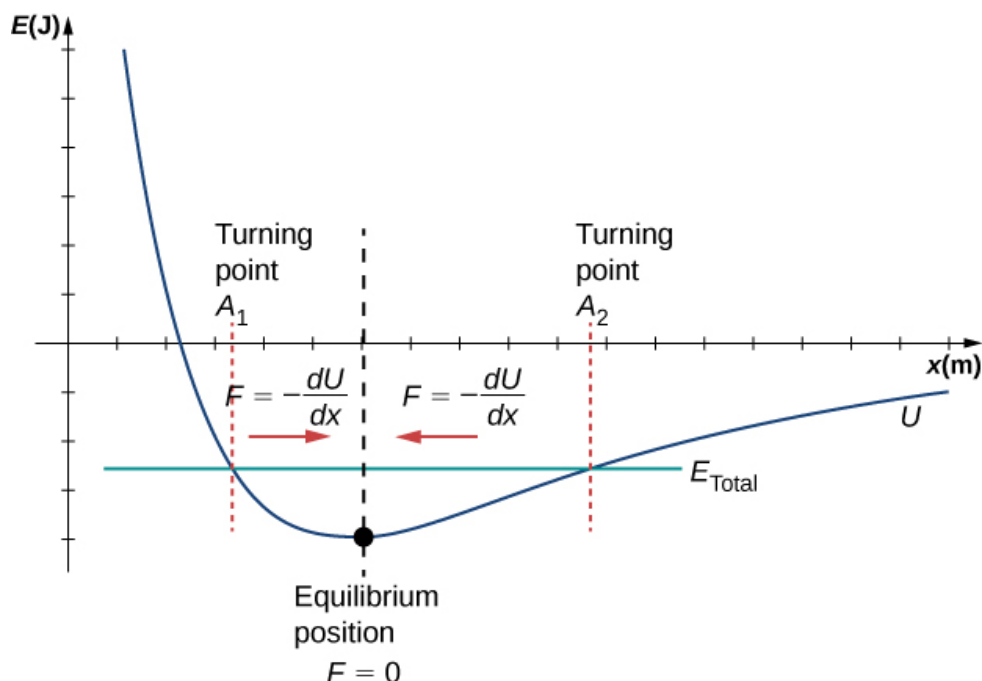
A practical application of the concept of stable equilibrium points is the force between two neutral atoms in a molecule. If two molecules are in close proximity, separated by a few atomic diameters, they can experience an attractive force. If the molecules move close enough so that the electron shells of the other electrons overlap, the force between the molecules becomes repulsive. The attractive force between the two atoms may cause the atoms to form a molecule. The force between the two molecules is not a linear force and cannot be modeled simply as two masses separated by a spring, but the atoms of the molecule can oscillate around an equilibrium point when displaced a small amount from the equilibrium position. The atoms oscillate due the attractive force and repulsive force between the two atoms.

Consider one example of the interaction between two atoms known as the van Der Waals interaction. It is beyond the scope of this chapter to discuss in depth the interactions of the two atoms, but the oscillations of the atoms can be examined by considering one example of a model of the potential energy of the system. One suggestion to model the potential energy of this molecule is with the Lennard-Jones 6-12 potential:

Equation:

$$U(x) = 4\epsilon \left[\left(\frac{\sigma}{x} \right)^{12} - \left(\frac{\sigma}{x} \right)^6 \right].$$

A graph of this function is shown in [\[link\]](#). The two parameters ε and σ are found experimentally.



The Lennard-Jones potential energy function for a system of two neutral atoms. If the energy is below some maximum energy, the system oscillates near the equilibrium position between the two turning points.

From the graph, you can see that there is a potential energy well, which has some similarities to the potential energy well of the potential energy function of the simple harmonic oscillator discussed in [\[link\]](#). The Lennard-Jones potential has a stable equilibrium point where the potential energy is minimum and the force on either side of the equilibrium point points toward equilibrium point. Note that unlike the simple harmonic oscillator, the potential well of the Lennard-Jones potential is not symmetric. This is due to the fact that the force between the atoms is not a Hooke's law force and is not linear. The atoms can still oscillate around the equilibrium position x_{min} because when $x < x_{min}$, the force is positive; when $x > x_{min}$, the force is negative. Notice that as x approaches zero, the slope is quite steep and negative, which means that the force is large and positive. This suggests that it takes a large force to try to push the atoms close together. As x becomes increasingly large, the slope becomes less steep and the force is smaller and negative. This suggests that if given a large enough energy, the atoms can be separated.

If you are interested in this interaction, find the force between the molecules by taking the derivative of the potential energy function. You will see immediately that the force does not resemble a Hooke's law force ($F = -kx$), but if you are familiar with the binomial theorem:

Equation:

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \dots,$$

the force can be approximated by a Hooke's law force.

Velocity and Energy Conservation

Getting back to the system of a block and a spring in [\[link\]](#), once the block is released from rest, it begins to move in the negative direction toward the equilibrium position. The potential energy decreases and the magnitude of the velocity and the kinetic energy increase. At time $t = T/4$, the block reaches the equilibrium position $x = 0.00$ m, where the force on the block and the potential energy are zero. At the equilibrium position, the block reaches a negative velocity with a magnitude equal to the maximum velocity $v = -A\omega$. The kinetic energy is maximum and equal to $K = \frac{1}{2}mv^2 = \frac{1}{2}mA^2\omega^2 = \frac{1}{2}kA^2$. At this point, the force on the block is zero, but momentum carries the block, and it continues in the negative direction toward $x = -A$. As the block continues to move, the force on it acts in the positive direction and the magnitude of the velocity and kinetic energy decrease. The potential energy increases as the spring compresses. At time $t = T/2$, the block reaches $x = -A$. Here the velocity and kinetic energy are equal to zero. The force on the block is $F = +kA$ and the potential energy stored in the spring is $U = \frac{1}{2}kA^2$. During the oscillations, the total energy is constant and equal to the sum of the potential energy and the kinetic energy of the system,

Note:

Equation:

$$E_{\text{Total}} = \frac{1}{2}kx^2 + \frac{1}{2}mv^2 = \frac{1}{2}kA^2.$$

The equation for the energy associated with SHM can be solved to find the magnitude of the velocity at any position:

Note:

Equation:

$$|v| = \sqrt{\frac{k}{m}(A^2 - x^2)}.$$

The energy in a simple harmonic oscillator is proportional to the square of the amplitude. When considering many forms of oscillations, you will find the energy proportional to the amplitude squared.

Note:

Exercise:

Problem:

Check Your Understanding Why would it hurt more if you snapped your hand with a ruler than with a loose spring, even if the displacement of each system is equal?

Solution:

The ruler is a stiffer system, which carries greater force for the same amount of displacement. The ruler snaps your hand with greater force, which hurts more.

Note:

Exercise:

Problem:

Check Your Understanding Identify one way you could decrease the maximum velocity of a simple harmonic oscillator.

Solution:

You could increase the mass of the object that is oscillating. Other options would be to reduce the amplitude, or use a less stiff spring.

Summary

- The simplest type of oscillations are related to systems that can be described by Hooke's law, $F = -kx$, where F is the restoring force, x is the displacement from equilibrium or deformation, and k is the force constant of the system.
- Elastic potential energy U stored in the deformation of a system that can be described by Hooke's law is given by $U = \frac{1}{2}kx^2$.
- Energy in the simple harmonic oscillator is shared between elastic potential energy and kinetic energy, with the total being constant:

Equation:

$$E_{\text{Total}} = \frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2}kA^2 = \text{constant}.$$

- The magnitude of the velocity as a function of position for the simple harmonic oscillator can be found by using

Equation:

$$|v| = \sqrt{\frac{k}{m}(A^2 - x^2)}.$$

Conceptual Questions

Exercise:

Problem: Describe a system in which elastic potential energy is stored.

Solution:

In a car, elastic potential energy is stored when the shock is extended or compressed. In some running shoes elastic potential energy is stored in the compression of the material of the soles of the running shoes. In pole vaulting, elastic potential energy is stored in the bending of the pole.

Exercise:

Problem:

Explain in terms of energy how dissipative forces such as friction reduce the amplitude of a harmonic oscillator. Also explain how a driving mechanism can compensate. (A pendulum clock is such a system.)

Exercise:

Problem:

The temperature of the atmosphere oscillates from a maximum near noontime and a minimum near sunrise. Would you consider the atmosphere to be in stable or unstable equilibrium?

Solution:

The overall system is stable. There may be times when the stability is interrupted by a storm, but the driving force provided by the sun bring the atmosphere back into a stable pattern.

Problems

Exercise:

Problem:

Fish are hung on a spring scale to determine their mass. (a) What is the force constant of the spring in such a scale if it the spring stretches 8.00 cm for a 10.0 kg load? (b) What is the mass of a fish that stretches the spring 5.50 cm? (c) How far apart are the half-kilogram marks on the scale?

Exercise:**Problem:**

It is weigh-in time for the local under-85-kg rugby team. The bathroom scale used to assess eligibility can be described by Hooke's law and is depressed 0.75 cm by its maximum load of 120 kg. (a) What is the spring's effective force constant? (b) A player stands on the scales and depresses it by 0.48 cm. Is he eligible to play on this under-85-kg team?

Solution:

a. $1.57 \times 10^5 \text{ N/m}$; b. 77 kg, yes, he is eligible to play

Exercise:**Problem:**

One type of BB gun uses a spring-driven plunger to blow the BB from its barrel. (a) Calculate the force constant of its plunger's spring if you must compress it 0.150 m to drive the 0.0500-kg plunger to a top speed of 20.0 m/s. (b) What force must be exerted to compress the spring?

Exercise:**Problem:**

When an 80.0-kg man stands on a pogo stick, the spring is compressed 0.120 m. (a) What is the force constant of the spring? (b) Will the spring be compressed more when he hops down the road?

Solution:

a. $6.53 \times 10^3 \text{ N/m}$; b. yes, when the man is at his lowest point in his hopping the spring will be compressed the most

Exercise:**Problem:**

A spring has a length of 0.200 m when a 0.300-kg mass hangs from it, and a length of 0.750 m when a 1.95-kg mass hangs from it. (a) What is the force constant of the spring? (b) What is the unloaded length of the spring?

Exercise:

Problem:

The length of nylon rope from which a mountain climber is suspended has an effective force constant of $1.40 \times 10^4 \text{ N/m}$. (a) What is the frequency at which he bounces, given his mass plus and the mass of his equipment are 90.0 kg? (b) How much would this rope stretch to break the climber's fall if he free-falls 2.00 m before the rope runs out of slack? (*Hint:* Use conservation of energy.) (c) Repeat both parts of this problem in the situation where twice this length of nylon rope is used.

Solution:

a. 1.99 Hz; b. 44.3 cm; c. 65.0 cm

Glossary

elastic potential energy

potential energy stored as a result of deformation of an elastic object, such as the stretching of a spring

restoring force

force acting in opposition to the force caused by a deformation

stable equilibrium point

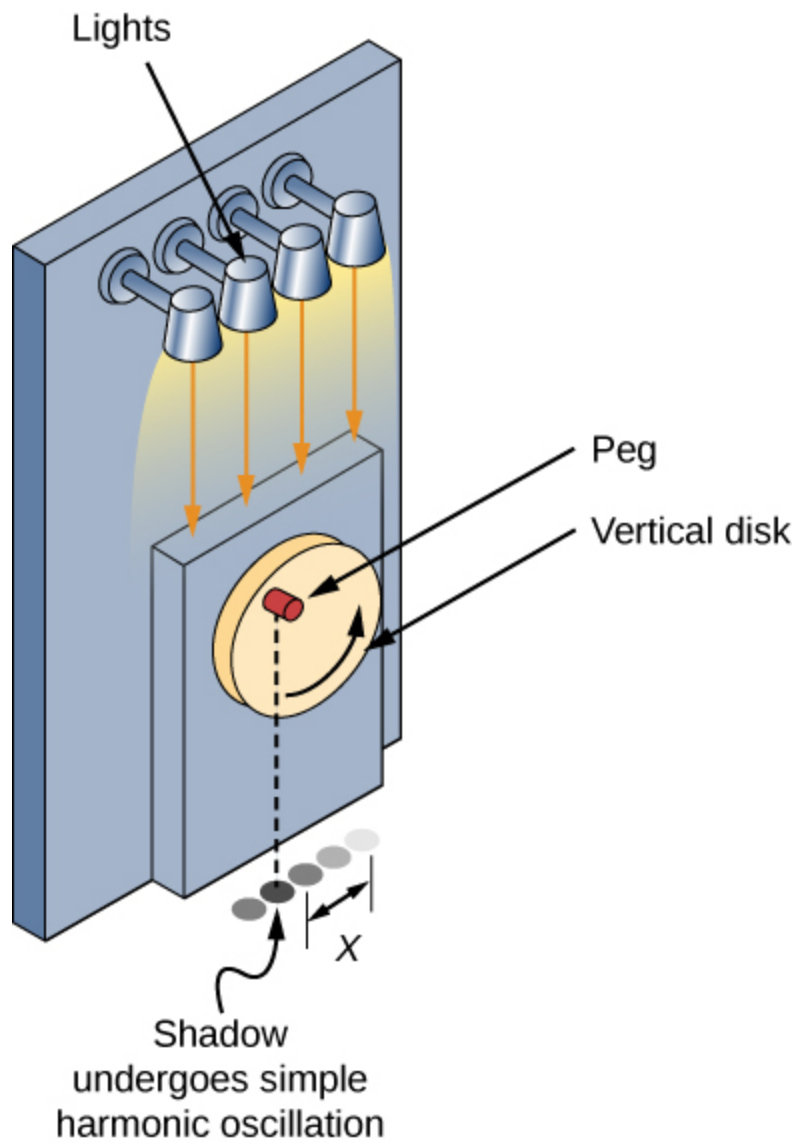
point where the net force on a system is zero, but a small displacement of the mass will cause a restoring force that points toward the equilibrium point

Comparing Simple Harmonic Motion and Circular Motion

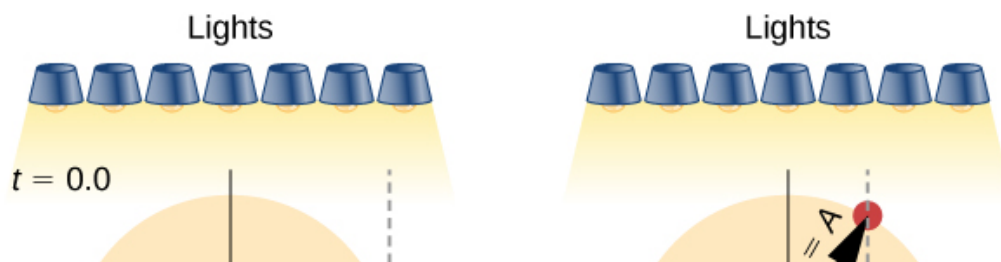
By the end of this section, you will be able to:

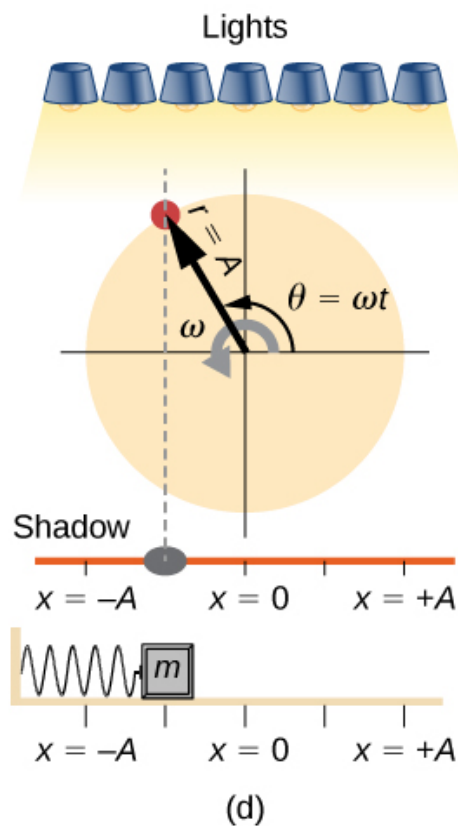
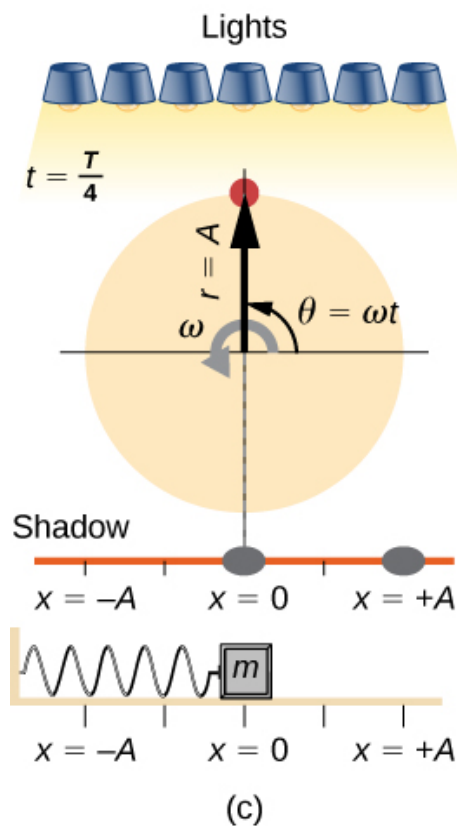
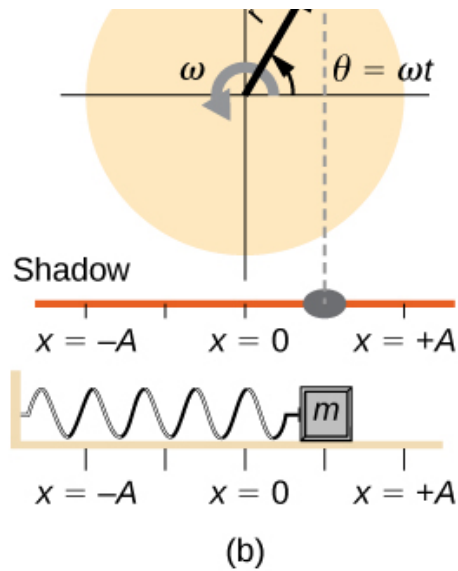
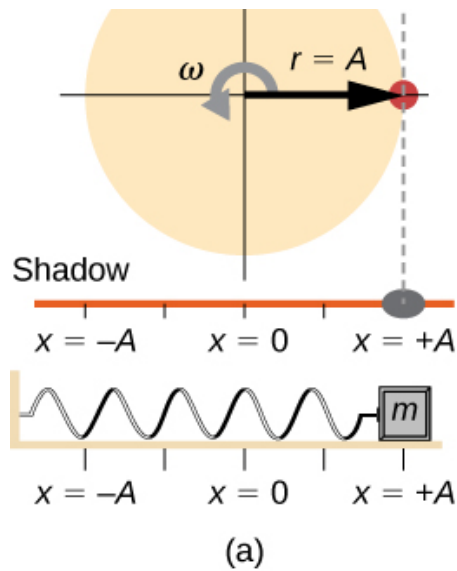
- Describe how the sine and cosine functions relate to the concepts of circular motion
- Describe the connection between simple harmonic motion and circular motion

An easy way to model SHM is by considering uniform circular motion. [\[link\]](#) shows one way of using this method. A peg (a cylinder of wood) is attached to a vertical disk, rotating with a constant angular frequency. [\[link\]](#) shows a side view of the disk and peg. If a lamp is placed above the disk and peg, the peg produces a shadow. Let the disk have a radius of $r = A$ and define the position of the shadow that coincides with the center line of the disk to be $x = 0.00\text{ m}$. As the disk rotates at a constant rate, the shadow oscillates between $x = +A$ and $x = -A$. Now imagine a block on a spring beneath the floor as shown in [\[link\]](#).



SHM can be modeled as rotational motion by looking at the shadow of a peg on a wheel rotating at a constant angular frequency.





Light shines down on the disk so that the peg makes a shadow. If the disk rotates at just the right angular frequency, the shadow follows the motion of the block on a spring. If there is no energy dissipated due to

nonconservative forces, the block and the shadow will oscillate back and forth in unison. In this figure, four snapshots are taken at four different times. (a) The wheel starts at $\theta = 0^\circ$ and the shadow of the peg is at $x = +A$, representing the mass at position $x = +A$. (b) As the disk rotates through an angle $\theta = \omega t$, the shadow of the peg is between $x = +A$ and $x = 0$. (c) The disk continues to rotate until $\theta = 90^\circ$, at which the shadow follows the mass to $x = 0$. (d) The disk continues to rotate, the shadow follows the position of the mass.

If the disk turns at the proper angular frequency, the shadow follows along with the block. The position of the shadow can be modeled with the equation

Note:

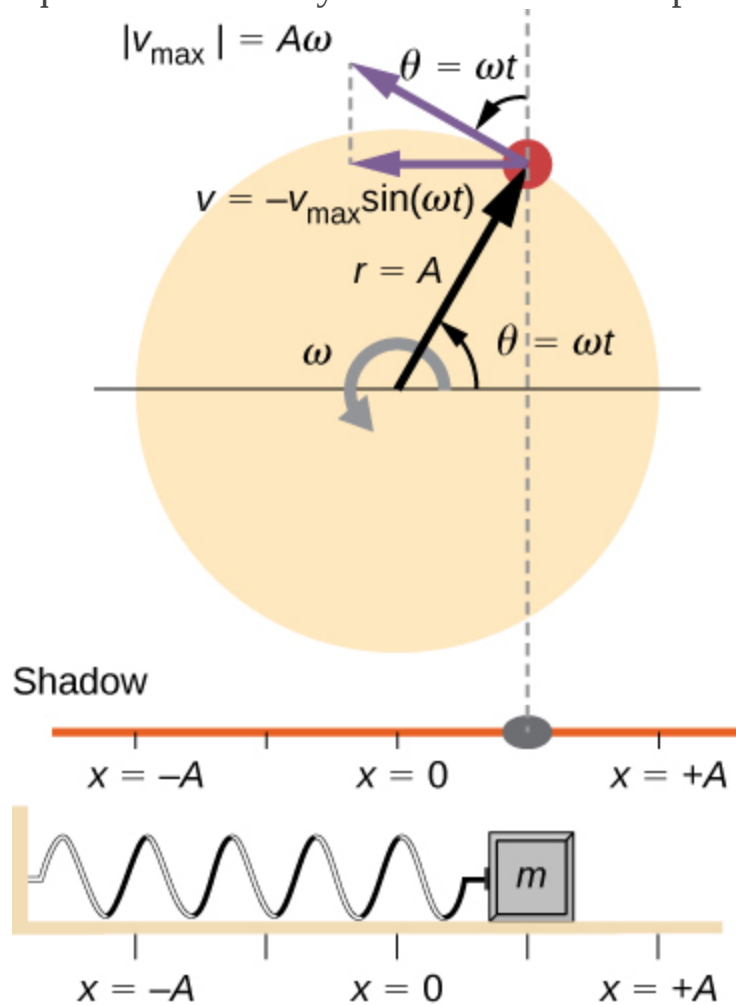
Equation:

$$x(t) = A\cos(\omega t).$$

Recall that the block attached to the spring does not move at a constant velocity. How often does the wheel have to turn to have the peg's shadow always on the block? The disk must turn at a constant angular frequency equal to 2π times the frequency of oscillation ($\omega = 2\pi f$).

[\[link\]](#) shows the basic relationship between uniform circular motion and SHM. The peg lies at the tip of the radius, a distance A from the center of the disk. The x -axis is defined by a line drawn parallel to the ground, cutting the disk in half. The y -axis (not shown) is defined by a line perpendicular to the ground, cutting the disk into a left half and a right half. The center of the disk is the point $(x = 0, y = 0)$. The projection of the

position of the peg onto the fixed x -axis gives the position of the shadow, which undergoes SHM analogous to the system of the block and spring. At the time shown in the figure, the projection has position x and moves to the left with velocity v . The tangential velocity of the peg around the circle equals \bar{v}_{\max} of the block on the spring. The x -component of the velocity is equal to the velocity of the block on the spring.



A peg moving on a circular path with a constant angular velocity ω is undergoing uniform circular motion. Its projection on the x -axis undergoes SHM. Also shown is the velocity of the peg around the circle, v_{\max} , and its projection, which is v . Note that these

velocities form a similar triangle to the displacement triangle.

We can use [\[link\]](#) to analyze the velocity of the shadow as the disk rotates. The peg moves in a circle with a speed of $v_{\max} = A\omega$. The shadow moves with a velocity equal to the component of the peg's velocity that is parallel to the surface where the shadow is being produced:

Note:

Equation:

$$v = -v_{\max}\sin(\omega t).$$

It follows that the acceleration is

Note:

Equation:

$$a = -a_{\max}\cos(\omega t).$$

Note:

Exercise:

Problem:

Check Your Understanding Identify an object that undergoes uniform circular motion. Describe how you could trace the SHM of this object.

Solution:

A ketchup bottle sits on a lazy Susan in the center of the dinner table. You set it rotating in uniform circular motion. A set of lights shine on the bottle, producing a shadow on the wall.

Summary

- A projection of uniform circular motion undergoes simple harmonic oscillation.
- Consider a circle with a radius A , moving at a constant angular speed ω . A point on the edge of the circle moves at a constant tangential speed of $v_{\max} = A\omega$. The projection of the radius onto the x -axis is $x(t) = A\cos(\omega t + \phi)$, where (ϕ) is the phase shift. The x -component of the tangential velocity is $v(t) = -A\omega\sin(\omega t + \phi)$.

Conceptual Questions**Exercise:****Problem:**

Can this analogy of SHM to circular motion be carried out with an object oscillating on a spring vertically hung from the ceiling? Why or why not? If given the choice, would you prefer to use a sine function or a cosine function to model the motion?

Exercise:**Problem:**

If the maximum speed of the mass attached to a spring, oscillating on a frictionless table, was increased, what characteristics of the rotating disk would need to be changed?

Solution:

The maximum speed is equal to $v_{\max} = A\omega$ and the angular frequency is independent of the amplitude, so the amplitude would be affected. The radius of the circle represents the amplitude of the circle, so make the amplitude larger.

Problems

Exercise:

Problem:

The motion of a mass on a spring hung vertically, where the mass oscillates up and down, can also be modeled using the rotating disk. Instead of the lights being placed horizontally along the top and pointing down, place the lights vertically and have the lights shine on the side of the rotating disk. A shadow will be produced on a nearby wall, and will move up and down. Write the equations of motion for the shadow taking the position at $t = 0.0$ s to be $y = 0.0$ m with the mass moving in the positive y -direction.

Exercise:

Problem:

(a) A novelty clock has a 0.0100-kg-mass object bouncing on a spring that has a force constant of 1.25 N/m. What is the maximum velocity of the object if the object bounces 3.00 cm above and below its equilibrium position? (b) How many joules of kinetic energy does the object have at its maximum velocity?

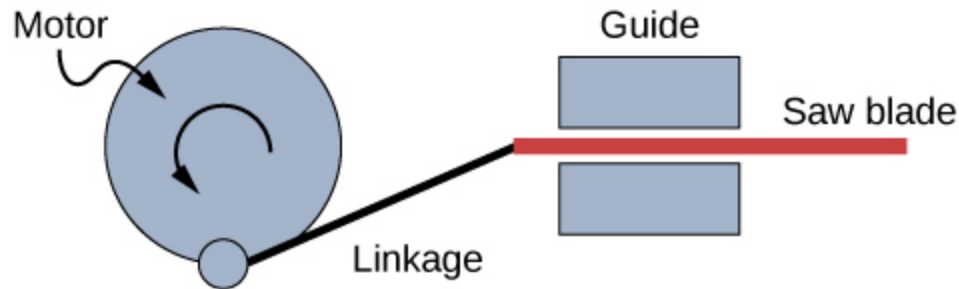
Solution:

a. 0.335 m/s; b. 5.61×10^{-4} J

Exercise:

Problem:

Reciprocating motion uses the rotation of a motor to produce linear motion up and down or back and forth. This is how a reciprocating saw operates, as shown below.



If the motor rotates at 60 Hz and has a radius of 3.0 cm, estimate the maximum speed of the saw blade as it moves left and right. This design is known as a scotch yoke.

Exercise:**Problem:**

A student stands on the edge of a merry-go-round which rotates five times a minute and has a radius of two meters one evening as the sun is setting. The student produces a shadow on the nearby building. (a) Write an equation for the position of the shadow. (b) Write an equation for the velocity of the shadow.

Solution:

a. $x(t) = 2 \text{ m} \cos(0.52 \text{ s}^{-1}t)$; b. $v(t) = (-1.05 \text{ m/s}) \sin(0.52 \text{ s}^{-1}t)$

Pendulums

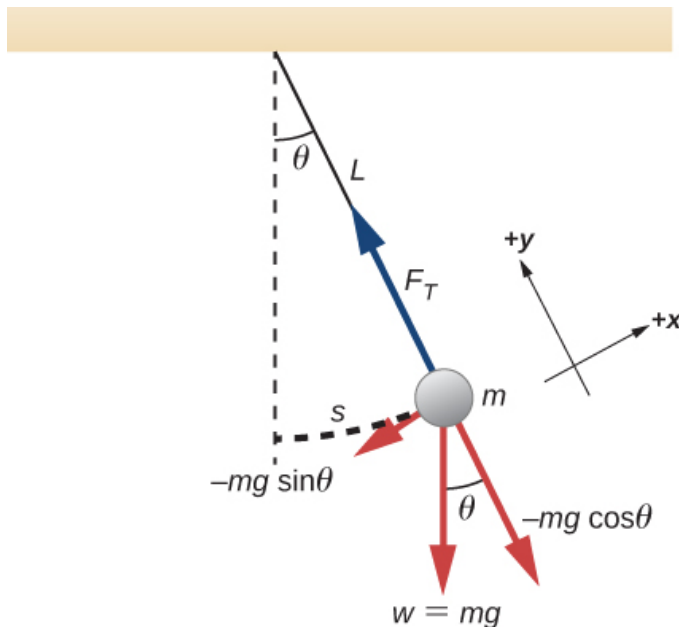
By the end of this section, you will be able to:

- State the forces that act on a simple pendulum
- Determine the angular frequency, frequency, and period of a simple pendulum in terms of the length of the pendulum and the acceleration due to gravity
- Define the period for a physical pendulum
- Define the period for a torsional pendulum

Pendulums are in common usage. Grandfather clocks use a pendulum to keep time and a pendulum can be used to measure the acceleration due to gravity. For small displacements, a pendulum is a simple harmonic oscillator.

The Simple Pendulum

A **simple pendulum** is defined to have a point mass, also known as the pendulum bob, which is suspended from a string of length L with negligible mass ([link](#)). Here, the only forces acting on the bob are the force of gravity (i.e., the weight of the bob) and tension from the string. The mass of the string is assumed to be negligible as compared to the mass of the bob.



A simple pendulum has a small-diameter bob and a string that has a very small mass but is strong enough not to stretch appreciably. The linear displacement from equilibrium is s , the length of the arc. Also shown are the forces on the bob, which result in a net force of $mg \sin \theta$ toward the equilibrium position.

— $mg \sin \theta$ toward the equilibrium position—
that is, a restoring force.

Consider the torque on the pendulum. The force providing the restoring torque is the component of the weight of the pendulum bob that acts along the arc length. The torque is the length of the string L times the component of the net force that is perpendicular to the radius of the arc. The minus sign indicates the torque acts in the opposite direction of the angular displacement:

Equation:

$$\begin{aligned}\tau &= -L(mg \sin \theta); \\ I\alpha &= -L(mg \sin \theta); \\ I \frac{d^2\theta}{dt^2} &= -L(mg \sin \theta); \\ mL^2 \frac{d^2\theta}{dt^2} &= -L(mg \sin \theta); \\ \frac{d^2\theta}{dt^2} &= -\frac{g}{L} \sin \theta.\end{aligned}$$

The solution to this differential equation involves advanced calculus, and is beyond the scope of this text. But note that for small angles (less than 15 degrees), $\sin \theta$ and θ differ by less than 1%, so we can use the small angle approximation $\sin \theta \approx \theta$. The angle θ describes the position of the pendulum. Using the small angle approximation gives an approximate solution for small angles,

Note:

Equation:

$$\frac{d^2\theta}{dt^2} = -\frac{g}{L} \theta.$$

Because this equation has the same form as the equation for SHM, the solution is easy to find. The angular frequency is

Note:

Equation:

$$\omega = \sqrt{\frac{g}{L}}$$

and the period is

Note:

Equation:

$$T = 2\pi\sqrt{\frac{L}{g}}.$$

The period of a simple pendulum depends on its length and the acceleration due to gravity. The period is completely independent of other factors, such as mass and the maximum displacement. As with simple harmonic oscillators, the period T for a pendulum is nearly independent of amplitude, especially if θ is less than about 15° . Even simple pendulum clocks can be finely adjusted and remain accurate.

Note the dependence of T on g . If the length of a pendulum is precisely known, it can actually be used to measure the acceleration due to gravity, as in the following example.

Example:

Measuring Acceleration due to Gravity by the Period of a Pendulum

What is the acceleration due to gravity in a region where a simple pendulum having a length 75.000 cm has a period of 1.7357 s?

Strategy

We are asked to find g given the period T and the length L of a pendulum. We can solve

$T = 2\pi\sqrt{\frac{L}{g}}$ for g , assuming only that the angle of deflection is less than 15° .

Solution

1. Square $T = 2\pi\sqrt{\frac{L}{g}}$ and solve for g :

Equation:

$$g = 4\pi^2 \frac{L}{T^2}.$$

2. Substitute known values into the new equation:

Equation:

$$g = 4\pi^2 \frac{0.75000 \text{ m}}{(1.7357 \text{ s})^2}.$$

3. Calculate to find g :

Equation:

$$g = 9.8281 \text{ m/s}^2.$$

Significance

This method for determining g can be very accurate, which is why length and period are given to five digits in this example. For the precision of the approximation $\sin \theta \approx \theta$ to be better than the precision of the pendulum length and period, the maximum displacement angle should be kept below about 0.5° .

Note:**Exercise:****Problem:**

Check Your Understanding An engineer builds two simple pendulums. Both are suspended from small wires secured to the ceiling of a room. Each pendulum hovers 2 cm above the floor. Pendulum 1 has a bob with a mass of 10 kg. Pendulum 2 has a bob with a mass of 100 kg. Describe how the motion of the pendulums will differ if the bobs are both displaced by 12° .

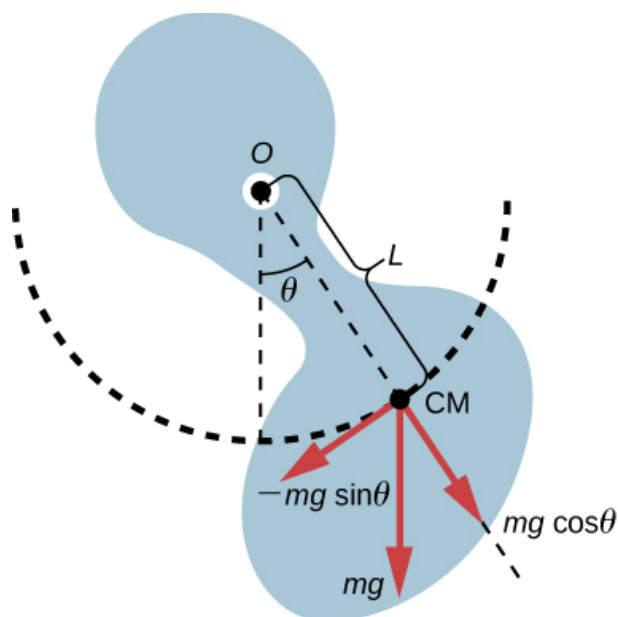
Solution:

The movement of the pendulums will not differ at all because the mass of the bob has no effect on the motion of a simple pendulum. The pendulums are only affected by the period (which is related to the pendulum's length) and by the acceleration due to gravity.

Physical Pendulum

Any object can oscillate like a pendulum. Consider a coffee mug hanging on a hook in the pantry. If the mug gets knocked, it oscillates back and forth like a pendulum until the oscillations die out. We have described a simple pendulum as a point mass and a string. A **physical pendulum** is any object whose oscillations are similar to those of the simple pendulum, but cannot be modeled as a point mass on a string, and the mass distribution must be included into the equation of motion.

As for the simple pendulum, the restoring force of the physical pendulum is the force of gravity. With the simple pendulum, the force of gravity acts on the center of the pendulum bob. In the case of the physical pendulum, the force of gravity acts on the center of mass (CM) of an object. The object oscillates about a point O . Consider an object of a generic shape as shown in [\[link\]](#).



A physical pendulum is any object that oscillates as a pendulum, but cannot be modeled as a point mass on a string. The force of gravity acts on the center of mass (CM) and provides the restoring force that causes the object to oscillate. The minus sign on the component of the weight that provides the restoring force is present because the force acts in the opposite direction of the increasing angle θ .

When a physical pendulum is hanging from a point but is free to rotate, it rotates because of the torque applied at the CM, produced by the component of the object's weight that acts tangent to the motion of the CM. Taking the counterclockwise direction to be positive, the component of the gravitational force that acts tangent to the motion is $-mg \sin \theta$. The minus sign is the result of the restoring force acting in the opposite direction of the increasing angle. Recall that the torque is equal to $\vec{\tau} = \vec{r} \times \vec{F}$. The magnitude of the torque is equal to the length of the radius arm times the tangential component of the force applied, $|\tau| = rF \sin \theta$. Here, the length L of the radius arm is the distance between the point of rotation and the CM. To analyze the motion, start with the net torque. Like the simple pendulum, consider only small angles so that $\sin \theta \approx \theta$. Recall from [Fixed-Axis Rotation](#) on rotation that the net torque is equal to the moment of inertia $I = \int r^2 dm$ times the angular acceleration α , where $\alpha = \frac{d^2\theta}{dt^2}$:

Equation:

$$I\alpha = \tau_{\text{net}} = L(-mg)\sin \theta.$$

Using the small angle approximation and rearranging:

Equation:

$$\begin{aligned}I\alpha &= -L(mg)\theta; \\I\frac{d^2\theta}{dt^2} &= -L(mg)\theta; \\\frac{d^2\theta}{dt^2} &= -\left(\frac{mgL}{I}\right)\theta.\end{aligned}$$

Once again, the equation says that the second time derivative of the position (in this case, the angle) equals minus a constant $\left(-\frac{mgL}{I}\right)$ times the position. The solution is

Equation:

$$\theta(t) = \Theta \cos(\omega t + \phi),$$

where Θ is the maximum angular displacement. The angular frequency is

Note:

Equation:

$$\omega = \sqrt{\frac{mgL}{I}}.$$

The period is therefore

Note:

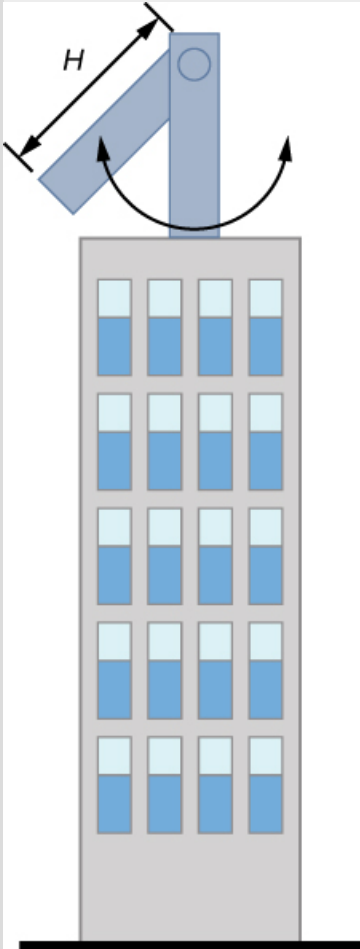
Equation:

$$T = 2\pi\sqrt{\frac{I}{mgL}}.$$

Note that for a simple pendulum, the moment of inertia is $I = \int r^2 dm = mL^2$ and the period reduces to $T = 2\pi\sqrt{\frac{L}{g}}$.

Example:**Reducing the Swaying of a Skyscraper**

In extreme conditions, skyscrapers can sway up to two meters with a frequency of up to 20.00 Hz due to high winds or seismic activity. Several companies have developed physical pendulums that are placed on the top of the skyscrapers. As the skyscraper sways to the right, the pendulum swings to the left, reducing the sway. Assuming the oscillations have a frequency of 0.50 Hz, design a pendulum that consists of a long beam, of constant density, with a mass of 100 metric tons and a pivot point at one end of the beam. What should be the length of the beam?

**Strategy**

We are asked to find the length of the physical pendulum with a known mass. We first need to find the moment of inertia of the beam. We can then use the equation for the period of a physical pendulum to find the length.

Solution

1. Find the moment of inertia for the CM:
2. Use the parallel axis theorem to find the moment of inertia about the point of rotation:

Equation:

$$I = I_{\text{CM}} + \frac{L^2}{4} M = \frac{1}{12} ML^2 + \frac{1}{4} ML^2 = \frac{1}{3} ML^2.$$

3. The period of a physical pendulum has a period of $T = 2\pi\sqrt{\frac{I}{mgL}}$. Use the moment of inertia to solve for the length L :

Equation:

$$T = 2\pi\sqrt{\frac{I}{MgL}} = 2\pi\sqrt{\frac{\frac{1}{3}ML^2}{MgL}} = 2\pi\sqrt{\frac{L}{3g}};$$

$$L = 3g\left(\frac{T}{2\pi}\right)^2 = 3\left(9.8\frac{\text{m}}{\text{s}^2}\right)\left(\frac{2\text{s}}{2\pi}\right)^2 = 2.98\text{ m}.$$

4. This length L is from the center of mass to the axis of rotation, which is half the length of the pendulum. Therefore the length H of the pendulum is:

Equation:

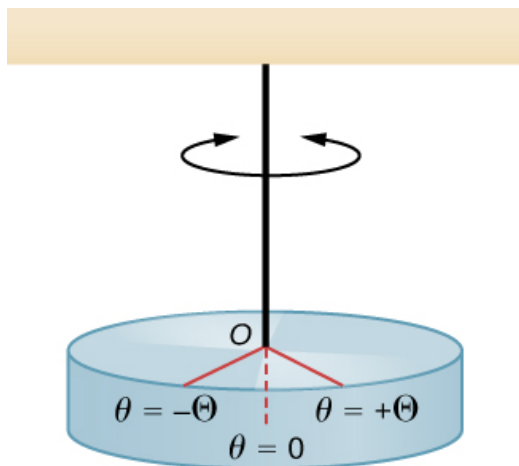
$$H = 2L = 5.96\text{ m}$$

Significance

There are many ways to reduce the oscillations, including modifying the shape of the skyscrapers, using multiple physical pendulums, and using tuned-mass dampers.

Torsional Pendulum

A **torsional pendulum** consists of a rigid body suspended by a light wire or spring ([link](#)). When the body is twisted some small maximum angle (Θ) and released from rest, the body oscillates between ($\theta = +\Theta$) and ($\theta = -\Theta$). The restoring torque is supplied by the shearing of the string or wire.



A torsional pendulum consists of a rigid body suspended by a string or wire. The rigid body oscillates between $\theta = +\Theta$ and $\theta = -\Theta$.

The restoring torque can be modeled as being proportional to the angle:

Equation:

$$\tau = -\kappa\theta.$$

The variable kappa (κ) is known as the torsion constant of the wire or string. The minus sign shows that the restoring torque acts in the opposite direction to increasing angular displacement. The net torque is equal to the moment of inertia times the angular acceleration:

Equation:

$$I \frac{d^2\theta}{dt^2} = -\kappa\theta;$$
$$\frac{d^2\theta}{dt^2} = -\frac{\kappa}{I}\theta.$$

This equation says that the second time derivative of the position (in this case, the angle) equals a negative constant times the position. This looks very similar to the equation of motion for the SHM $\frac{d^2x}{dt^2} = -\frac{k}{m}x$, where the period was found to be $T = 2\pi\sqrt{\frac{m}{k}}$. Therefore, the period of the torsional pendulum can be found using

Note:

Equation:

$$T = 2\pi\sqrt{\frac{I}{\kappa}}.$$

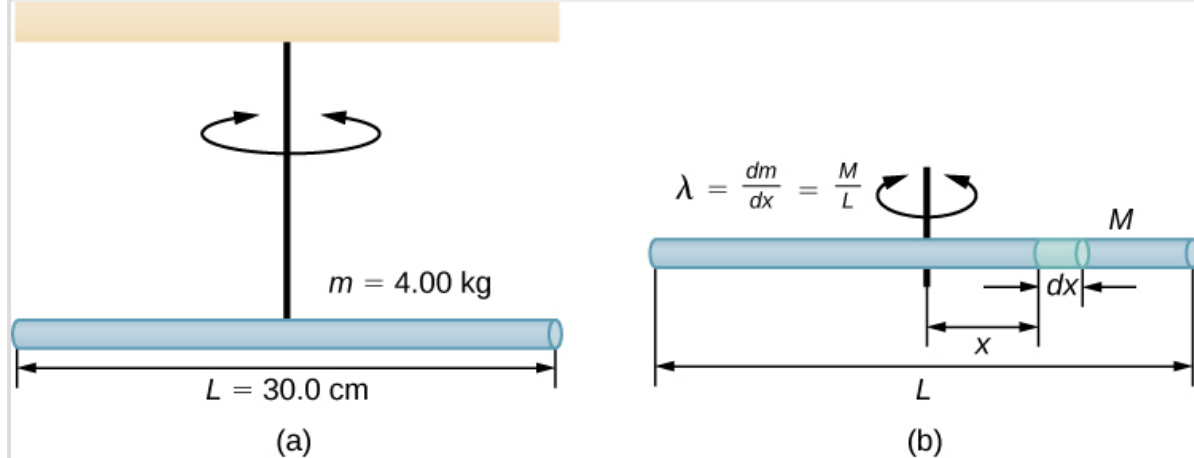
The units for the torsion constant are $[\kappa] = \text{N}\cdot\text{m} = (\text{kg}\frac{\text{m}}{\text{s}^2})\text{m} = \text{kg}\frac{\text{m}^2}{\text{s}^2}$ and the units for the moment of inertia are $[I] = \text{kg}\cdot\text{m}^2$, which show that the unit for the period is the second.

Example:

Measuring the Torsion Constant of a String

A rod has a length of $l = 0.30$ m and a mass of 4.00 kg. A string is attached to the CM of the rod and the system is hung from the ceiling ([link](#)). The rod is displaced 10 degrees from the

equilibrium position and released from rest. The rod oscillates with a period of 0.5 s. What is the torsion constant κ ?



(a) A rod suspended by a string from the ceiling. (b) Finding the rod's moment of inertia.

Strategy

We are asked to find the torsion constant of the string. We first need to find the moment of inertia.

Solution

1. Find the moment of inertia for the CM:

Equation:

$$I_{\text{CM}} = \int x^2 dm = \int_{-L/2}^{+L/2} x^2 \lambda dx = \lambda \left[\frac{x^3}{3} \right]_{-L/2}^{+L/2} = \lambda \frac{2L^3}{24} = \left(\frac{M}{L} \right) \frac{2L^3}{24} = \frac{1}{12} ML^2.$$

2. Calculate the torsion constant using the equation for the period:

Equation:

$$\begin{aligned} T &= 2\pi \sqrt{\frac{I}{\kappa}}; \\ \kappa &= I \left(\frac{2\pi}{T} \right)^2 = \left(\frac{1}{12} ML^2 \right) \left(\frac{2\pi}{T} \right)^2; \\ &= \left(\frac{1}{12} (4.00 \text{ kg}) (0.30 \text{ m})^2 \right) \left(\frac{2\pi}{0.50 \text{ s}} \right)^2 = 4.73 \text{ N} \cdot \text{m}. \end{aligned}$$

Significance

Like the force constant of the system of a block and a spring, the larger the torsion constant, the shorter the period.

Summary

- A mass m suspended by a wire of length L and negligible mass is a simple pendulum and undergoes SHM for amplitudes less than about 15° . The period of a simple pendulum is $T = 2\pi\sqrt{\frac{L}{g}}$, where L is the length of the string and g is the acceleration due to gravity.
- The period of a physical pendulum $T = 2\pi\sqrt{\frac{I}{mgL}}$ can be found if the moment of inertia is known. The length between the point of rotation and the center of mass is L .
- The period of a torsional pendulum $T = 2\pi\sqrt{\frac{I}{\kappa}}$ can be found if the moment of inertia and torsion constant are known.

Conceptual Questions

Exercise:

Problem:

Pendulum clocks are made to run at the correct rate by adjusting the pendulum's length. Suppose you move from one city to another where the acceleration due to gravity is slightly greater, taking your pendulum clock with you, will you have to lengthen or shorten the pendulum to keep the correct time, other factors remaining constant? Explain your answer.

Exercise:

Problem:

A pendulum clock works by measuring the period of a pendulum. In the springtime the clock runs with perfect time, but in the summer and winter the length of the pendulum changes. When most materials are heated, they expand. Does the clock run too fast or too slow in the summer? What about the winter?

Solution:

The period of the pendulum is $T = 2\pi\sqrt{L/g}$. In summer, the length increases, and the period increases. If the period should be one second, but period is longer than one second in the summer, it will oscillate fewer than 60 times a minute and clock will run slow. In the winter it will run fast.

Exercise:

Problem:

With the use of a phase shift, the position of an object may be modeled as a cosine or sine function. If given the option, which function would you choose? Assuming that the phase shift is zero, what are the initial conditions of function; that is, the initial position, velocity, and acceleration, when using a sine function? How about when a cosine function is used?

Problems

Exercise:

Problem: What is the length of a pendulum that has a period of 0.500 s?

Exercise:

Problem:

Some people think a pendulum with a period of 1.00 s can be driven with “mental energy” or psycho kinetically, because its period is the same as an average heartbeat. True or not, what is the length of such a pendulum?

Solution:

24.8 cm

Exercise:

Problem: What is the period of a 1.00-m-long pendulum?

Exercise:

Problem:

How long does it take a child on a swing to complete one swing if her center of gravity is 4.00 m below the pivot?

Solution:

4.01 s

Exercise:

Problem: The pendulum on a cuckoo clock is 5.00-cm long. What is its frequency?

Exercise:

Problem:

Two parakeets sit on a swing with their combined CMs 10.0 cm below the pivot. At what frequency do they swing?

Solution:

1.58 s

Exercise:

Problem:

(a) A pendulum that has a period of 3.00000 s and that is located where the acceleration due to gravity is 9.79 m/s^2 is moved to a location where the acceleration due to gravity is 9.82 m/s^2 . What is its new period? (b) Explain why so many digits are needed in the value for the period, based on the relation between the period and the acceleration due to gravity.

Exercise:**Problem:**

A pendulum with a period of 2.00000 s in one location ($g = 9.80 \text{ m/s}^2$) is moved to a new location where the period is now 1.99796 s. What is the acceleration due to gravity at its new location?

Solution:

9.82002 m/s^2

Exercise:**Problem:**

(a) What is the effect on the period of a pendulum if you double its length? (b) What is the effect on the period of a pendulum if you decrease its length by 5.00%?

Glossary

physical pendulum

any extended object that swings like a pendulum

simple pendulum

point mass, called a pendulum bob, attached to a near massless string

torsional pendulum

any suspended object that oscillates by twisting its suspension

Damped Oscillations

By the end of this section, you will be able to:

- Describe the motion of damped harmonic motion
- Write the equations of motion for damped harmonic oscillations
- Describe the motion of driven, or forced, damped harmonic motion
- Write the equations of motion for forced, damped harmonic motion

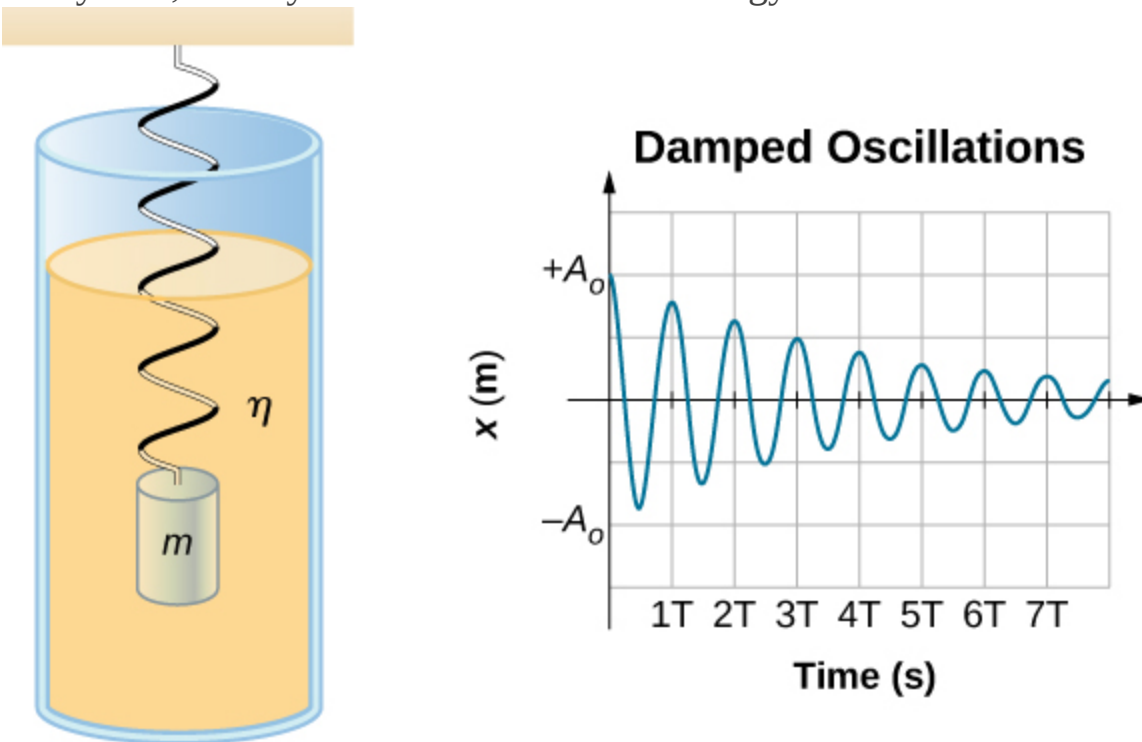
In the real world, oscillations seldom follow true SHM. Friction of some sort usually acts to dampen the motion so it dies away, or needs more force to continue. In this section, we examine some examples of damped harmonic motion and see how to modify the equations of motion to describe this more general case.

A guitar string stops oscillating a few seconds after being plucked. To keep swinging on a playground swing, you must keep pushing ([link](#)). Although we can often make friction and other nonconservative forces small or negligible, completely undamped motion is rare. In fact, we may even want to damp oscillations, such as with car shock absorbers.



To counteract dampening forces, you need to keep pumping a swing. (credit: Bob Mical)

[\[link\]](#) shows a mass m attached to a spring with a force constant k . The mass is raised to a position A_0 , the initial amplitude, and then released. The mass oscillates around the equilibrium position in a fluid with viscosity but the amplitude decreases for each oscillation. For a system that has a small amount of damping, the period and frequency are constant and are nearly the same as for SHM, but the amplitude gradually decreases as shown. This occurs because the non-conservative damping force removes energy from the system, usually in the form of thermal energy.



For a mass on a spring oscillating in a viscous fluid, the period remains constant, but the amplitudes of the oscillations decrease due to the damping caused by the fluid.

Consider the forces acting on the mass. Note that the only contribution of the weight is to change the equilibrium position, as discussed earlier in the chapter. Therefore, the net force is equal to the force of the spring and the damping force (F_D). If the magnitude of the velocity is small, meaning the

mass oscillates slowly, the damping force is proportional to the velocity and acts against the direction of motion ($F_D = -bv$). The net force on the mass is therefore

Equation:

$$ma = -bv - kx.$$

Writing this as a differential equation in x , we obtain

Note:

Equation:

$$m \frac{d^2 x}{dt^2} + b \frac{dx}{dt} + kx = 0.$$

To determine the solution to this equation, consider the plot of position versus time shown in [\[link\]](#). The curve resembles a cosine curve oscillating in the envelope of an exponential function $A_0 e^{-\alpha t}$ where $\alpha = \frac{b}{2m}$. The solution is

Note:

Equation:

$$x(t) = A_0 e^{-\frac{b}{2m}t} \cos(\omega t + \phi).$$

It is left as an exercise to prove that this is, in fact, the solution. To prove that it is the right solution, take the first and second derivatives with respect

to time and substitute them into [\[link\]](#). It is found that [\[link\]](#) is the solution if

Equation:

$$\omega = \sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2}.$$

Recall that the angular frequency of a mass undergoing SHM is equal to the square root of the force constant divided by the mass. This is often referred to as the **natural angular frequency**, which is represented as

Note:

Equation:

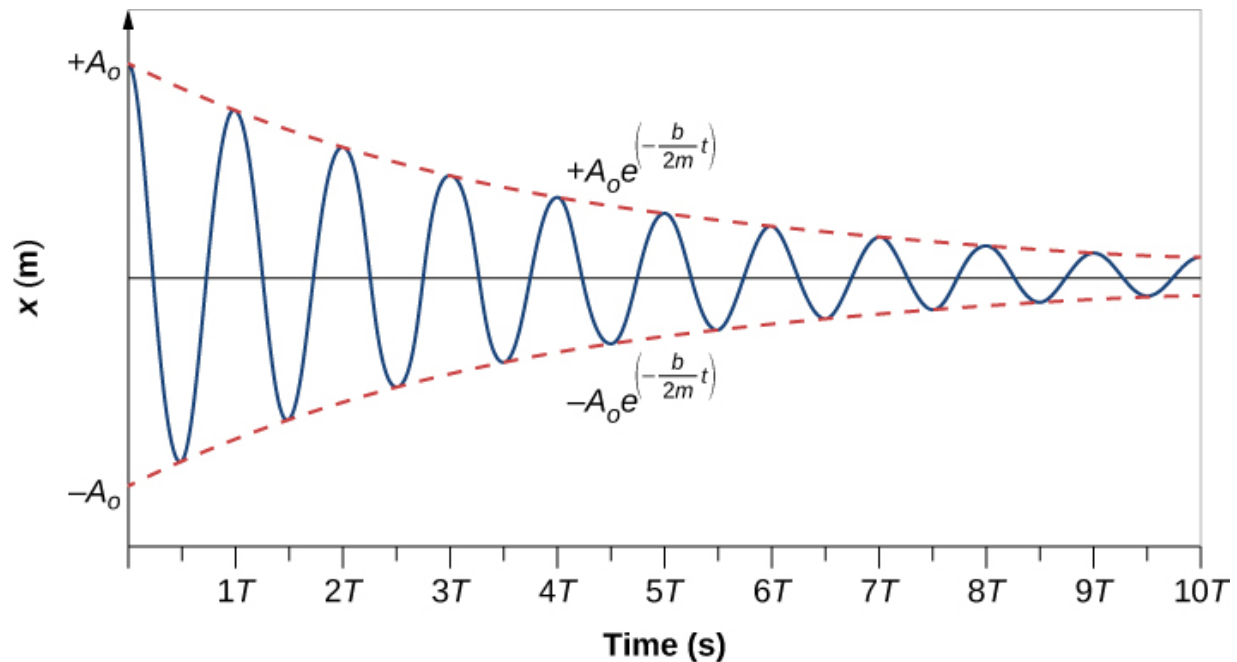
$$\omega_0 = \sqrt{\frac{k}{m}}.$$

The angular frequency for damped harmonic motion becomes

Note:

Equation:

$$\omega = \sqrt{\omega_0^2 - \left(\frac{b}{2m}\right)^2}.$$



Position versus time for the mass oscillating on a spring in a viscous fluid. Notice that the curve appears to be a cosine function inside an exponential envelope.

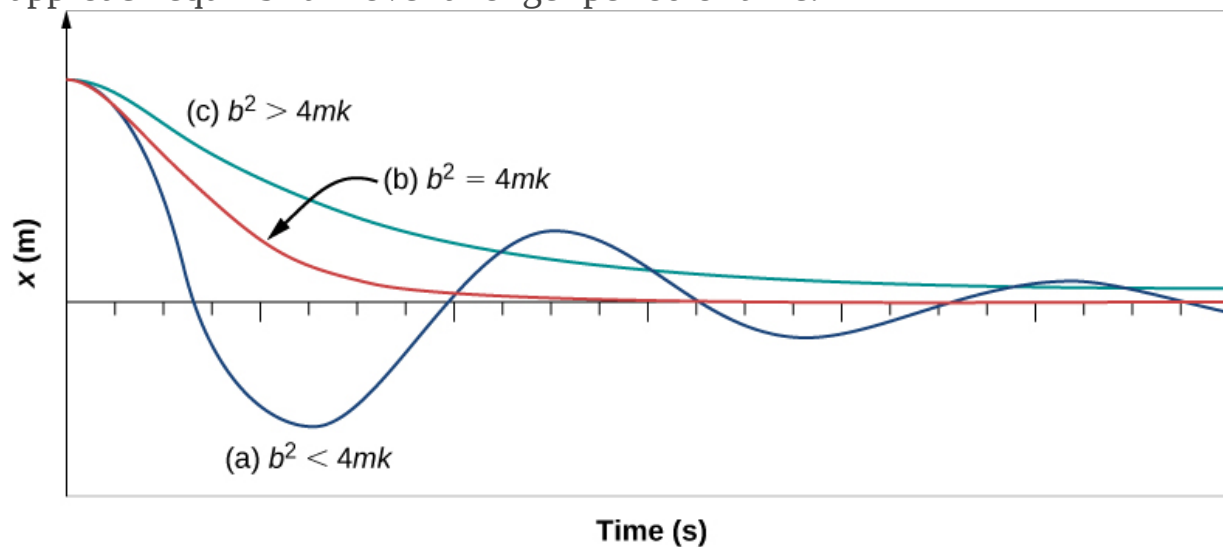
Recall that when we began this description of damped harmonic motion, we stated that the damping must be small. Two questions come to mind. Why must the damping be small? And how small is small? If you gradually *increase* the amount of damping in a system, the period and frequency begin to be affected, because damping opposes and hence slows the back and forth motion. (The net force is smaller in both directions.) If there is very large damping, the system does not even oscillate—it slowly moves toward equilibrium. The angular frequency is equal to

Equation:

$$\omega = \sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2}.$$

As b increases, $\frac{k}{m} - \left(\frac{b}{2m}\right)^2$ becomes smaller and eventually reaches zero when $b = \sqrt{4mk}$. If b becomes any larger, $\frac{k}{m} - \left(\frac{b}{2m}\right)^2$ becomes a negative number and $\sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2}$ is a complex number.

[\[link\]](#) shows the displacement of a harmonic oscillator for different amounts of damping. When the damping constant is small, $b < \sqrt{4mk}$, the system oscillates while the amplitude of the motion decays exponentially. This system is said to be **underdamped**, as in curve (a). Many systems are underdamped, and oscillate while the amplitude decreases exponentially, such as the mass oscillating on a spring. The damping may be quite small, but eventually the mass comes to rest. If the damping constant is $b = \sqrt{4mk}$, the system is said to be **critically damped**, as in curve (b). An example of a critically damped system is the shock absorbers in a car. It is advantageous to have the oscillations decay as fast as possible. Here, the system does not oscillate, but asymptotically approaches the equilibrium condition as quickly as possible. Curve (c) in [\[link\]](#) represents an **overdamped** system where $b > \sqrt{4mk}$. An overdamped system will approach equilibrium over a longer period of time.



The position versus time for three systems consisting of a mass and a spring in a viscous fluid. (a) If the damping is small ($b < \sqrt{4mk}$), the mass oscillates, slowly losing amplitude as the energy is dissipated by the non-conservative force(s). The limiting case is (b) where the

damping is $(b = \sqrt{4mk})$. (c) If the damping is very large $(b > \sqrt{4mk})$, the mass does not oscillate when displaced, but attempts to return to the equilibrium position.

Critical damping is often desired, because such a system returns to equilibrium rapidly and remains at equilibrium as well. In addition, a constant force applied to a critically damped system moves the system to a new equilibrium position in the shortest time possible without overshooting or oscillating about the new position.

Note:

Exercise:

Problem:

Check Your Understanding Why are completely undamped harmonic oscillators so rare?

Solution:

Friction often comes into play whenever an object is moving. Friction causes damping in a harmonic oscillator.

Summary

- Damped harmonic oscillators have non-conservative forces that dissipate their energy.
- Critical damping returns the system to equilibrium as fast as possible without overshooting.

- An underdamped system will oscillate through the equilibrium position.
- An overdamped system moves more slowly toward equilibrium than one that is critically damped.

Conceptual Questions

Exercise:

Problem:

Give an example of a damped harmonic oscillator. (They are more common than undamped or simple harmonic oscillators.)

Solution:

A car shock absorber.

Exercise:

Problem:

How would a car bounce after a bump under each of these conditions?

(a) overdamping

(b) underdamping

(c) critical damping

Exercise:

Problem:

Most harmonic oscillators are damped and, if undriven, eventually come to a stop. Why?

Solution:

The second law of thermodynamics states that perpetual motion machines are impossible. Eventually the ordered motion of the system

decreases and returns to equilibrium.

Problems

Exercise:

Problem:

The amplitude of a lightly damped oscillator decreases by 3.0% during each cycle. What percentage of the mechanical energy of the oscillator is lost in each cycle?

Solution:

6%

Glossary

critically damped

condition in which the damping of an oscillator causes it to return as quickly as possible to its equilibrium position without oscillating back and forth about this position

natural angular frequency

angular frequency of a system oscillating in SHM

overdamped

condition in which damping of an oscillator causes it to return to equilibrium without oscillating; oscillator moves more slowly toward equilibrium than in the critically damped system

underdamped

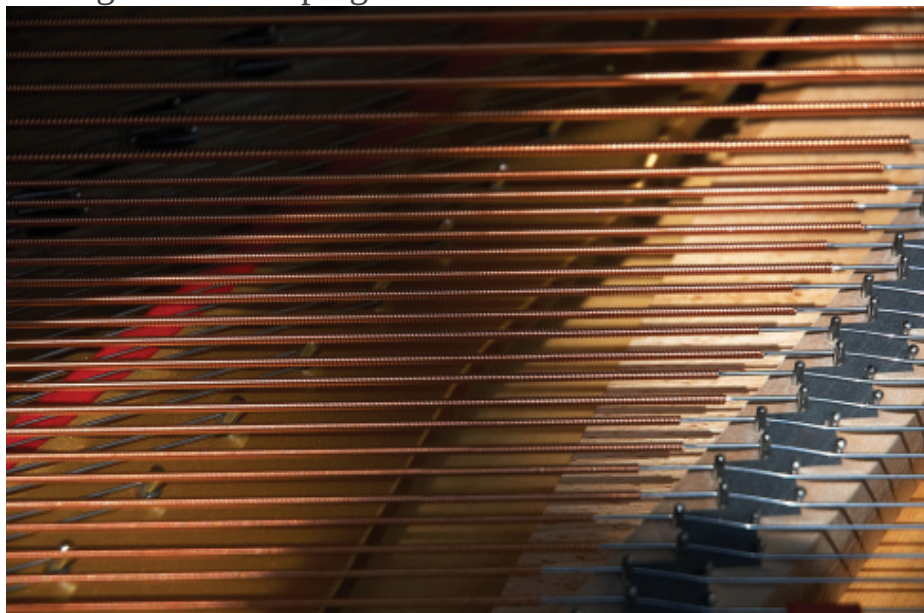
condition in which damping of an oscillator causes the amplitude of oscillations of a damped harmonic oscillator to decrease over time, eventually approaching zero

Forced Oscillations

By the end of this section, you will be able to:

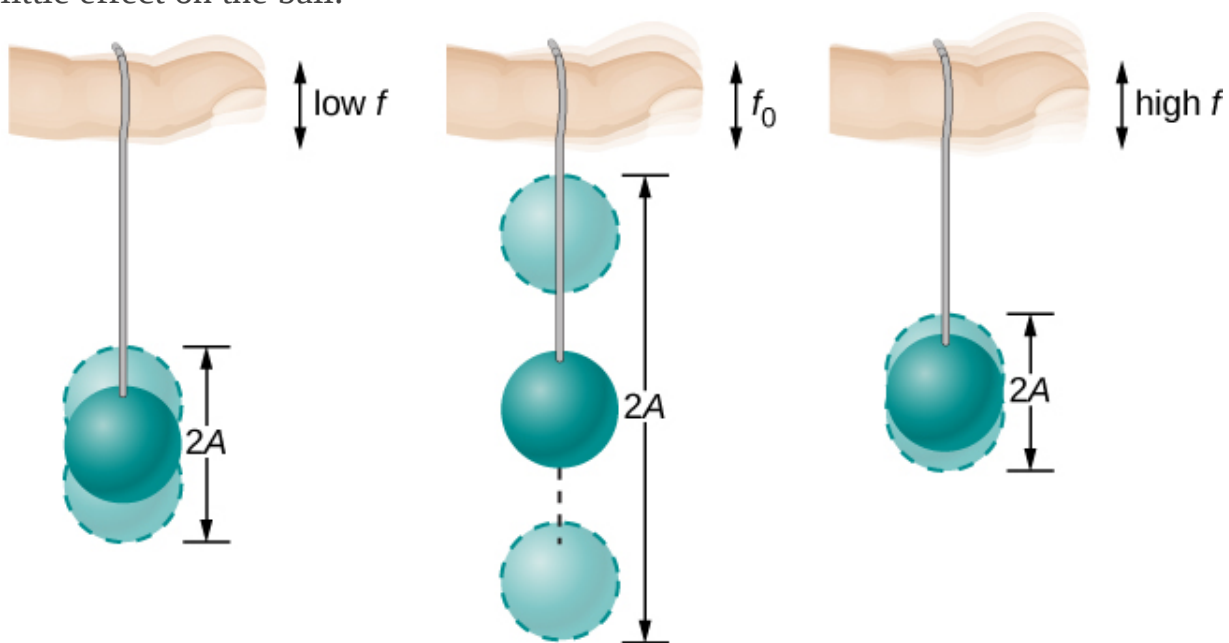
- Define forced oscillations
- List the equations of motion associated with forced oscillations
- Explain the concept of resonance and its impact on the amplitude of an oscillator
- List the characteristics of a system oscillating in resonance

Sit in front of a piano sometime and sing a loud brief note at it with the dampers off its strings ([link](#)). It will sing the same note back at you—the strings, having the same frequencies as your voice, are resonating in response to the forces from the sound waves that you sent to them. This is a good example of the fact that objects—in this case, piano strings—can be forced to oscillate, and oscillate most easily at their natural frequency. In this section, we briefly explore applying a periodic driving force acting on a simple harmonic oscillator. The driving force puts energy into the system at a certain frequency, not necessarily the same as the natural frequency of the system. Recall that the natural frequency is the frequency at which a system would oscillate if there were no driving and no damping force.



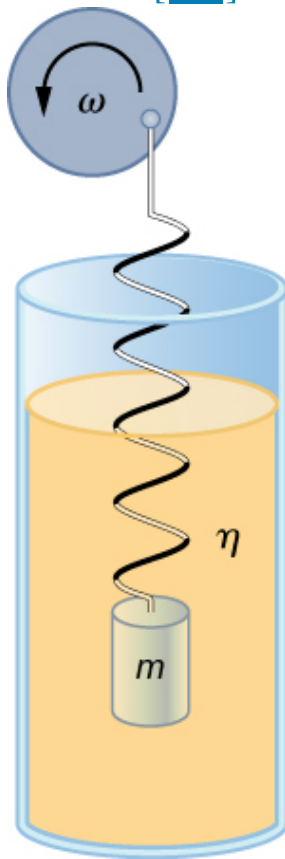
You can cause the strings in a piano to vibrate simply by producing sound waves from your voice. (credit: Matt Billings)

Most of us have played with toys involving an object supported on an elastic band, something like the paddle ball suspended from a finger in [\[link\]](#). Imagine the finger in the figure is your finger. At first, you hold your finger steady, and the ball bounces up and down with a small amount of damping. If you move your finger up and down slowly, the ball follows along without bouncing much on its own. As you increase the frequency at which you move your finger up and down, the ball responds by oscillating with increasing amplitude. When you drive the ball at its natural frequency, the ball's oscillations increase in amplitude with each oscillation for as long as you drive it. The phenomenon of driving a system with a frequency equal to its natural frequency is called **resonance**. A system being driven at its natural frequency is said to *resonate*. As the driving frequency gets progressively higher than the resonant or natural frequency, the amplitude of the oscillations becomes smaller until the oscillations nearly disappear, and your finger simply moves up and down with little effect on the ball.



The paddle ball on its rubber band moves in response to the finger supporting it. If the finger moves with the natural frequency f_0 of the ball on the rubber band, then a resonance is achieved, and the amplitude of the ball's oscillations increases dramatically. At higher and lower driving frequencies, energy is transferred to the ball less efficiently, and it responds with lower-amplitude oscillations.

Consider a simple experiment. Attach a mass m to a spring in a viscous fluid, similar to the apparatus discussed in the damped harmonic oscillator. This time, instead of fixing the free end of the spring, attach the free end to a disk that is driven by a variable-speed motor. The motor turns with an angular driving frequency of ω . The rotating disk provides energy to the system by the work done by the driving force ($F_d = F_0 \sin(\omega t)$). The experimental apparatus is shown in [\[link\]](#).



Forced,
damped
harmonic
motion
produced by
driving a
spring and
mass with a
disk driven

by a
variable-
speed motor.

Using Newton's second law ($\vec{\mathbf{F}}_{\text{net}} = m\vec{\mathbf{a}}$), we can analyze the motion of the mass. The resulting equation is similar to the force equation for the damped harmonic oscillator, with the addition of the driving force:

Note:

Equation:

$$-kx - b\frac{dx}{dt} + F_0\sin(\omega t) = m\frac{d^2x}{dt^2}.$$

When an oscillator is forced with a periodic driving force, the motion may seem chaotic. The motions of the oscillator is known as transients. After the transients die out, the oscillator reaches a steady state, where the motion is periodic. After some time, the steady state solution to this differential equation is

Note:

Equation:

$$x(t) = A\cos(\omega t + \phi).$$

Once again, it is left as an exercise to prove that this equation is a solution. Taking the first and second time derivative of $x(t)$ and substituting them into the force equation shows that $x(t) = A\sin(\omega t + \phi)$ is a solution as long as the amplitude is equal to

Note:

Equation:

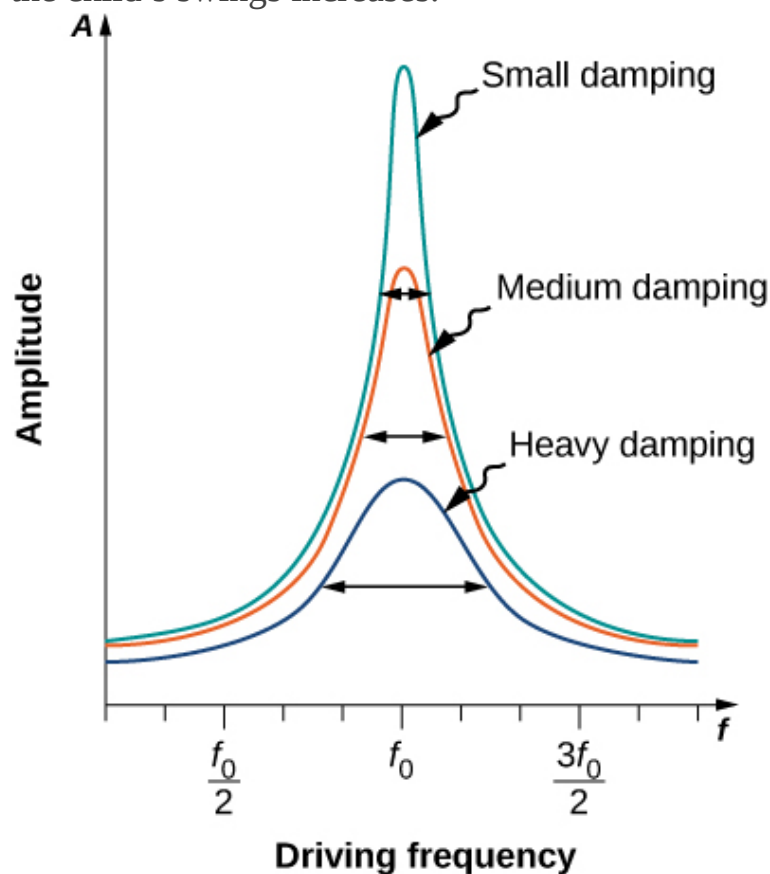
$$A = \frac{F_0}{\sqrt{m^2(\omega^2 - \omega_0^2)^2 + b^2\omega^2}}$$

where $\omega_0 = \sqrt{\frac{k}{m}}$ is the angular frequency of the driving force. Recall that the angular frequency, and therefore the frequency, of the motor can be adjusted. Looking at the denominator of the equation for the amplitude, when the driving frequency is much smaller, or much larger, than the natural frequency, the square of the difference of the two angular frequencies $(\omega^2 - \omega_0^2)^2$ is positive and large, making the denominator large, and the result is a small amplitude for the oscillations of the mass. As the frequency of the driving force approaches the natural frequency of the system, the denominator becomes small and the amplitude of the oscillations becomes large. The maximum amplitude results when the frequency of the driving force equals the natural frequency of the system $\left(A_{\max} = \frac{F_0}{b\omega}\right)$.

[\[link\]](#) shows a graph of the amplitude of a damped harmonic oscillator as a function of the frequency of the periodic force driving it. Each of the three curves on the graph represents a different amount of damping. All three curves peak at the point where the frequency of the driving force equals the natural frequency of the harmonic oscillator. The highest peak, or greatest response, is for the least amount of damping, because less energy is removed by the damping force. Note that since the amplitude grows as the damping decreases, taking this to the limit where there is no damping ($b = 0$), the amplitude becomes infinite.

Note that a small-amplitude driving force can produce a large-amplitude response. This phenomenon is known as resonance. A common example of resonance is a parent pushing a small child on a swing. When the child wants to go higher, the parent does not move back and then, getting a running start, slam into the child, applying a great force in a short interval. Instead, the parent

applies small pushes to the child at just the right frequency, and the amplitude of the child's swings increases.

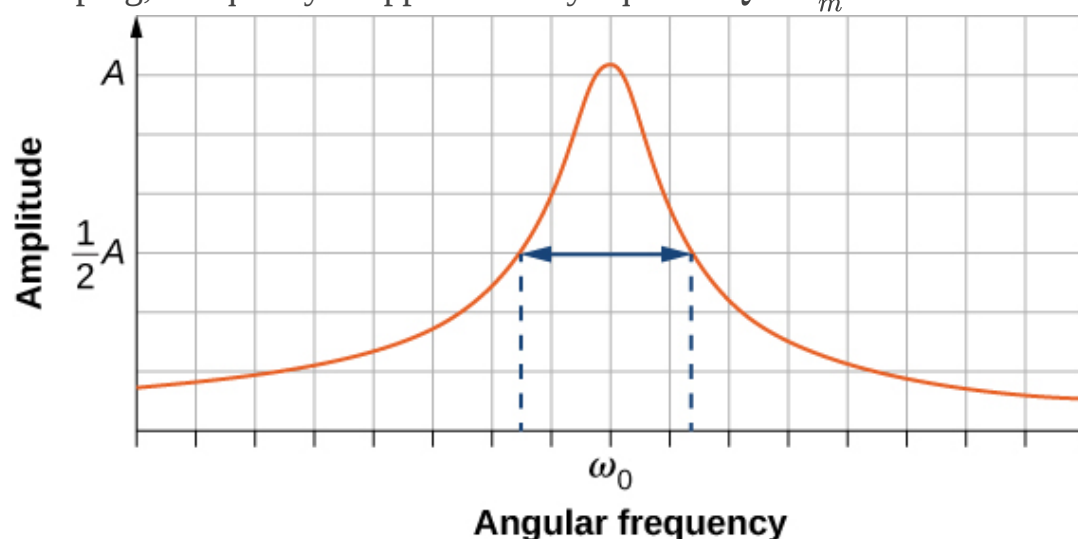


Amplitude of a harmonic oscillator as a function of the frequency of the driving force. The curves represent the same oscillator with the same natural frequency but with different amounts of damping.

Resonance occurs when the driving frequency equals the natural frequency, and the greatest response is for the least amount of damping. The narrowest response is also for the least damping.

It is interesting to note that the widths of the resonance curves shown in [\[link\]](#) depend on damping: the less the damping, the narrower the resonance. The consequence is that if you want a driven oscillator to resonate at a very specific

frequency, you need as little damping as possible. For instance, a radio has a circuit that is used to choose a particular radio station. In this case, the forced damped oscillator consists of a resistor, capacitor, and inductor, which will be discussed later in this course. The circuit is “tuned” to pick a particular radio station. Here it is desirable to have the resonance curve be very narrow, to pick out the exact frequency of the radio station chosen. The narrowness of the graph, and the ability to pick out a certain frequency, is known as the quality of the system. The quality is defined as the spread of the angular frequency, or equivalently, the spread in the frequency, at half the maximum amplitude, divided by the natural frequency ($Q = \frac{\omega_0}{\Delta\omega}$) as shown in [\[link\]](#). For a small damping, the quality is approximately equal to $Q \approx \frac{2b}{m}$.



The quality of a system is defined as the spread in the frequencies at half the amplitude divided by the natural frequency.

These features of driven harmonic oscillators apply to a huge variety of systems. For instance, magnetic resonance imaging (MRI) is a widely used medical diagnostic tool in which atomic nuclei (mostly hydrogen nuclei or protons) are made to resonate by incoming radio waves (on the order of 100 MHz). In all of these cases, the efficiency of energy transfer from the driving force into the oscillator is best at resonance. [\[link\]](#) shows the London Millennium Footbridge that allows pedestrians to cross the River Thames in London. This bridge was nicknamed “Wobbly Bridge” when pedestrians experienced swaying motion

while crossing it. The bridge was closed for roughly two years to get rid of this motion.



Initially when people crossed the London Millennium Footbridge, they experienced a swaying motion. People continuing to cross reinforced the oscillation's amplitude, thereby increasing the problematic swaying.
(credit: Adrian Pingstone/Wikimedia Commons)

Note:

Exercise:

Problem:

Check Your Understanding A famous magic trick involves a performer singing a note toward a crystal glass until the glass shatters. Explain why the trick works in terms of resonance and natural frequency.

Solution:

The performer must be singing a note that corresponds to the natural frequency of the glass. As the sound wave is directed at the glass, the glass responds by resonating at the same frequency as the sound wave. With enough energy introduced into the system, the glass begins to vibrate and eventually shatters.

Summary

- A system's natural frequency is the frequency at which the system oscillates if not affected by driving or damping forces.
- A periodic force driving a harmonic oscillator at its natural frequency produces resonance. The system is said to resonate.
- The less damping a system has, the higher the amplitude of the forced oscillations near resonance. The more damping a system has, the broader response it has to varying driving frequencies.

Key Equations

Relationship between frequency and period	$f = \frac{1}{T}$
Position in SHM with $\phi = 0.00$	$x(t) = A \cos(\omega t)$
General position in SHM	$x(t) = A \cos(\omega t + \phi)$
General velocity in SHM	$v(t) = -A\omega \sin(\omega t + \phi)$
General acceleration in SHM	$a(t) = -A\omega^2 \cos(\omega t + \phi)$

Maximum displacement (amplitude) of SHM	$x_{\max} = A$
Maximum velocity of SHM	$ v_{\max} = A\omega$
Maximum acceleration of SHM	$ a_{\max} = A\omega^2$
Angular frequency of a mass-spring system in SHM	$\omega = \sqrt{\frac{k}{m}}$
Period of a mass-spring system in SHM	$T = 2\pi\sqrt{\frac{m}{k}}$
Frequency of a mass-spring system in SHM	$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$
Energy in a mass-spring system in SHM	$E_{\text{Total}} = \frac{1}{2}kx^2 + \frac{1}{2}mv^2 = \frac{1}{2}kA^2$
The velocity of the mass in a spring-mass system in SHM	$v = \pm\sqrt{\frac{k}{m}(A^2 - x^2)}$
The x-component of the radius of a rotating disk	$x(t) = A\cos(\omega t + \phi)$
The x-component of the velocity of the edge of a rotating disk	$v(t) = -v_{\max}\sin(\omega t + \phi)$
The x-component of the acceleration of the edge of a rotating disk	$a(t) = -a_{\max}\cos(\omega t + \phi)$
Force equation for a simple pendulum	$\frac{d^2\theta}{dt^2} = -\frac{g}{L}\theta$
Angular frequency for a simple pendulum	$\omega = \sqrt{\frac{g}{L}}$

Period of a simple pendulum	$T = 2\pi\sqrt{\frac{L}{g}}$
Angular frequency of a physical pendulum	$\omega = \sqrt{\frac{mgL}{I}}$
Period of a physical pendulum	$T = 2\pi\sqrt{\frac{I}{mgL}}$
Period of a torsional pendulum	$T = 2\pi\sqrt{\frac{I}{\kappa}}$
Newton's second law for harmonic motion	$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = 0$
Solution for underdamped harmonic motion	$x(t) = A_0e^{-\frac{b}{2m}t}\cos(\omega t + \phi)$
Natural angular frequency of a mass-spring system	$\omega_0 = \sqrt{\frac{k}{m}}$
Angular frequency of underdamped harmonic motion	$\omega = \sqrt{\omega_0^2 - \left(\frac{b}{2m}\right)^2}$
Newton's second law for forced, damped oscillation	$-kx - b\frac{dx}{dt} + F_o\sin(\omega t) = m\frac{d^2x}{dt^2}$
Solution to Newton's second law for forced, damped oscillations	$x(t) = A\cos(\omega t + \phi)$
Amplitude of system undergoing forced, damped oscillations	$A = \frac{F_o}{\sqrt{m^2(\omega^2 - \omega_0^2)^2 + b^2\omega^2}}$

Conceptual Questions

Exercise:

Problem:

Why are soldiers in general ordered to “route step” (walk out of step) across a bridge?

Exercise:**Problem:**

Do you think there is any harmonic motion in the physical world that is not damped harmonic motion? Try to make a list of five examples of undamped harmonic motion and damped harmonic motion. Which list was easier to make?

Solution:

All harmonic motion is damped harmonic motion, but the damping may be negligible. This is due to friction and drag forces. It is easy to come up with five examples of damped motion: (1) A mass oscillating on a hanging on a spring (it eventually comes to rest). (2) Shock absorbers in a car (thankfully they also come to rest). (3) A pendulum in a grandfather clock (weights are added to add energy to the oscillations). (4) A child on a swing (eventually comes to rest unless energy is added by pushing the child). (5) A marble rolling in a bowl (eventually comes to rest). As for the undamped motion, even a mass on a spring in a vacuum will eventually come to rest due to internal forces in the spring. Damping may be negligible, but cannot be eliminated.

Exercise:**Problem:**

Some engineers use sound to diagnose performance problems with car engines. Occasionally, a part of the engine is designed that resonates at the frequency of the engine. The unwanted oscillations can cause noise that irritates the driver or could lead to the part failing prematurely. In one case, a part was located that had a length L made of a material with a mass M . What can be done to correct this problem?

Problems

Exercise:**Problem:**

How much energy must the shock absorbers of a 1200-kg car dissipate in order to damp a bounce that initially has a velocity of 0.800 m/s at the equilibrium position? Assume the car returns to its original vertical position.

Exercise:**Problem:**

If a car has a suspension system with a force constant of $5.00 \times 10^4 \text{ N/m}$, how much energy must the car's shocks remove to dampen an oscillation starting with a maximum displacement of 0.0750 m?

Solution:

141 J

Exercise:**Problem:**

(a) How much will a spring that has a force constant of 40.0 N/m be stretched by an object with a mass of 0.500 kg when hung motionless from the spring? (b) Calculate the decrease in gravitational potential energy of the 0.500-kg object when it descends this distance. (c) Part of this gravitational energy goes into the spring. Calculate the energy stored in the spring by this stretch, and compare it with the gravitational potential energy. Explain where the rest of the energy might go.

Exercise:

Problem:

Suppose you have a 0.750-kg object on a horizontal surface connected to a spring that has a force constant of 150 N/m. There is simple friction between the object and surface with a static coefficient of friction $\mu_s = 0.100$. (a) How far can the spring be stretched without moving the mass? (b) If the object is set into oscillation with an amplitude twice the distance found in part (a), and the kinetic coefficient of friction is $\mu_k = 0.0850$, what total distance does it travel before stopping? Assume it starts at the maximum amplitude.

Solution:

a. 4.90×10^{-3} m; b. 1.15×10^{-2} m

Additional Problems**Exercise:****Problem:**

Suppose you attach an object with mass m to a vertical spring originally at rest, and let it bounce up and down. You release the object from rest at the spring's original rest length, the length of the spring in equilibrium, without the mass attached. The amplitude of the motion is the distance between the equilibrium position of the spring without the mass attached and the equilibrium position of the spring with the mass attached. (a) Show that the spring exerts an upward force of $2.00mg$ on the object at its lowest point. (b) If the spring has a force constant of 10.0 N/m, is hung horizontally, and the position of the free end of the spring is marked as $y = 0.00$ m, where is the new equilibrium position if a 0.25-kg-mass object is hung from the spring? (c) If the spring has a force constant of 10.0 N/m and a 0.25-kg-mass object is set in motion as described, find the amplitude of the oscillations. (d) Find the maximum velocity.

Exercise:

Problem:

A diver on a diving board is undergoing SHM. Her mass is 55.0 kg and the period of her motion is 0.800 s. The next diver is a male whose period of simple harmonic oscillation is 1.05 s. What is his mass if the mass of the board is negligible?

Solution:

94.7 kg

Exercise:**Problem:**

Suppose a diving board with no one on it bounces up and down in a SHM with a frequency of 4.00 Hz. The board has an effective mass of 10.0 kg. What is the frequency of the SHM of a 75.0-kg diver on the board?

Exercise:**Problem:**

The device pictured in the following figure entertains infants while keeping them from wandering. The child bounces in a harness suspended from a door frame by a spring. (a) If the spring stretches 0.250 m while supporting an 8.0-kg child, what is its force constant? (b) What is the time for one complete bounce of this child? (c) What is the child's maximum velocity if the amplitude of her bounce is 0.200 m?



(credit: Lisa Doehnert)

Solution:

a. 314 N/m; b. 1.00 s; c. 1.25 m/s

Exercise:

Problem:

A mass is placed on a frictionless, horizontal table. A spring ($k = 100 \text{ N/m}$), which can be stretched or compressed, is placed on the table. A 5.00-kg mass is attached to one end of the spring, the other end is anchored to the wall. The equilibrium position is marked at zero. A student moves the mass out to $x = 4.00 \text{ cm}$ and releases it from rest. The mass oscillates in SHM. (a) Determine the equations of motion. (b) Find the position, velocity, and acceleration of the mass at time $t = 3.00 \text{ s}$.

Exercise:

Problem:

Find the ratio of the new/old periods of a pendulum if the pendulum were transported from Earth to the Moon, where the acceleration due to gravity is 1.63 m/s^2 .

Solution:

ratio of 2.45

Exercise:**Problem:**

At what rate will a pendulum clock run on the Moon, where the acceleration due to gravity is 1.63 m/s^2 , if it keeps time accurately on Earth? That is, find the time (in hours) it takes the clock's hour hand to make one revolution on the Moon.

Exercise:**Problem:**

If a pendulum-driven clock gains 5.00 s/day , what fractional change in pendulum length must be made for it to keep perfect time?

Solution:

The length must increase by 0.0116% .

Exercise:**Problem:**

A 2.00-kg object hangs, at rest, on a 1.00-m -long string attached to the ceiling. A 100-g mass is fired with a speed of 20 m/s at the 2.00-kg mass, and the 100.00-g mass collides perfectly elastically with the 2.00-kg mass. Write an equation for the motion of the hanging mass after the collision. Assume air resistance is negligible.

Exercise:

Problem:

A 2.00-kg object hangs, at rest, on a 1.00-m-long string attached to the ceiling. A 100-g object is fired with a speed of 20 m/s at the 2.00-kg object, and the two objects collide and stick together in a totally inelastic collision. Write an equation for the motion of the system after the collision. Assume air resistance is negligible.

Solution:

$$\theta = (0.31 \text{ rad})\sin(3.13 \text{ s}^{-1}t)$$

Exercise:**Problem:**

Assume that a pendulum used to drive a grandfather clock has a length $L_0 = 1.00 \text{ m}$ and a mass M at temperature $T = 20.00^\circ\text{C}$. It can be modeled as a physical pendulum as a rod oscillating around one end. By what percentage will the period change if the temperature increases by 10°C ? Assume the length of the rod changes linearly with temperature, where $L = L_0(1 + \alpha\Delta T)$ and the rod is made of brass ($\alpha = 18 \times 10^{-6}^\circ\text{C}^{-1}$).

Exercise:**Problem:**

A 2.00-kg block lies at rest on a frictionless table. A spring, with a spring constant of 100 N/m is attached to the wall and to the block. A second block of 0.50 kg is placed on top of the first block. The 2.00-kg block is gently pulled to a position $x = +A$ and released from rest. There is a coefficient of friction of 0.45 between the two blocks. (a) What is the period of the oscillations? (b) What is the largest amplitude of motion that will allow the blocks to oscillate without the 0.50-kg block sliding off?

Solution:

a. 0.99 s; b. 0.11 m

Challenge Problems

Exercise:

Problem:

A suspension bridge oscillates with an effective force constant of $1.00 \times 10^8 \text{ N/m}$. (a) How much energy is needed to make it oscillate with an amplitude of 0.100 m? (b) If soldiers march across the bridge with a cadence equal to the bridge's natural frequency and impart $1.00 \times 10^4 \text{ J}$ of energy each second, how long does it take for the bridge's oscillations to go from 0.100 m to 0.500 m amplitude.

Exercise:

Problem:

Near the top of the Citigroup Center building in New York City, there is an object with mass of $4.00 \times 10^5 \text{ kg}$ on springs that have adjustable force constants. Its function is to dampen wind-driven oscillations of the building by oscillating at the same frequency as the building is being driven—the driving force is transferred to the object, which oscillates instead of the entire building. (a) What effective force constant should the springs have to make the object oscillate with a period of 2.00 s? (b) What energy is stored in the springs for a 2.00-m displacement from equilibrium?

Solution:

a. $3.95 \times 10^6 \text{ N/m}$; b. $7.90 \times 10^6 \text{ J}$

Exercise:

Problem:

Parcels of air (small volumes of air) in a stable atmosphere (where the temperature increases with height) can oscillate up and down, due to the restoring force provided by the buoyancy of the air parcel. The frequency of the oscillations are a measure of the stability of the atmosphere.

Assuming that the acceleration of an air parcel can be modeled as

$\frac{\partial^2 z'}{\partial t^2} = \frac{g}{\rho_0} \frac{\partial \rho(z)}{\partial z} z'$, prove that $z' = z_0' e^{t\sqrt{-N^2}}$ is a solution, where N is

known as the Brunt-Väisälä frequency. Note that in a stable atmosphere, the density decreases with height and parcel oscillates up and down.

Exercise:**Problem:**

Consider the van der Waals potential $U(r) = U_o \left[\left(\frac{R_o}{r} \right)^{12} - 2 \left(\frac{R_o}{r} \right)^6 \right]$, used to model the potential energy function of two molecules, where the minimum potential is at $r = R_o$. Find the force as a function of r . Consider a small displacement $r = R_o + r'$ and use the binomial theorem:

$$(1 + x)^n = 1 + nx + \frac{n(n-1)}{2!} x^2 + \frac{n(n-1)(n-2)}{3!} x^3 + \dots,$$

to show that the force does approximate a Hooke's law force.

Solution:

$$F \approx -\text{constant } r'$$

Exercise:**Problem:**

Suppose the length of a clock's pendulum is changed by 1.000%, exactly at noon one day. What time will the clock read 24.00 hours later, assuming it the pendulum has kept perfect time before the change? Note that there are two answers, and perform the calculation to four-digit precision.

Exercise:**Problem:**

(a) The springs of a pickup truck act like a single spring with a force constant of $1.30 \times 10^5 \text{ N/m}$. By how much will the truck be depressed by its maximum load of 1000 kg? (b) If the pickup truck has four identical springs, what is the force constant of each?

Solution:

a. 7.54 cm; b. $3.25 \times 10^4 \text{ N/m}$

Glossary

resonance

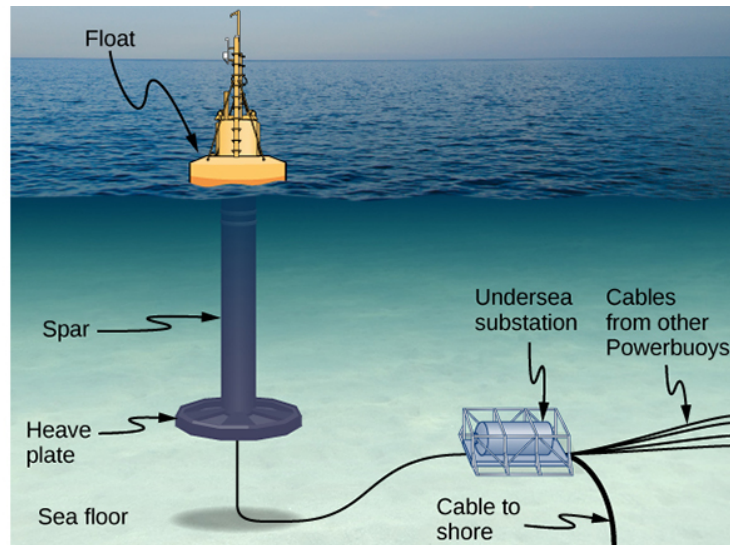
large amplitude oscillations in a system produced by a small amplitude driving force, which has a frequency equal to the natural frequency

Introduction

class="introduction"

From the world of renewable energy sources comes the electric power-generating buoy. Although there are many versions, this one converts the up-and-down motion, as well as side-to-side motion, of the buoy into rotational motion in order to turn an electric generator, which stores the

energy in
batteries.



In this chapter, we study the physics of wave motion. We concentrate on mechanical waves, which are disturbances that move through a medium such as air or water. Like simple harmonic motion studied in the preceding chapter, the energy transferred through the medium is proportional to the amplitude squared. Surface water waves in the ocean are transverse waves in which the energy of the wave travels horizontally while the water oscillates up and down due to some restoring force. In the picture above, a buoy is used to convert the awesome power of ocean waves into electricity. The up-and-down motion of the buoy generated as the waves pass is converted into rotational motion that turns a rotor in an electric generator. The generator charges batteries, which are in turn used to provide a consistent energy source for the end user. This model was successfully tested by the US Navy in a project to provide power to coastal security networks and was able to provide an average power of 350 W. The buoy survived the difficult ocean environment, including operation off the New Jersey coast through Hurricane Irene in 2011.

The concepts presented in this chapter will be the foundation for many interesting topics, from the transmission of information to the concepts of quantum mechanics.

Traveling Waves

By the end of this section, you will be able to:

- Describe the basic characteristics of wave motion
- Define the terms wavelength, amplitude, period, frequency, and wave speed
- Explain the difference between longitudinal and transverse waves, and give examples of each type
- List the different types of waves

We saw in [Oscillations](#) that oscillatory motion is an important type of behavior that can be used to model a wide range of physical phenomena. Oscillatory motion is also important because oscillations can generate waves, which are of fundamental importance in physics. Many of the terms and equations we studied in the chapter on oscillations apply equally well to wave motion ([\[link\]](#)).



An ocean wave is probably the first picture that comes to mind when

you hear the word “wave.” Although this breaking wave, and ocean waves in general, have apparent similarities to the basic wave characteristics we will discuss, the mechanisms driving ocean waves are highly complex and beyond the scope of this chapter. It may seem natural, and even advantageous, to apply the concepts in this chapter to ocean waves, but ocean waves are nonlinear, and the simple models presented in this chapter do not fully explain them. (credit: Steve Jurvetson)

Types of Waves

A **wave** is a disturbance that propagates, or moves from the place it was created. There are three basic types of waves: mechanical waves, electromagnetic waves, and matter waves.

Basic **mechanical waves** are governed by Newton’s laws and require a medium. A medium is the substance a mechanical waves propagates through, and the medium produces an elastic restoring force when it is deformed. Mechanical waves transfer energy and momentum, without transferring mass. Some examples of mechanical waves are water waves, sound waves, and seismic waves. The medium for water waves is water; for sound waves, the medium is usually air. (Sound waves can travel in other media as well; we will look at that in more detail in [Sound](#).) For surface water waves, the disturbance occurs on the surface of the water, perhaps created by a rock thrown into a pond or by a swimmer splashing the surface repeatedly. For sound waves, the disturbance is a change in air pressure, perhaps created by the oscillating cone inside a speaker or a vibrating tuning fork. In both cases, the disturbance is the oscillation of the molecules of the fluid. In mechanical waves, energy and momentum transfer with the motion of the wave, whereas the mass oscillates around an equilibrium point. (We discuss this in [Energy and Power of a Wave](#).) Earthquakes generate seismic waves from several types of disturbances, including the disturbance of Earth’s surface and pressure disturbances under the surface. Seismic waves travel through the solids and liquids that form Earth. In this chapter, we focus on mechanical waves.

Electromagnetic waves are associated with oscillations in electric and magnetic fields and do not require a medium. Examples include gamma rays, X-rays, ultraviolet waves, visible light, infrared waves, microwaves, and radio waves. Electromagnetic waves can travel through a vacuum at the speed of light, $v = c = 2.99792458 \times 10^8$ m/s. For example, light from distant stars travels through the vacuum of space and reaches Earth. Electromagnetic waves have some characteristics that are similar to mechanical waves; they are covered in more detail in [Electromagnetic Waves](#).

Matter waves are a central part of the branch of physics known as quantum mechanics. These waves are associated with protons, electrons, neutrons, and other fundamental particles found in nature. The theory that all types of matter have wave-like properties was first proposed by Louis de Broglie in 1924. Matter waves are discussed in [Photons and Matter Waves](#).

Mechanical Waves

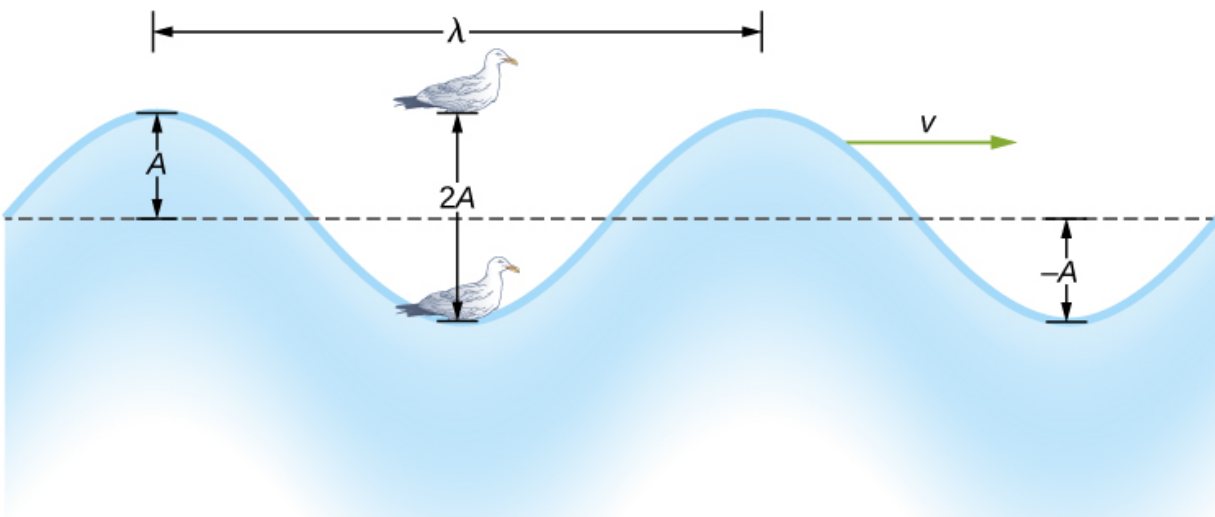
Mechanical waves exhibit characteristics common to all waves, such as amplitude, wavelength, period, frequency, and energy. All wave characteristics can be described by a small set of underlying principles.

The simplest mechanical waves repeat themselves for several cycles and are associated with simple harmonic motion. These simple harmonic waves can be modeled using some combination of sine and cosine functions. For example, consider the simplified surface water wave that moves across the surface of water as illustrated in [\[link\]](#). Unlike complex ocean waves, in surface water waves, the medium, in this case water, moves vertically, oscillating up and down, whereas the disturbance of the wave moves horizontally through the medium. In [\[link\]](#), the waves causes a seagull to move up and down in simple harmonic motion as the wave crests and troughs (peaks and valleys) pass under the bird. The crest is the highest point of the wave, and the trough is the lowest part of the wave. The time for one complete oscillation of the up-and-down motion is the wave's period T . The wave's frequency is the number of waves that pass through a point per unit time and is equal to $f = 1/T$. The period can be expressed

using any convenient unit of time but is usually measured in seconds; frequency is usually measured in hertz (Hz), where $1 \text{ Hz} = 1 \text{ s}^{-1}$.

The length of the wave is called the **wavelength** and is represented by the Greek letter lambda (λ), which is measured in any convenient unit of length, such as a centimeter or meter. The wavelength can be measured between any two similar points along the medium that have the same height and the same slope. In [\[link\]](#), the wavelength is shown measured between two crests. As stated above, the period of the wave is equal to the time for one oscillation, but it is also equal to the time for one wavelength to pass through a point along the wave's path.

The amplitude of the wave (A) is a measure of the maximum displacement of the medium from its equilibrium position. In the figure, the equilibrium position is indicated by the dotted line, which is the height of the water if there were no waves moving through it. In this case, the wave is symmetrical, the crest of the wave is a distance $+A$ above the equilibrium position, and the trough is a distance $-A$ below the equilibrium position. The units for the amplitude can be centimeters or meters, or any convenient unit of distance.



An idealized surface water wave passes under a seagull that bobs up and down in simple harmonic motion. The wave has a wavelength λ , which is the distance between adjacent identical parts of the wave. The amplitude A of the wave is the maximum displacement of the wave from the equilibrium position, which is indicated by the dotted line. In this example, the medium moves up and down, whereas the disturbance of the surface propagates parallel to the surface at a speed v .

The water wave in the figure moves through the medium with a propagation velocity \vec{v} . The magnitude of the **wave velocity** is the distance the wave travels in a given time, which is one wavelength in the time of one period, and the **wave speed** is the magnitude of wave velocity. In equation form, this is

Note:

Equation:

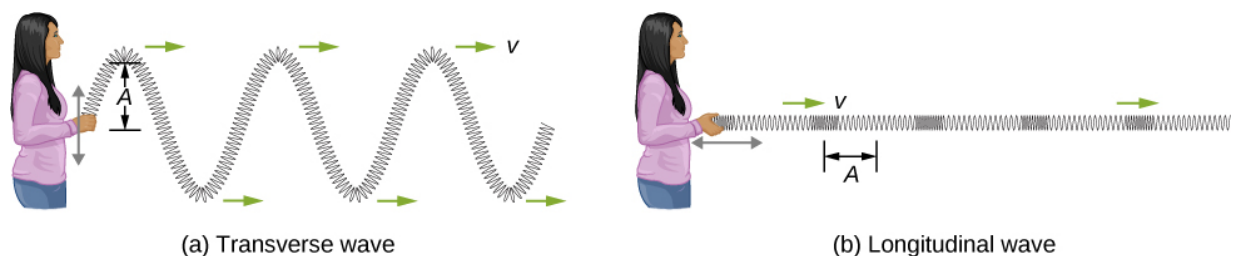
$$v = \frac{\lambda}{T} = \lambda f.$$

This fundamental relationship holds for all types of waves. For water waves, v is the speed of a surface wave; for sound, v is the speed of sound; and for visible light, v is the speed of light.

Transverse and Longitudinal Waves

We have seen that a simple mechanical wave consists of a periodic disturbance that propagates from one place to another through a medium. In [\[link\]](#)(a), the wave propagates in the horizontal direction, whereas the medium is disturbed in the vertical direction. Such a wave is called a

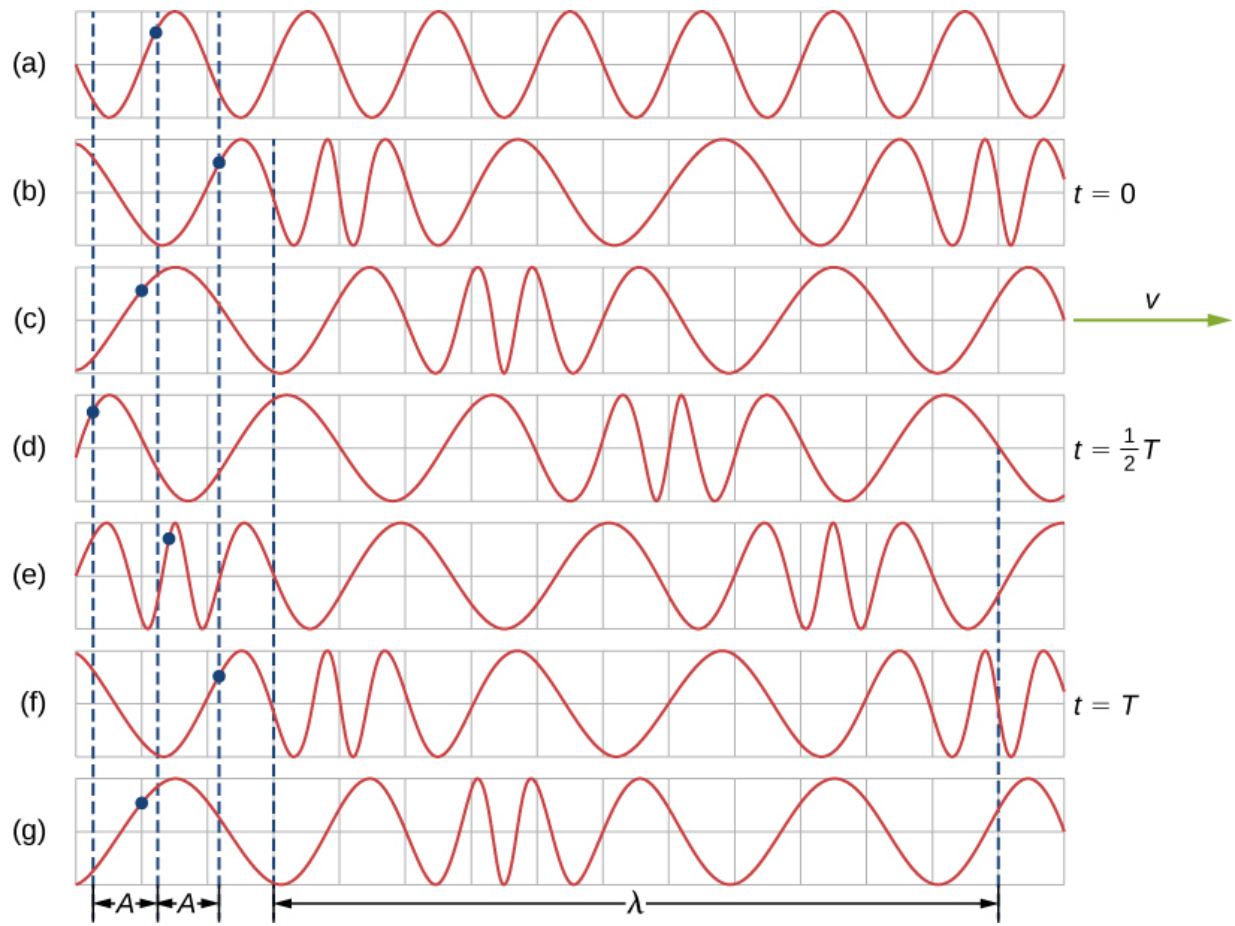
transverse wave. In a transverse wave, the wave may propagate in any direction, but the disturbance of the medium is perpendicular to the direction of propagation. In contrast, in a **longitudinal wave** or compressional wave, the disturbance is parallel to the direction of propagation. [\[link\]](#)(b) shows an example of a longitudinal wave. The size of the disturbance is its amplitude A and is completely independent of the speed of propagation v .



(a) In a transverse wave, the medium oscillates perpendicular to the wave velocity. Here, the spring moves vertically up and down, while the wave propagates horizontally to the right. (b) In a longitudinal wave, the medium oscillates parallel to the propagation of the wave. In this case, the spring oscillates back and forth, while the wave propagates to the right.

A simple graphical representation of a section of the spring shown in [\[link\]](#) (b) is shown in [\[link\]](#). [\[link\]](#)(a) shows the equilibrium position of the spring before any waves move down it. A point on the spring is marked with a blue dot. [\[link\]](#)(b) through (g) show snapshots of the spring taken one-quarter of a period apart, sometime after the end of the spring is oscillated back and forth in the x -direction at a constant frequency. The disturbance of the wave is seen as the compressions and the expansions of the spring. Note that the blue dot oscillates around its equilibrium position a distance A , as the longitudinal wave moves in the positive x -direction with a constant speed. The distance A is the amplitude of the wave. The y -position of the dot does not change as the wave moves through the spring. The wavelength

of the wave is measured in part (d). The wavelength depends on the speed of the wave and the frequency of the driving force.



(a) This is a simple, graphical representation of a section of the stretched spring shown in [\[link\]](#)(b), representing the spring's equilibrium position before any waves are induced on the spring. A point on the spring is marked by a blue dot. (b–g) Longitudinal waves are created by oscillating the end of the spring (not shown) back and forth along the x -axis. The longitudinal wave, with a wavelength λ , moves along the spring in the $+x$ -direction with a wave speed v . For convenience, the wavelength is measured in (d). Note that the point on the spring that was marked with the blue dot moves back and forth a distance A from the equilibrium position, oscillating around the equilibrium position of the point.

Waves may be transverse, longitudinal, or a combination of the two. Examples of transverse waves are the waves on stringed instruments or surface waves on water, such as ripples moving on a pond. Sound waves in air and water are longitudinal. With sound waves, the disturbances are periodic variations in pressure that are transmitted in fluids. Fluids do not have appreciable shear strength, and for this reason, the sound waves in them are longitudinal waves. Sound in solids can have both longitudinal and transverse components, such as those in a seismic wave. Earthquakes generate seismic waves under Earth's surface with both longitudinal and transverse components (called compressional or P-waves and shear or S-waves, respectively). The components of seismic waves have important individual characteristics—they propagate at different speeds, for example. Earthquakes also have surface waves that are similar to surface waves on water. Ocean waves also have both transverse and longitudinal components.

Example:**Wave on a String**

A student takes a 30.00-m-long string and attaches one end to the wall in the physics lab. The student then holds the free end of the rope, keeping the tension constant in the rope. The student then begins to send waves down the string by moving the end of the string up and down with a frequency of 2.00 Hz. The maximum displacement of the end of the string is 20.00 cm. The first wave hits the lab wall 6.00 s after it was created. (a) What is the speed of the wave? (b) What is the period of the wave? (c) What is the wavelength of the wave?

Strategy

- The speed of the wave can be derived by dividing the distance traveled by the time.
- The period of the wave is the inverse of the frequency of the driving force.
- The wavelength can be found from the speed and the period
 $v = \lambda/T$.

Solution

- a. The first wave traveled 30.00 m in 6.00 s:

Equation:

$$v = \frac{30.00 \text{ m}}{6.00 \text{ s}} = 5.00 \frac{\text{m}}{\text{s}}.$$

- b. The period is equal to the inverse of the frequency:

Equation:

$$T = \frac{1}{f} = \frac{1}{2.00 \text{ s}^{-1}} = 0.50 \text{ s}.$$

- c. The wavelength is equal to the velocity times the period:

Equation:

$$\lambda = vT = 5.00 \frac{\text{m}}{\text{s}} (0.50 \text{ s}) = 2.50 \text{ m}.$$

Significance

The frequency of the wave produced by an oscillating driving force is equal to the frequency of the driving force.

Note:

Exercise:

Problem:

Check Your Understanding When a guitar string is plucked, the guitar string oscillates as a result of waves moving through the string. The vibrations of the string cause the air molecules to oscillate, forming sound waves. The frequency of the sound waves is equal to the frequency of the vibrating string. Is the wavelength of the sound wave always equal to the wavelength of the waves on the string?

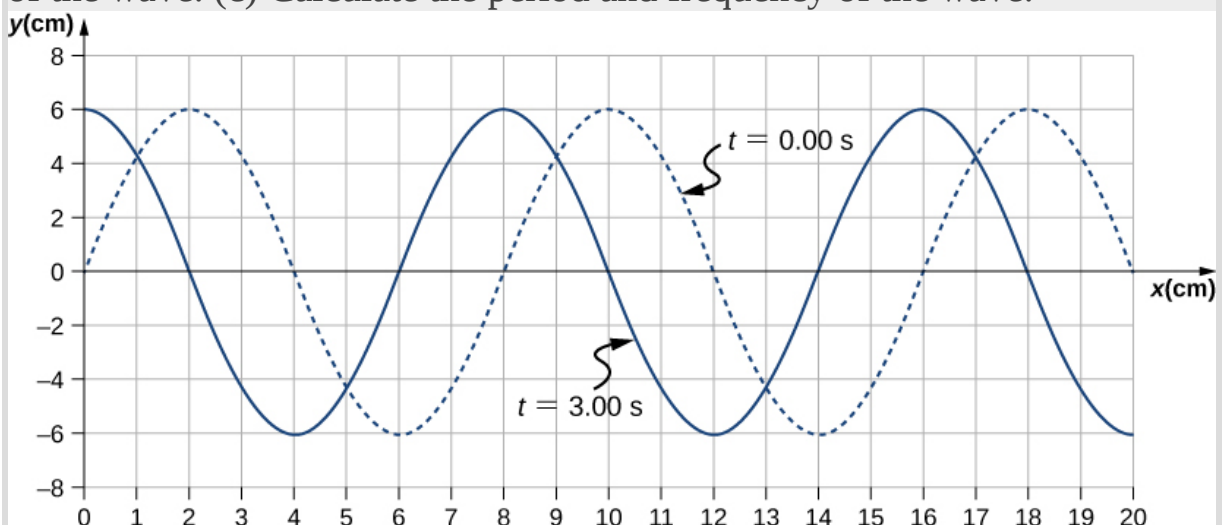
Solution:

The wavelength of the waves depends on the frequency and the velocity of the wave. The frequency of the sound wave is equal to the frequency of the wave on the string. The wavelengths of the sound waves and the waves on the string are equal only if the velocities of the waves are the same, which is not always the case. If the speed of the sound wave is different from the speed of the wave on the string, the wavelengths are different. This velocity of sound waves will be discussed in [Sound](#).

Example:

Characteristics of a Wave

A transverse mechanical wave propagates in the positive x -direction through a spring (as shown in [link](#)(a)) with a constant wave speed, and the medium oscillates between $+A$ and $-A$ around an equilibrium position. The graph in [link](#) shows the height of the spring (y) versus the position (x), where the x -axis points in the direction of propagation. The figure shows the height of the spring versus the x -position at $t = 0.00$ s as a dotted line and the wave at $t = 3.00$ s as a solid line. Assume the wave has not traveled more than 1 wavelength in this time. (a) Determine the wavelength and amplitude of the wave. (b) Find the propagation velocity of the wave. (c) Calculate the period and frequency of the wave.



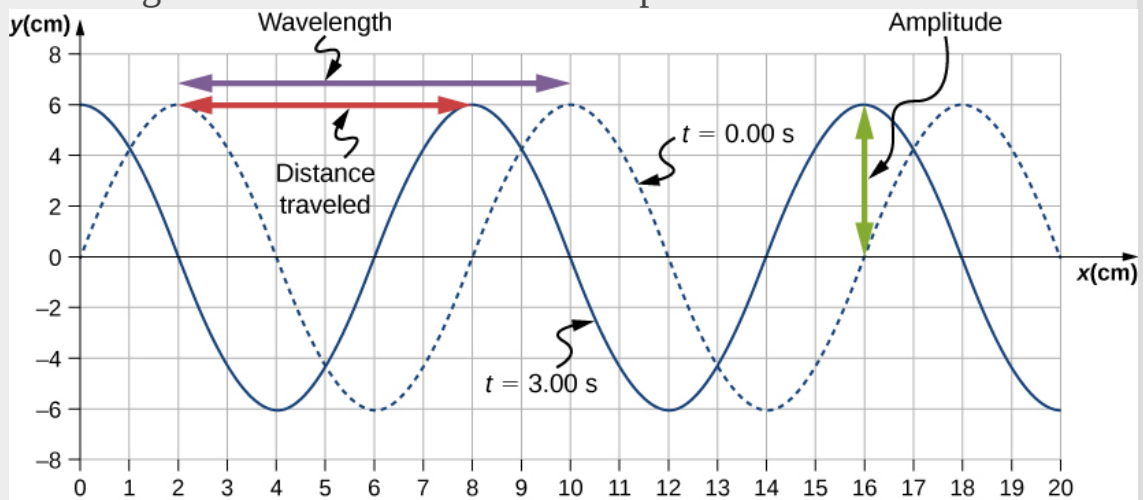
A transverse wave shown at two instants of time.

Strategy

- The amplitude and wavelength can be determined from the graph.
- Since the velocity is constant, the velocity of the wave can be found by dividing the distance traveled by the wave by the time it took the wave to travel the distance.
- The period can be found from $v = \frac{\lambda}{T}$ and the frequency from $f = \frac{1}{T}$.

Solution

- Read the wavelength from the graph, looking at the purple arrow in [\[link\]](#). Read the amplitude by looking at the green arrow. The wavelength is $\lambda = 8.00$ cm and the amplitude is $A = 6.00$ cm.



Characteristics of the wave marked on a graph of its displacement.

- The distance the wave traveled from time $t = 0.00$ s to time $t = 3.00$ s can be seen in the graph. Consider the red arrow, which shows the distance the crest has moved in 3 s. The distance is 8.00 cm $-$ 2.00 cm $=$ 6.00 cm. The velocity is

Equation:

$$v = \frac{\Delta x}{\Delta t} = \frac{8.00 \text{ cm} - 2.00 \text{ cm}}{3.00 \text{ s} - 0.00 \text{ s}} = 2.00 \text{ cm/s}.$$

c. The period is $T = \frac{\lambda}{v} = \frac{8.00 \text{ cm}}{2.00 \text{ cm/s}} = 4.00 \text{ s}$ and the frequency is $f = \frac{1}{T} = \frac{1}{4.00 \text{ s}} = 0.25 \text{ Hz}$.

Significance

Note that the wavelength can be found using any two successive identical points that repeat, having the same height and slope. You should choose two points that are most convenient. The displacement can also be found using any convenient point.

Note:

Exercise:

Problem:

Check Your Understanding The propagation velocity of a transverse or longitudinal mechanical wave may be constant as the wave disturbance moves through the medium. Consider a transverse mechanical wave: Is the velocity of the medium also constant?

Solution:

In a transverse wave, the wave may move at a constant propagation velocity through the medium, but the medium oscillates perpendicular to the motion of the wave. If the wave moves in the positive x -direction, the medium oscillates up and down in the y -direction. The velocity of the medium is therefore not constant, but the medium's velocity and acceleration are similar to that of the simple harmonic motion of a mass on a spring.

Summary

- A wave is a disturbance that moves from the point of origin with a wave velocity v .
- A wave has a wavelength λ , which is the distance between adjacent identical parts of the wave. Wave velocity and wavelength are related to the wave's frequency and period by $v = \frac{\lambda}{T} = \lambda f$.
- Mechanical waves are disturbances that move through a medium and are governed by Newton's laws.
- Electromagnetic waves are disturbances in the electric and magnetic fields, and do not require a medium.
- Matter waves are a central part of quantum mechanics and are associated with protons, electrons, neutrons, and other fundamental particles found in nature.
- A transverse wave has a disturbance perpendicular to the wave's direction of propagation, whereas a longitudinal wave has a disturbance parallel to its direction of propagation.

Conceptual Questions

Exercise:

Problem:

Give one example of a transverse wave and one example of a longitudinal wave, being careful to note the relative directions of the disturbance and wave propagation in each.

Solution:

A wave on a guitar string is an example of a transverse wave. The disturbance of the string moves perpendicular to the propagation of the wave. The sound produced by the string is a longitudinal wave where the disturbance of the air moves parallel to the propagation of the wave.

Exercise:

Problem:

A sinusoidal transverse wave has a wavelength of 2.80 m. It takes 0.10 s for a portion of the string at a position x to move from a maximum position of $y = 0.03$ m to the equilibrium position $y = 0$. What are the period, frequency, and wave speed of the wave?

Exercise:**Problem:**

What is the difference between propagation speed and the frequency of a mechanical wave? Does one or both affect wavelength? If so, how?

Solution:

Propagation speed is the speed of the wave propagating through the medium. If the wave speed is constant, the speed can be found by $v = \frac{\lambda}{T} = \lambda f$. The frequency is the number of wave that pass a point per unit time. The wavelength is directly proportional to the wave speed and inversely proportional to the frequency.

Exercise:**Problem:**

Consider a stretched spring, such as a slinky. The stretched spring can support longitudinal waves and transverse waves. How can you produce transverse waves on the spring? How can you produce longitudinal waves on the spring?

Exercise:**Problem:**

Consider a wave produced on a stretched spring by holding one end and shaking it up and down. Does the wavelength depend on the distance you move your hand up and down?

Solution:

No, the distance you move your hand up and down will determine the amplitude of the wave. The wavelength will depend on the frequency you move your hand up and down, and the speed of the wave through the spring.

Exercise:

Problem:

A sinusoidal, transverse wave is produced on a stretched spring, having a period T . Each section of the spring moves perpendicular to the direction of propagation of the wave, in simple harmonic motion with an amplitude A . Does each section oscillate with the same period as the wave or a different period? If the amplitude of the transverse wave were doubled but the period stays the same, would your answer be the same?

Exercise:

Problem:

An electromagnetic wave, such as light, does not require a medium. Can you think of an example that would support this claim?

Solution:

Light from the Sun and stars reach Earth through empty space where there is no medium present.

Problems

Exercise:

Problem:

Storms in the South Pacific can create waves that travel all the way to the California coast, 12,000 km away. How long does it take them to travel this distance if they travel at 15.0 m/s?

Exercise:

Problem:

Waves on a swimming pool propagate at 0.75 m/s. You splash the water at one end of the pool and observe the wave go to the opposite end, reflect, and return in 30.00 s. How far away is the other end of the pool?

Solution:

$$2d = vt \Rightarrow d = 11.25 \text{ m}$$

Exercise:**Problem:**

Wind gusts create ripples on the ocean that have a wavelength of 5.00 cm and propagate at 2.00 m/s. What is their frequency?

Exercise:**Problem:**

How many times a minute does a boat bob up and down on ocean waves that have a wavelength of 40.0 m and a propagation speed of 5.00 m/s?

Solution:

$$v = f\lambda, \text{ so that } f = 0.125 \text{ Hz, so that}$$
$$N = 7.50 \text{ times}$$

Exercise:**Problem:**

Scouts at a camp shake the rope bridge they have just crossed and observe the wave crests to be 8.00 m apart. If they shake the bridge twice per second, what is the propagation speed of the waves?

Exercise:

Problem:

What is the wavelength of the waves you create in a swimming pool if you splash your hand at a rate of 2.00 Hz and the waves propagate at a wave speed of 0.800 m/s?

Solution:

$$v = f\lambda \Rightarrow \lambda = 0.400 \text{ m}$$

Exercise:**Problem:**

What is the wavelength of an earthquake that shakes you with a frequency of 10.0 Hz and gets to another city 84.0 km away in 12.0 s?

Exercise:**Problem:**

Radio waves transmitted through empty space at the speed of light $v = c = 3.00 \times 10^8 \text{ m/s}$ by the *Voyager* spacecraft have a wavelength of 0.120 m. What is their frequency?

Solution:

$$v = f\lambda \Rightarrow f = 2.50 \times 10^9 \text{ Hz}$$

Exercise:

Problem:

Your ear is capable of differentiating sounds that arrive at each ear just 0.34 ms apart, which is useful in determining where low frequency sound is originating from. (a) Suppose a low-frequency sound source is placed to the right of a person, whose ears are approximately 18 cm apart, and the speed of sound generated is 340 m/s. How long is the interval between when the sound arrives at the right ear and the sound arrives at the left ear? (b) Assume the same person was scuba diving and a low-frequency sound source was to the right of the scuba diver. How long is the interval between when the sound arrives at the right ear and the sound arrives at the left ear, if the speed of sound in water is 1500 m/s? (c) What is significant about the time interval of the two situations?

Exercise:**Problem:**

(a) Seismographs measure the arrival times of earthquakes with a precision of 0.100 s. To get the distance to the epicenter of the quake, geologists compare the arrival times of S- and P-waves, which travel at different speeds. If S- and P-waves arrive at 4.00 and 7.20 km/s, respectively, in the region considered, how precisely can the distance to the source of the earthquake be determined? (b) Seismic waves from underground detonations of nuclear bombs can be used to locate the test site and detect violations of test bans. Discuss whether your answer to (a) implies a serious limit to such detection. (Note also that the uncertainty is greater if there is an uncertainty in the propagation speeds of the S- and P-waves.)

Solution:

a. The P-waves outrun the S-waves by a speed of $v = 3.20$ km/s; therefore, $\Delta d = 0.320$ km. b. Since the uncertainty in the distance is less than a kilometer, our answer to part (a) does not seem to limit the detection of nuclear bomb detonations. However, if the velocities are

uncertain, then the uncertainty in the distance would increase and could then make it difficult to identify the source of the seismic waves.

Exercise:

Problem:

A Girl Scout is taking a 10.00-km hike to earn a merit badge. While on the hike, she sees a cliff some distance away. She wishes to estimate the time required to walk to the cliff. She knows that the speed of sound is approximately 343 meters per second. She yells and finds that the echo returns after approximately 2.00 seconds. If she can hike 1.00 km in 10 minutes, how long would it take her to reach the cliff?

Exercise:

Problem:

A quality assurance engineer at a frying pan company is asked to qualify a new line of nonstick-coated frying pans. The coating needs to be 1.00 mm thick. One method to test the thickness is for the engineer to pick a percentage of the pans manufactured, strip off the coating, and measure the thickness using a micrometer. This method is a destructive testing method. Instead, the engineer decides that every frying pan will be tested using a nondestructive method. An ultrasonic transducer is used that produces sound waves with a frequency of $f = 25$ kHz. The sound waves are sent through the coating and are reflected by the interface between the coating and the metal pan, and the time is recorded. The wavelength of the ultrasonic waves in the coating is 0.076 m. What should be the time recorded if the coating is the correct thickness (1.00 mm)?

Solution:

$$\begin{aligned}v &= 1900 \text{ m/s} \\ \Delta t &= 1.05 \mu\text{s}\end{aligned}$$

Glossary

longitudinal wave

wave in which the disturbance is parallel to the direction of propagation

mechanical wave

wave that is governed by Newton's laws and requires a medium

transverse wave

wave in which the disturbance is perpendicular to the direction of propagation

wave

disturbance that moves from its source and carries energy

wave velocity

velocity at which the disturbance moves; also called the propagation velocity

wave speed

magnitude of the wave velocity

wavelength

distance between adjacent identical parts of a wave

Mathematics of Waves

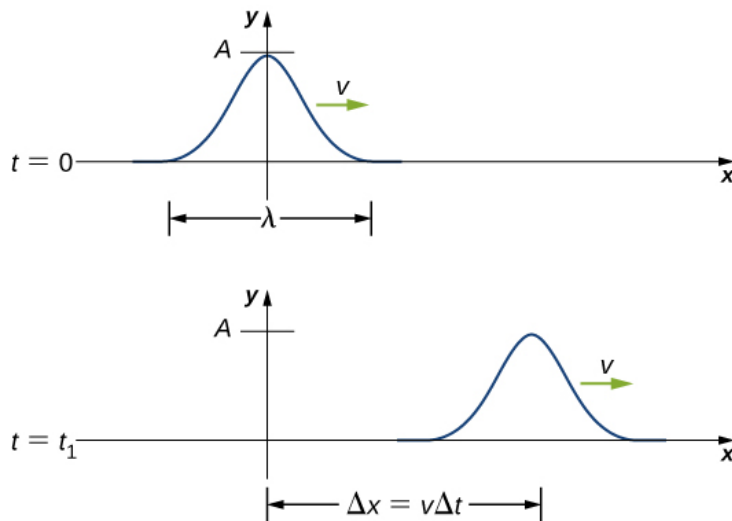
By the end of this section, you will be able to:

- Model a wave, moving with a constant wave velocity, with a mathematical expression
- Calculate the velocity and acceleration of the medium
- Show how the velocity of the medium differs from the wave velocity (propagation velocity)

In the previous section, we described periodic waves by their characteristics of wavelength, period, amplitude, and wave speed of the wave. Waves can also be described by the motion of the particles of the medium through which the waves move. The position of particles of the medium can be mathematically modeled as **wave functions**, which can be used to find the position, velocity, and acceleration of the particles of the medium of the wave at any time.

Pulses

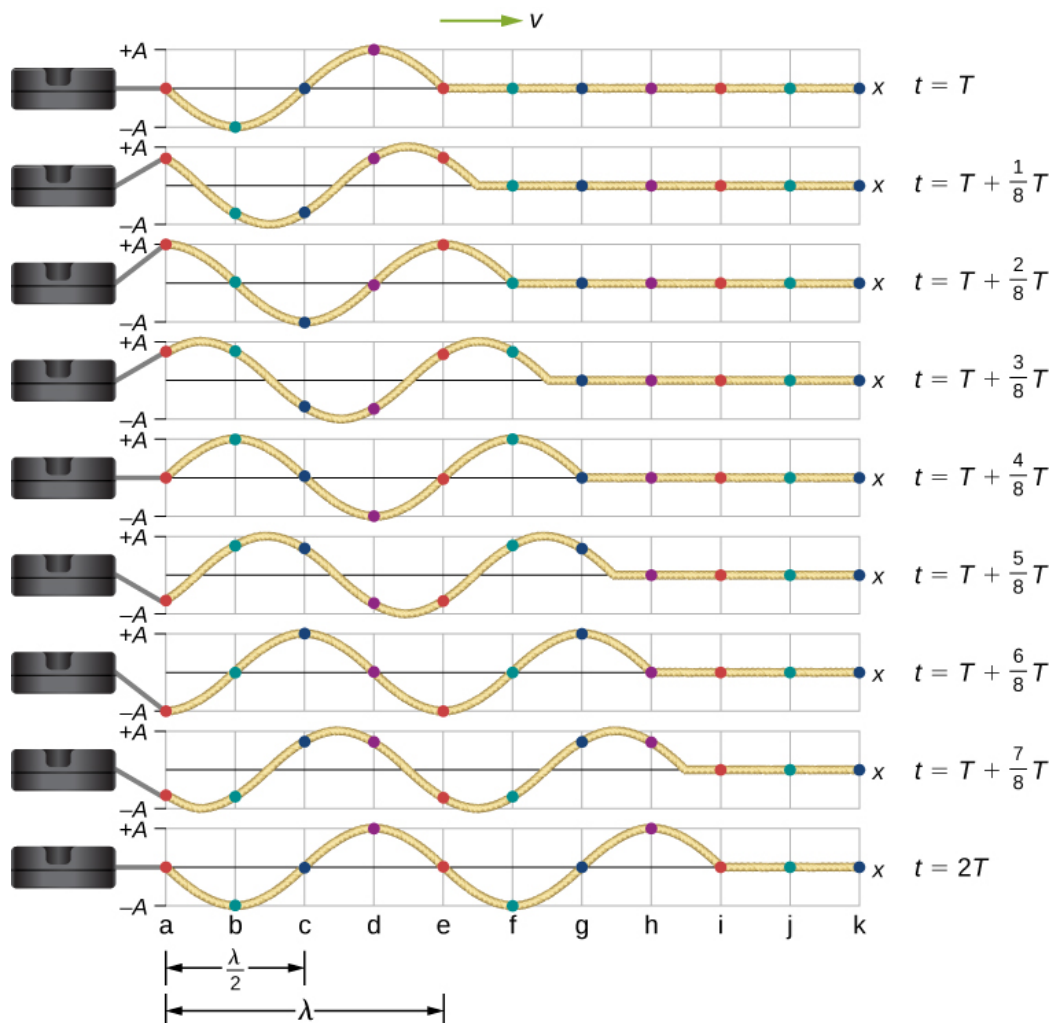
A **pulse** can be described as wave consisting of a single disturbance that moves through the medium with a constant amplitude. The pulse moves as a pattern that maintains its shape as it propagates with a constant wave speed. Because the wave speed is constant, the distance the pulse moves in a time Δt is equal to $\Delta x = v\Delta t$ ([link](#)).



The pulse at time $t = 0$ is centered on $x = 0$ with amplitude A . The pulse moves as a pattern with a constant shape, with a constant maximum value A . The velocity is constant and the pulse moves a distance $\Delta x = v\Delta t$ in a time Δt . The distance traveled is measured with any convenient point on the pulse. In this figure, the crest is used.

Modeling a One-Dimensional Sinusoidal Wave using a Wave Function

Consider a string kept at a constant tension F_T where one end is fixed and the free end is oscillated between $y = +A$ and $y = -A$ by a mechanical device at a constant frequency. [link](#) shows snapshots of the wave at an interval of an eighth of a period, beginning after one period ($t = T$).



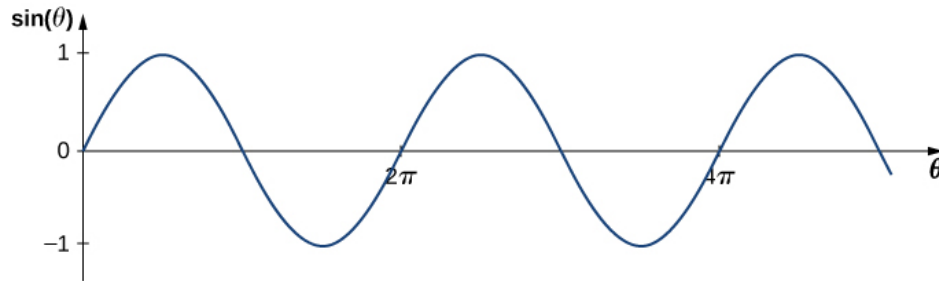
Snapshots of a transverse wave moving through a string under tension, beginning at time $t = T$ and taken at intervals of $\frac{1}{8}T$. Colored dots are used to highlight points on the string. Points that are a wavelength apart in the x -direction are highlighted with the same color dots.

Notice that each select point on the string (marked by colored dots) oscillates up and down in simple harmonic motion, between $y = +A$ and $y = -A$, with a period T . The wave on the string is sinusoidal and is translating in the positive x -direction as time progresses.

At this point, it is useful to recall from your study of algebra that if $f(x)$ is some function, then $f(x - d)$ is the same function translated in the positive x -direction by a distance d . The function $f(x + d)$ is the same function translated in the negative x -direction by a distance d . We want to define a wave function that will give the y -position of each segment of the string for every position x along the string for every time t .

Looking at the first snapshot in [\[link\]](#), the y -position of the string between $x = 0$ and $x = \lambda$ can be modeled as a sine function. This wave propagates down the string one wavelength in one period, as seen in the last snapshot. The wave therefore moves with a constant wave speed of $v = \lambda/T$.

Recall that a sine function is a function of the angle θ , oscillating between $+1$ and -1 , and repeating every 2π radians ([link](#)). However, the y -position of the medium, or the wave function, oscillates between $+A$ and $-A$, and repeats every wavelength λ .



A sine function oscillates between $+1$ and -1 every 2π radians.

To construct our model of the wave using a periodic function, consider the ratio of the angle and the position,
Equation:

$$\begin{aligned}\frac{\theta}{x} &= \frac{2\pi}{\lambda}, \\ \theta &= \frac{2\pi}{\lambda}x.\end{aligned}$$

Using $\theta = \frac{2\pi}{\lambda}x$ and multiplying the sine function by the amplitude A , we can now model the y -position of the string as a function of the position x :

Equation:

$$y(x) = A \sin\left(\frac{2\pi}{\lambda}x\right).$$

The wave on the string travels in the positive x -direction with a constant velocity v , and moves a distance vt in a time t . The wave function can now be defined by

Equation:

$$y(x, t) = A \sin\left(\frac{2\pi}{\lambda}(x - vt)\right).$$

It is often convenient to rewrite this wave function in a more compact form. Multiplying through by the ratio $\frac{2\pi}{\lambda}$ leads to the equation

Equation:

$$y(x, t) = A \sin\left(\frac{2\pi}{\lambda}x - \frac{2\pi}{\lambda}vt\right).$$

The value $\frac{2\pi}{\lambda}$ is defined as the **wave number**. The symbol for the wave number is k and has units of inverse meters, m^{-1} :

Note:

Equation:

$$k \equiv \frac{2\pi}{\lambda}$$

Recall from [Oscillations](#) that the angular frequency is defined as $\omega \equiv \frac{2\pi}{T}$. The second term of the wave function becomes

Equation:

$$\frac{2\pi}{\lambda}vt = \frac{2\pi}{\lambda}\left(\frac{\lambda}{T}\right)t = \frac{2\pi}{T}t = \omega t.$$

The wave function for a simple harmonic wave on a string reduces to

Equation:

$$y(x, t) = A \sin(kx \mp \omega t),$$

where A is the amplitude, $k = \frac{2\pi}{\lambda}$ is the wave number, $\omega = \frac{2\pi}{T}$ is the angular frequency, the minus sign is for waves moving in the positive x -direction, and the plus sign is for waves moving in the negative x -direction. The velocity of the wave is equal to

Note:

Equation:

$$v = \frac{\lambda}{T} = \frac{\lambda}{T} \left(\frac{2\pi}{2\pi} \right) = \frac{\omega}{k}.$$

Think back to our discussion of a mass on a spring, when the position of the mass was modeled as $x(t) = A \cos(\omega t + \phi)$. The angle ϕ is a phase shift, added to allow for the fact that the mass may have initial conditions other than $x = +A$ and $v = 0$. For similar reasons, the initial phase is added to the wave function. The wave function modeling a sinusoidal wave, allowing for an initial phase shift ϕ , is

Note:

Equation:

$$y(x, t) = A \sin(kx \mp \omega t + \phi)$$

The value

Note:

Equation:

$$(kx \mp \omega t + \phi)$$

is known as the phase of the wave, where ϕ is the initial phase of the wave function. Whether the temporal term ωt is negative or positive depends on the direction of the wave. First consider the minus sign for a wave with an initial phase equal to zero ($\phi = 0$). The phase of the wave would be $(kx - \omega t)$. Consider following a point on a wave, such as a crest. A crest will occur when $\sin(kx - \omega t) = 1.00$, that is, when $kx - \omega t = n\pi + \frac{\pi}{2}$, for any integral value of n . For instance, one particular crest occurs at $kx - \omega t = \frac{\pi}{2}$. As the wave moves, time increases and x must also increase to keep the phase equal to $\frac{\pi}{2}$. Therefore, the minus sign is for a wave moving in the positive x -direction. Using the plus sign, $kx + \omega t = \frac{\pi}{2}$. As time increases, x must decrease to keep the phase equal to $\frac{\pi}{2}$. The plus sign is used for waves moving in the negative x -direction. In summary, $y(x, t) = A \sin(kx - \omega t + \phi)$ models a wave moving in the positive x -direction and $y(x, t) = A \sin(kx + \omega t + \phi)$ models a wave moving in the negative x -direction.

[\[link\]](#) is known as a simple harmonic wave function. A wave function is any function such that $f(x, t) = f(x - vt)$. Later in this chapter, we will see that it is a solution to the linear wave equation. Note that $y(x, t) = A \cos(kx + \omega t + \phi)$ works equally well because it corresponds to a different phase shift $\phi' = \phi - \frac{\pi}{2}$.

Note:

Finding the Characteristics of a Sinusoidal Wave

1. To find the amplitude, wavelength, period, and frequency of a sinusoidal wave, write down the wave function in the form $y(x, t) = A \sin(kx - \omega t + \phi)$.
2. The amplitude can be read straight from the equation and is equal to A .
3. The period of the wave can be derived from the angular frequency ($T = \frac{2\pi}{\omega}$).
4. The frequency can be found using $f = \frac{1}{T}$.
5. The wavelength can be found using the wave number ($\lambda = \frac{2\pi}{k}$).

Example:

Characteristics of a Traveling Wave on a String

A transverse wave on a taut string is modeled with the wave function

Equation:

$$y(x, t) = A \sin(kx - \omega t) = 0.2 \text{ m} \sin(6.28 \text{ m}^{-1}x - 1.57 \text{ s}^{-1}t).$$

Find the amplitude, wavelength, period, and speed of the wave.

Strategy

All these characteristics of the wave can be found from the constants included in the equation or from simple combinations of these constants.

Solution

1. The amplitude, wave number, and angular frequency can be read directly from the wave equation:

Equation:

$$y(x, t) = A \sin(kx - \omega t) = 0.2 \text{ m} \sin(6.28 \text{ m}^{-1}x - 1.57 \text{ s}^{-1}t).$$

Equation:

$$(A = 0.2 \text{ m}; k = 6.28 \text{ m}^{-1}; \omega = 1.57 \text{ s}^{-1})$$

2. The wave number can be used to find the wavelength:

Equation:

$$k = \frac{2\pi}{\lambda}.$$

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{6.28 \text{ m}^{-1}} = 1.0 \text{ m}.$$

3. The period of the wave can be found using the angular frequency:

Equation:

$$\omega = \frac{2\pi}{T}.$$

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{1.57 \text{ s}^{-1}} = 4 \text{ s}.$$

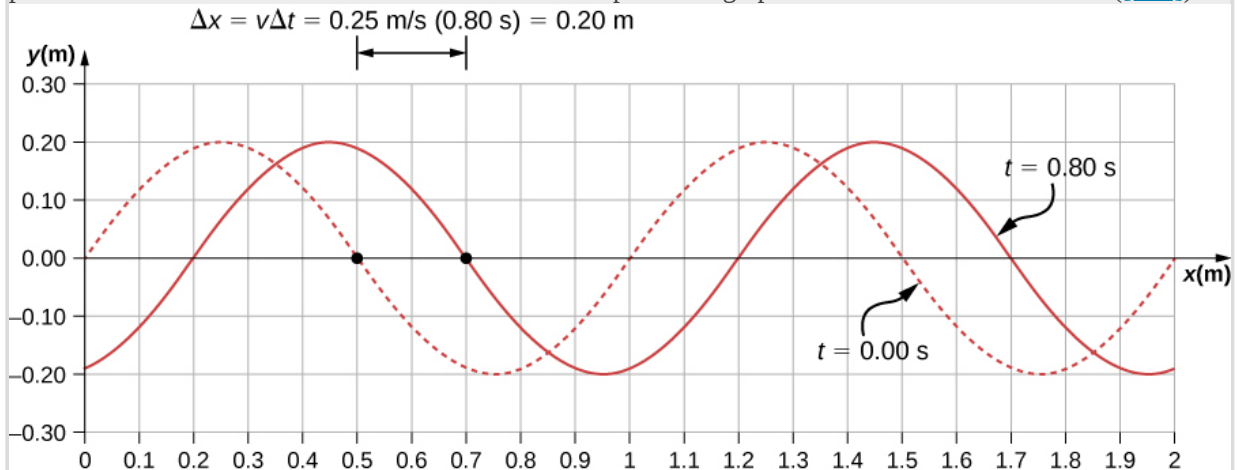
4. The speed of the wave can be found using the wave number and the angular frequency. The direction of the wave can be determined by considering the sign of $kx \mp \omega t$: A negative sign suggests that the wave is moving in the positive x -direction:

Equation:

$$|v| = \frac{\omega}{k} = \frac{1.57 \text{ s}^{-1}}{6.28 \text{ m}^{-1}} = 0.25 \text{ m/s}.$$

Significance

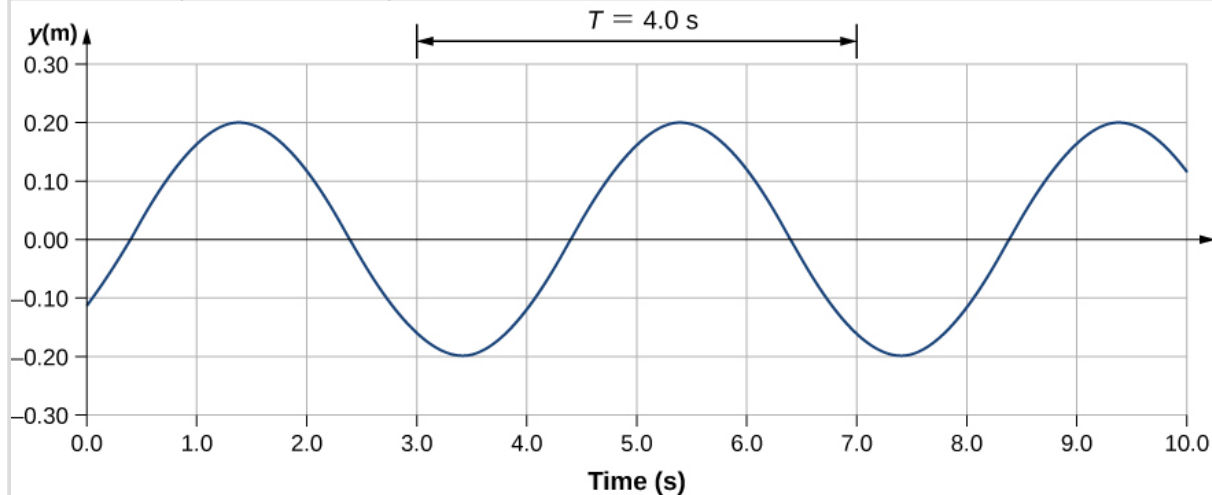
All of the characteristics of the wave are contained in the wave function. Note that the wave speed is the speed of the wave in the direction parallel to the motion of the wave. Plotting the height of the medium y versus the position x for two times $t = 0.00 \text{ s}$ and $t = 0.80 \text{ s}$ can provide a graphical visualization of the wave ([link](#)).



A graph of height of the wave y as a function of position x for snapshots of the wave at two times. The dotted line represents the wave at time $t = 0.00 \text{ s}$ and the solid line represents the wave at $t = 0.80 \text{ s}$.

Since the wave velocity is constant, the distance the wave travels is the wave velocity times the time interval. The black dots indicate the points used to measure the displacement of the wave. The medium moves up and down, whereas the wave moves to the right.

There is a second velocity to the motion. In this example, the wave is transverse, moving horizontally as the medium oscillates up and down perpendicular to the direction of motion. The graph in [\[link\]](#) shows the motion of the medium at point $x = 0.60 \text{ m}$ as a function of time. Notice that the medium of the wave oscillates up and down between $y = +0.20 \text{ m}$ and $y = -0.20 \text{ m}$ every period of 4.0 seconds.



A graph of height of the wave y as a function of time t for the position $x = 0.6 \text{ m}$. The medium oscillates between $y = +0.20 \text{ m}$ and $y = -0.20 \text{ m}$ every period. The period represented picks two convenient points in the oscillations to measure the period. The period can be measured between any two adjacent points with the same amplitude and the same velocity, $(\partial y / \partial t)$. The velocity can be found by looking at the slope tangent to the point on a y -versus- t plot. Notice that at times $t = 3.00 \text{ s}$ and $t = 7.00 \text{ s}$, the heights and the velocities are the same and the period of the oscillation is 4.00 s.

Note:

Exercise:

Problem:

Check Your Understanding The wave function above is derived using a sine function. Can a cosine function be used instead?

Solution:

Yes, a cosine function is equal to a sine function with a phase shift, and either function can be used in a wave function. Which function is more convenient to use depends on the initial conditions. In [\[link\]](#), the wave has an initial height of $y(0.00, 0.00) = 0$ and then the wave height increases to the maximum height at the crest. If the initial height at the initial time was equal to the amplitude of the wave $y(0.00, 0.00) = +A$, then it might be more convenient to model the wave with a cosine function.

Velocity and Acceleration of the Medium

As seen in [\[link\]](#), the wave speed is constant and represents the speed of the wave as it propagates through the medium, not the speed of the particles that make up the medium. The particles of the medium oscillate around

an equilibrium position as the wave propagates through the medium. In the case of the transverse wave propagating in the x -direction, the particles oscillate up and down in the y -direction, perpendicular to the motion of the wave. The velocity of the particles of the medium is not constant, which means there is an acceleration. The velocity of the medium, which is perpendicular to the wave velocity in a transverse wave, can be found by taking the partial derivative of the position equation with respect to time. The partial derivative is found by taking the derivative of the function, treating all variables as constants, except for the variable in question. In the case of the partial derivative with respect to time t , the position x is treated as a constant. Although this may sound strange if you haven't seen it before, the object of this exercise is to find the transverse velocity at a point, so in this sense, the x -position is not changing. We have

Equation:

$$\begin{aligned} y(x, t) &= A \sin(kx - \omega t + \phi) \\ v_y(x, t) &= \frac{\partial y(x, t)}{\partial t} = \frac{\partial}{\partial t}(A \sin(kx - \omega t + \phi)) \\ &= -A\omega \cos(kx - \omega t + \phi) \\ &= -v_{y \max} \cos(kx - \omega t + \phi). \end{aligned}$$

The magnitude of the maximum velocity of the medium is $|v_{y \max}| = A\omega$. This may look familiar from the [Oscillations](#) and a mass on a spring.

We can find the acceleration of the medium by taking the partial derivative of the velocity equation with respect to time,

Equation:

$$\begin{aligned} a_y(x, t) &= \frac{\partial v_y}{\partial t} = \frac{\partial}{\partial t}(-A\omega \cos(kx - \omega t + \phi)) \\ &= -A\omega^2 \sin(kx - \omega t + \phi) \\ &= -a_{y \max} \sin(kx - \omega t + \phi). \end{aligned}$$

The magnitude of the maximum acceleration is $|a_{y \max}| = A\omega^2$. The particles of the medium, or the mass elements, oscillate in simple harmonic motion for a mechanical wave.

The Linear Wave Equation

We have just determined the velocity of the medium at a position x by taking the partial derivative, with respect to time, of the position y . For a transverse wave, this velocity is perpendicular to the direction of propagation of the wave. We found the acceleration by taking the partial derivative, with respect to time, of the velocity, which is the second time derivative of the position:

Equation:

$$a_y(x, t) = \frac{\partial^2 y(x, t)}{\partial t^2} = \frac{\partial^2}{\partial t^2}(A \sin(kx - \omega t + \phi)) = -A\omega^2 \sin(kx - \omega t + \phi).$$

Now consider the partial derivatives with respect to the other variable, the position x , holding the time constant. The first derivative is the slope of the wave at a point x at a time t ,

Equation:

$$\text{slope} = \frac{\partial y(x, t)}{\partial x} = \frac{\partial}{\partial x}(A \sin(kx - \omega t + \phi)) = Ak \cos(kx - \omega t + \phi).$$

The second partial derivative expresses how the slope of the wave changes with respect to position—in other words, the curvature of the wave, where

Equation:

$$\text{curvature} = \frac{\partial^2 y(x, t)}{\partial x^2} = \frac{\partial^2}{\partial x^2} (A \sin(kx - \omega t + \phi)) = -Ak^2 \sin(kx - \omega t + \phi).$$

The ratio of the acceleration and the curvature leads to a very important relationship in physics known as the **linear wave equation**. Taking the ratio and using the equation $v = \omega/k$ yields the linear wave equation (also known simply as the wave equation or the equation of a vibrating string),

Equation:

$$\begin{aligned} \frac{\frac{\partial^2 y(x, t)}{\partial t^2}}{\frac{\partial^2 y(x, t)}{\partial x^2}} &= \frac{-A\omega^2 \sin(kx - \omega t + \phi)}{-Ak^2 \sin(kx - \omega t + \phi)} \\ &= \frac{\omega^2}{k^2} = v^2, \end{aligned}$$

Note:

Equation:

$$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y(x, t)}{\partial t^2}.$$

[\[link\]](#) is the linear wave equation, which is one of the most important equations in physics and engineering. We derived it here for a transverse wave, but it is equally important when investigating longitudinal waves. This relationship was also derived using a sinusoidal wave, but it successfully describes any wave or pulse that has the form $y(x, t) = f(x \mp vt)$. These waves result due to a linear restoring force of the medium—thus, the name linear wave equation. Any wave function that satisfies this equation is a linear wave function.

An interesting aspect of the linear wave equation is that if two wave functions are individually solutions to the linear wave equation, then the sum of the two linear wave functions is also a solution to the wave equation. Consider two transverse waves that propagate along the x -axis, occupying the same medium. Assume that the individual waves can be modeled with the wave functions $y_1(x, t) = f(x \mp vt)$ and $y_2(x, t) = g(x \mp vt)$, which are solutions to the linear wave equations and are therefore linear wave functions. The sum of the wave functions is the wave function

Equation:

$$y_1(x, t) + y_2(x, t) = f(x \mp vt) + g(x \mp vt).$$

Consider the linear wave equation:

Equation:

$$\begin{aligned} \frac{\partial^2(f+g)}{\partial x^2} &= \frac{1}{v^2} \frac{\partial^2(f+g)}{\partial t^2} \\ \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 g}{\partial x^2} &= \frac{1}{v^2} \left[\frac{\partial^2 f}{\partial t^2} + \frac{\partial^2 g}{\partial t^2} \right]. \end{aligned}$$

This has shown that if two linear wave functions are added algebraically, the resulting wave function is also linear. This wave function models the displacement of the medium of the resulting wave at each position along the x -axis. If two linear waves occupy the same medium, they are said to interfere. If these waves can be modeled with a linear wave function, these wave functions add to form the wave equation of the wave resulting from the interference of the individual waves. The displacement of the medium at every point of the resulting wave is the algebraic sum of the displacements due to the individual waves.

Taking this analysis a step further, if wave functions $y_1(x, t) = f(x \mp vt)$ and $y_2(x, t) = g(x \mp vt)$ are solutions to the linear wave equation, then $Ay_1(x, t) + By_2(x, t)$, where A and B are constants, is also a solution to the linear wave equation. This property is known as the principle of superposition. Interference and superposition are covered in more detail in [Interference of Waves](#).

Example:

Interference of Waves on a String

Consider a very long string held taut by two students, one on each end. Student A oscillates the end of the string producing a wave modeled with the wave function $y_1(x, t) = A \sin(kx - \omega t)$ and student B oscillates the string producing at twice the frequency, moving in the opposite direction. Both waves move at the same speed $v = \frac{\omega}{k}$. The two waves interfere to form a resulting wave whose wave function is

$y_R(x, t) = y_1(x, t) + y_2(x, t)$. Find the velocity of the resulting wave using the linear wave equation

$$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y(x, t)}{\partial t^2}.$$

Strategy

First, write the wave function for the wave created by the second student. Note that the angular frequency of the second wave is twice the frequency of the first wave (2ω), and since the velocity of the two waves are the same, the wave number of the second wave is twice that of the first wave ($2k$). Next, write the wave equation for the resulting wave function, which is the sum of the two individual wave functions. Then find the second partial derivative with respect to position and the second partial derivative with respect to time. Use the linear wave equation to find the velocity of the resulting wave.

Solution

1. Write the wave function of the second wave: $y_2(x, t) = A \sin(2kx + 2\omega t)$.

2. Write the resulting wave function:

Equation:

$$y_R(x, t) = y_1(x, t) + y_2(x, t) = A \sin(kx - \omega t) + A \sin(2kx + 2\omega t).$$

3. Find the partial derivatives:

Equation:

$$\begin{aligned} \frac{\partial y_R(x, t)}{\partial x} &= -Ak \cos(kx - \omega t) + 2Ak \cos(2kx + 2\omega t), \\ \frac{\partial^2 y_R(x, t)}{\partial x^2} &= -Ak^2 \sin(kx - \omega t) - 4Ak^2 \sin(2kx + 2\omega t), \\ \frac{\partial y_R(x, t)}{\partial t} &= -A\omega \cos(kx - \omega t) + 2A\omega \cos(2kx + 2\omega t), \\ \frac{\partial^2 y_R(x, t)}{\partial t^2} &= -A\omega^2 \sin(kx - \omega t) - 4A\omega^2 \sin(2kx + 2\omega t). \end{aligned}$$

4. Use the wave equation to find the velocity of the resulting wave:

$$\begin{aligned}
\frac{\partial^2 y(x,t)}{\partial x^2} &= \frac{1}{v^2} \frac{\partial^2 y(x,t)}{\partial t^2}, \\
-Ak^2 \sin(kx - \omega t) - 4Ak^2 \sin(2kx + 2\omega t) &= \frac{1}{v^2} (-A\omega^2 \sin(kx - \omega t) - 4A\omega^2 \sin(2kx + 2\omega t)), \\
k^2 (-A \sin(kx - \omega t) - 4A \sin(2kx + 2\omega t)) &= \frac{\omega^2}{v^2} (-A \sin(kx - \omega t) - 4A \sin(2kx + 2\omega t)), \\
k^2 &= \frac{\omega^2}{v^2}, |v| = \frac{\omega}{k}.
\end{aligned}$$

Significance

The speed of the resulting wave is equal to the speed of the original waves ($v = \frac{\omega}{k}$). We will show in the next section that the speed of a simple harmonic wave on a string depends on the tension in the string and the mass per length of the string. For this reason, it is not surprising that the component waves as well as the resultant wave all travel at the same speed.

Note:

Exercise:

Problem:

Check Your Understanding The wave equation $\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y(x,t)}{\partial t^2}$ works for any wave of the form $y(x,t) = f(x \mp vt)$. In the previous section, we stated that a cosine function could also be used to model a simple harmonic mechanical wave. Check if the wave

Equation:

$$y(x,t) = 0.50 \text{ m} \cos\left(0.20\pi \text{ m}^{-1}x - 4.00\pi \text{ s}^{-1}t + \frac{\pi}{10}\right)$$

is a solution to the wave equation.

Solution:

This wave, with amplitude $A = 0.5 \text{ m}$, wavelength $\lambda = 10.00 \text{ m}$, period $T = 0.50 \text{ s}$, is a solution to the wave equation with a wave velocity $v = 20.00 \text{ m/s}$.

Any disturbance that complies with the wave equation can propagate as a wave moving along the x -axis with a wave speed v . It works equally well for waves on a string, sound waves, and electromagnetic waves. This equation is extremely useful. For example, it can be used to show that electromagnetic waves move at the speed of light.

Summary

- A wave is an oscillation (of a physical quantity) that travels through a medium, accompanied by a transfer of energy. Energy transfers from one point to another in the direction of the wave motion. The particles of the medium oscillate up and down, back and forth, or both up and down and back and forth, around an equilibrium position.
- A snapshot of a sinusoidal wave at time $t = 0.00 \text{ s}$ can be modeled as a function of position. Two examples of such functions are $y(x) = A \sin(kx + \phi)$ and $y(x) = A \cos(kx + \phi)$.
- Given a function of a wave that is a snapshot of the wave, and is only a function of the position x , the motion of the pulse or wave moving at a constant velocity can be modeled with the function, replacing x with $x \mp vt$. The minus sign is for motion in the positive direction and the plus sign for the negative direction.

- The wave function is given by $y(x, t) = A \sin(kx - \omega t + \phi)$ where $k = 2\pi/\lambda$ is defined as the wave number, $\omega = 2\pi/T$ is the angular frequency, and ϕ is the phase shift.
- The wave moves with a constant velocity v_w , where the particles of the medium oscillate about an equilibrium position. The constant velocity of a wave can be found by $v = \frac{\lambda}{T} = \frac{\omega}{k}$.

Conceptual Questions

Exercise:

Problem:

If you were to shake the end of a taut spring up and down 10 times a second, what would be the frequency and the period of the sinusoidal wave produced on the spring?

Exercise:

Problem:

If you shake the end of a stretched spring up and down with a frequency f , you can produce a sinusoidal, transverse wave propagating down the spring. Does the wave number depend on the frequency you are shaking the spring?

Solution:

The wavelength is equal to the velocity of the wave times the frequency and the wave number is equal to $k = \frac{2\pi}{\lambda}$, so yes, the wave number will depend on the frequency and also depend on the velocity of the wave propagating through the spring.

Exercise:

Problem:

Does the vertical speed of a segment of a horizontal taut string through which a sinusoidal, transverse wave is propagating depend on the wave speed of the transverse wave?

Exercise:

Problem:

In this section, we have considered waves that move at a constant wave speed. Does the medium accelerate?

Solution:

The medium moves in simple harmonic motion as the wave propagates through the medium, continuously changing speed, therefore it accelerates. The acceleration of the medium is due to the restoring force of the medium, which acts in the opposite direction of the displacement.

Exercise:

Problem:

If you drop a pebble in a pond you may notice that several concentric ripples are produced, not just a single ripple. Why do you think that is?

Problems

Exercise:

Problem:

A pulse can be described as a single wave disturbance that moves through a medium. Consider a pulse that is defined at time $t = 0.00$ s by the equation $y(x) = \frac{6.00 \text{ m}^3}{x^2 + 2.00 \text{ m}^2}$ centered around $x = 0.00$ m. The pulse moves with a velocity of $v = 3.00$ m/s in the positive x -direction. (a) What is the amplitude of the pulse? (b) What is the equation of the pulse as a function of position and time? (c) Where is the pulse centered at time $t = 5.00$ s?

Exercise:**Problem:**

A transverse wave on a string is modeled with the wave function $y(x, t) = (0.20 \text{ cm})\sin(2.00 \text{ m}^{-1}x - 3.00 \text{ s}^{-1}t + \frac{\pi}{16})$. What is the height of the string with respect to the equilibrium position at a position $x = 4.00$ m and a time $t = 10.00$ s?

Solution:

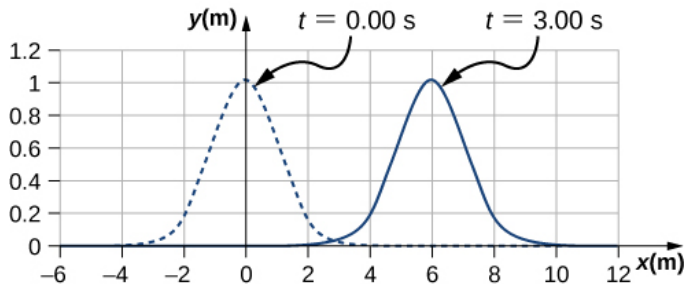
$$y(x, t) = -0.037 \text{ cm}$$

Exercise:**Problem:**

Consider the wave function $y(x, t) = (3.00 \text{ cm})\sin(0.4 \text{ m}^{-1}x + 2.00 \text{ s}^{-1}t + \frac{\pi}{10})$. What are the period, wavelength, speed, and initial phase shift of the wave modeled by the wave function?

Exercise:**Problem:**

A pulse is defined as $y(x, t) = e^{-2.77 \left(\frac{2.00(x - 2.00 \text{ m/s}(t))}{5.00 \text{ m}} \right)^2}$. Use a spreadsheet, or other computer program, to plot the pulse as the height of medium y as a function of position x . Plot the pulse at times $t = 0.00$ s and $t = 3.00$ s on the same graph. Where is the pulse centered at time $t = 3.00$ s? Use your spreadsheet to check your answer.

Solution:

The pulse will move $\Delta x = 6.00$ m.

Exercise:**Problem:**

A wave is modeled at time $t = 0.00$ s with a wave function that depends on position. The equation is $y(x) = (0.30 \text{ m})\sin(6.28 \text{ m}^{-1}x)$. The wave travels a distance of 4.00 meters in 0.50 s in the positive x -direction. Write an equation for the wave as a function of position and time.

Exercise:**Problem:**

A wave is modeled with the function $y(x, t) = (0.25 \text{ m})\cos(0.30 \text{ m}^{-1}x - 0.90 \text{ s}^{-1}t + \frac{\pi}{3})$. Find the (a) amplitude, (b) wave number, (c) angular frequency, (d) wave speed, (e) initial phase shift, (f) wavelength, and (g) period of the wave.

Solution:

a. $A = 0.25 \text{ m}$; b. $k = 0.30 \text{ m}^{-1}$; c. $\omega = 0.90 \text{ s}^{-1}$; d. $v = 3.0 \text{ m/s}$; e. $\phi = \pi/3 \text{ rad}$; f. $\lambda = 20.93 \text{ m}$; g. $T = 6.98 \text{ s}$

Exercise:**Problem:**

A surface ocean wave has an amplitude of 0.60 m and the distance from trough to trough is 8.00 m. It moves at a constant wave speed of 1.50 m/s propagating in the positive x -direction. At $t = 0$, the water displacement at $x = 0$ is zero, and v_y is positive. (a) Assuming the wave can be modeled as a sine wave, write a wave function to model the wave. (b) Use a spreadsheet to plot the wave function at times $t = 0.00 \text{ s}$ and $t = 2.00 \text{ s}$ on the same graph. Verify that the wave moves 3.00 m in those 2.00 s.

Exercise:**Problem:**

A wave is modeled by the wave function $y(x, t) = (0.30 \text{ m})\sin[\frac{2\pi}{4.50 \text{ m}}(x - 18.00 \frac{\text{m}}{\text{s}}t)]$. What are the amplitude, wavelength, wave speed, period, and frequency of the wave?

Solution:

$A = 0.30 \text{ m}$, $\lambda = 4.50 \text{ m}$, $v = 18.00 \text{ m/s}$, $f = 4.00 \text{ Hz}$, $T = 0.25 \text{ s}$

Exercise:**Problem:**

A transverse wave on a string is described with the wave function $y(x, t) = (0.50 \text{ cm})\sin(1.57 \text{ m}^{-1}x - 6.28 \text{ s}^{-1}t)$. (a) What is the wave velocity of the wave? (b) What is the magnitude of the maximum velocity of the string perpendicular to the direction of the motion?

Exercise:**Problem:**

A swimmer in the ocean observes one day that the ocean surface waves are periodic and resemble a sine wave. The swimmer estimates that the vertical distance between the crest and the trough of each wave is approximately 0.45 m, and the distance between each crest is approximately 1.8 m. The swimmer counts that 12 waves pass every two minutes. Determine the simple harmonic wave function that would describes these waves.

Solution:

$y(x, t) = 0.23 \text{ m} \sin(3.49 \text{ m}^{-1}x - 0.63 \text{ s}^{-1}t)$

Exercise:

Problem:

Consider a wave described by the wave function $y(x, t) = 0.3 \text{ m} \sin(2.00 \text{ m}^{-1}x - 628.00 \text{ s}^{-1}t)$. (a) How many crests pass by an observer at a fixed location in 2.00 minutes? (b) How far has the wave traveled in that time?

Exercise:**Problem:**

Consider two waves defined by the wave functions $y_1(x, t) = 0.50 \text{ m} \sin\left(\frac{2\pi}{3.00 \text{ m}}x + \frac{2\pi}{4.00 \text{ s}}t\right)$ and $y_2(x, t) = 0.50 \text{ m} \sin\left(\frac{2\pi}{6.00 \text{ m}}x - \frac{2\pi}{4.00 \text{ s}}t\right)$. What are the similarities and differences between the two waves?

Solution:

They have the same angular frequency, frequency, and period. They are traveling in opposite directions and $y_2(x, t)$ has twice the wavelength as $y_1(x, t)$ and is moving at half the wave speed.

Exercise:**Problem:**

Consider two waves defined by the wave functions $y_1(x, t) = 0.20 \text{ m} \sin\left(\frac{2\pi}{6.00 \text{ m}}x - \frac{2\pi}{4.00 \text{ s}}t\right)$ and $y_2(x, t) = 0.20 \text{ m} \cos\left(\frac{2\pi}{6.00 \text{ m}}x - \frac{2\pi}{4.00 \text{ s}}t\right)$. What are the similarities and differences between the two waves?

Exercise:**Problem:**

The speed of a transverse wave on a string is 300.00 m/s, its wavelength is 0.50 m, and the amplitude is 20.00 cm. How much time is required for a particle on the string to move through a distance of 5.00 km?

Solution:

Each particle of the medium moves a distance of $4A$ each period. The period can be found by dividing the velocity by the wavelength: $t = 10.42 \text{ s}$

Glossary

linear wave equation

equation describing waves that result from a linear restoring force of the medium; any function that is a solution to the wave equation describes a wave moving in the positive x -direction or the negative x -direction with a constant wave speed v

pulse

single disturbance that moves through a medium, transferring energy but not mass

wave function

mathematical model of the position of particles of the medium

wave number

$$\frac{2\pi}{\lambda}$$

Wave Speed on a Stretched String

By the end of this section, you will be able to:

- Determine the factors that affect the speed of a wave on a string
- Write a mathematical expression for the speed of a wave on a string and generalize these concepts for other media

The speed of a wave depends on the characteristics of the medium. For example, in the case of a guitar, the strings vibrate to produce the sound. The speed of the waves on the strings, and the wavelength, determine the frequency of the sound produced. The strings on a guitar have different thickness but may be made of similar material. They have different *linear densities*, where the linear density is defined as the mass per length,

Note:

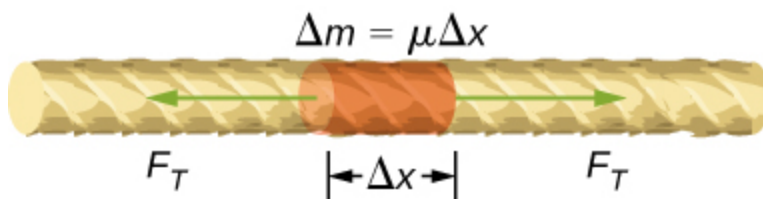
Equation:

$$\mu = \frac{\text{mass of string}}{\text{length of string}} = \frac{m}{l}.$$

In this chapter, we consider only string with a constant linear density. If the linear density is constant, then the mass (Δm) of a small length of string (Δx) is $\Delta m = \mu \Delta x$. For example, if the string has a length of 2.00 m and a mass of 0.06 kg, then the linear density is $\mu = \frac{0.06 \text{ kg}}{2.00 \text{ m}} = 0.03 \frac{\text{kg}}{\text{m}}$. If a 1.00-mm section is cut from the string, the mass of the 1.00-mm length is $\Delta m = \mu \Delta x = \left(0.03 \frac{\text{kg}}{\text{m}}\right) 0.001 \text{ m} = 3.00 \times 10^{-5} \text{ kg}$. The guitar also has a method to change the tension of the strings. The tension of the strings is adjusted by turning spindles, called the tuning pegs, around which the strings are wrapped. For the guitar, the linear density of the string and the tension in the string determine the speed of the waves in the string and the frequency of the sound produced is proportional to the wave speed.

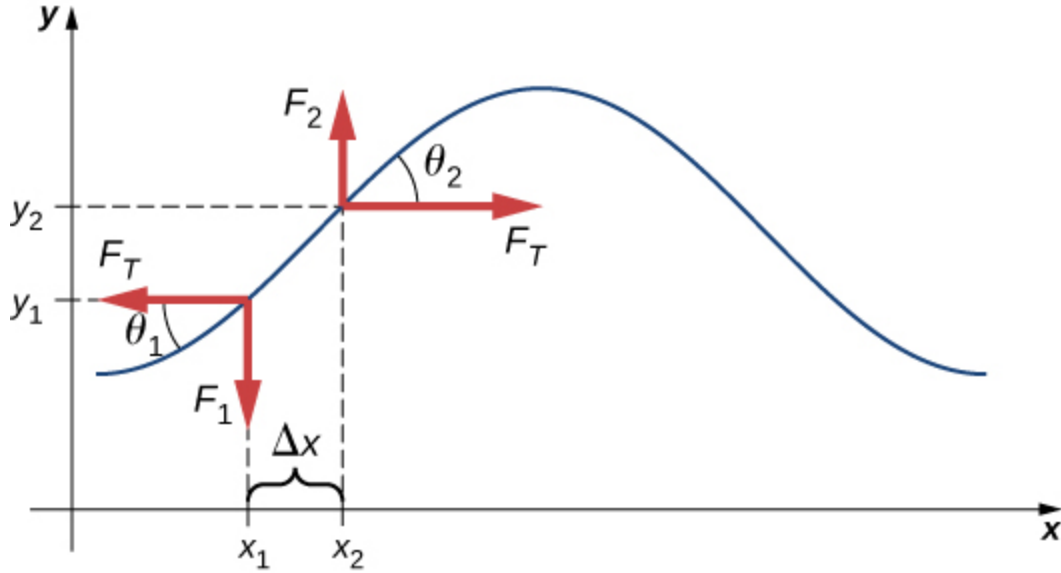
Wave Speed on a String under Tension

To see how the speed of a wave on a string depends on the tension and the linear density, consider a pulse sent down a taut string ([\[link\]](#)). When the taut string is at rest at the equilibrium position, the tension in the string F_T is constant. Consider a small element of the string with a mass equal to $\Delta m = \mu \Delta x$. The mass element is at rest and in equilibrium and the force of tension of either side of the mass element is equal and opposite.



Mass element of a string kept taut with a tension F_T . The mass element is in static equilibrium, and the force of tension acting on either side of the mass element is equal in magnitude and opposite in direction.

If you pluck a string under tension, a transverse wave moves in the positive x -direction, as shown in [\[link\]](#). The mass element is small but is enlarged in the figure to make it visible. The small mass element oscillates perpendicular to the wave motion as a result of the restoring force provided by the string and does not move in the x -direction. The tension F_T in the string, which acts in the positive and negative x -direction, is approximately constant and is independent of position and time.



A string under tension is plucked, causing a pulse to move along the string in the positive x -direction.

Assume that the inclination of the displaced string with respect to the horizontal axis is small. The net force on the element of the string, acting parallel to the string, is the sum of the tension in the string and the restoring force. The x -components of the force of tension cancel, so the net force is equal to the sum of the y -components of the force. The magnitude of the x -component of the force is equal to the horizontal force of tension of the string F_T as shown in [\[link\]](#). To obtain the y -components of the force, note that $\tan \theta_1 = \frac{-F_1}{F_T}$ and $\tan \theta_2 = \frac{F_2}{F_T}$. The $\tan \theta$ is equal to the slope of a function at a point, which is equal to the partial derivative of y with respect to x at that point. Therefore, $\frac{F_1}{F_T}$ is equal to the negative slope of the string at x_1 and $\frac{F_2}{F_T}$ is equal to the slope of the string at x_2 :

Equation:

$$\frac{F_1}{F_T} = -\left(\frac{\partial y}{\partial x}\right)_{x_1} \text{ and } \frac{F_2}{F_T} = \left(\frac{\partial y}{\partial x}\right)_{x_2}.$$

The net force is on the small mass element can be written as

Equation:

$$F_{\text{net}} = F_1 + F_2 = F_T \left[\left(\frac{\partial y}{\partial x} \right)_{x_2} - \left(\frac{\partial y}{\partial x} \right)_{x_1} \right].$$

Using Newton's second law, the net force is equal to the mass times the acceleration. The linear density of the string μ is the mass per length of the string, and the mass of the portion of the string is $\mu\Delta x$,

Equation:

$$\begin{aligned} F_T \left[\left(\frac{\partial y}{\partial x} \right)_{x_2} - \left(\frac{\partial y}{\partial x} \right)_{x_1} \right] &= \Delta m a, \\ F_T \left[\left(\frac{\partial y}{\partial x} \right)_{x_2} - \left(\frac{\partial y}{\partial x} \right)_{x_1} \right] &= \mu \Delta x \frac{\partial^2 y}{\partial t^2}. \end{aligned}$$

Dividing by $F_T\Delta x$ and taking the limit as Δx approaches zero,

Equation:

$$\begin{aligned} \frac{\left[\left(\frac{\partial y}{\partial x} \right)_{x_2} - \left(\frac{\partial y}{\partial x} \right)_{x_1} \right]}{\Delta x} &= \frac{\mu}{F_T} \frac{\partial^2 y}{\partial t^2} \\ \lim_{\Delta x \rightarrow 0} \frac{\left[\left(\frac{\partial y}{\partial x} \right)_{x_2} - \left(\frac{\partial y}{\partial x} \right)_{x_1} \right]}{\Delta x} &= \frac{\mu}{F_T} \frac{\partial^2 y}{\partial t^2} \\ \frac{\partial^2 y}{\partial x^2} &= \frac{\mu}{F_T} \frac{\partial^2 y}{\partial t^2}. \end{aligned}$$

Recall that the linear wave equation is

Equation:

$$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y(x, t)}{\partial t^2}.$$

Therefore,
Equation:

$$\frac{1}{v^2} = \frac{\mu}{F_T}.$$

Solving for v , we see that the speed of the wave on a string depends on the tension and the linear density.

Note:

Speed of a Wave on a String Under Tension

The speed of a pulse or wave on a string under tension can be found with the equation

Equation:

$$|v| = \sqrt{\frac{F_T}{\mu}}$$

where F_T is the tension in the string and μ is the mass per length of the string.

Example:

The Wave Speed of a Guitar String

On a six-string guitar, the high E string has a linear density of

$\mu_{\text{High E}} = 3.09 \times 10^{-4} \text{ kg/m}$ and the low E string has a linear density of

$\mu_{\text{Low E}} = 5.78 \times 10^{-3} \text{ kg/m}$. (a) If the high E string is plucked,

producing a wave in the string, what is the speed of the wave if the tension

of the string is 56.40 N? (b) The linear density of the low E string is

approximately 20 times greater than that of the high E string. For waves to

travel through the low E string at the same wave speed as the high E,

would the tension need to be larger or smaller than the high E string? What

would be the approximate tension? (c) Calculate the tension of the low E string needed for the same wave speed.

Strategy

- The speed of the wave can be found from the linear density and the tension $v = \sqrt{\frac{F_T}{\mu}}$.
- From the equation $v = \sqrt{\frac{F_T}{\mu}}$, if the linear density is increased by a factor of almost 20, the tension would need to be increased by a factor of 20.
- Knowing the velocity and the linear density, the velocity equation can be solved for the force of tension $F_T = \mu v^2$.

Solution

- Use the velocity equation to find the speed:

Equation:

$$v = \sqrt{\frac{F_T}{\mu}} = \sqrt{\frac{56.40 \text{ N}}{3.09 \times 10^{-4} \text{ kg/m}}} = 427.23 \text{ m/s}.$$

- The tension would need to be increased by a factor of approximately 20. The tension would be slightly less than 1128 N.
- Use the velocity equation to find the actual tension:

Equation:

$$F_T = \mu v^2 = 5.78 \times 10^{-3} \text{ kg/m} (427.23 \text{ m/s})^2 = 1055.00 \text{ N}.$$

This solution is within 7% of the approximation.

Significance

The standard notes of the six string (high E, B, G, D, A, low E) are tuned to vibrate at the fundamental frequencies (329.63 Hz, 246.94Hz, 196.00Hz, 146.83Hz, 110.00Hz, and 82.41Hz) when plucked. The frequencies depend on the speed of the waves on the string and the wavelength of the waves. The six strings have different linear densities and are “tuned” by changing the tensions in the strings. We will see in [Interference of Waves](#) that the

wavelength depends on the length of the strings and the boundary conditions. To play notes other than the fundamental notes, the lengths of the strings are changed by pressing down on the strings.

Note:

Exercise:

Problem:

Check Your Understanding The wave speed of a wave on a string depends on the tension and the linear mass density. If the tension is doubled, what happens to the speed of the waves on the string?

Solution:

Since the speed of a wave on a taut string is proportional to the square root of the tension divided by the linear density, the wave speed would increase by $\sqrt{2}$.

Speed of Compression Waves in a Fluid

The speed of a wave on a string depends on the square root of the tension divided by the mass per length, the linear density. In general, the speed of a wave through a medium depends on the elastic property of the medium and the inertial property of the medium.

Equation:

$$|v| = \sqrt{\frac{\text{elastic property}}{\text{inertial property}}}$$

The elastic property describes the tendency of the particles of the medium to return to their initial position when perturbed. The inertial property

describes the tendency of the particle to resist changes in velocity.

The speed of a longitudinal wave through a liquid or gas depends on the density of the fluid and the bulk modulus of the fluid,

Note:

Equation:

$$v = \sqrt{\frac{B}{\rho}}.$$

Here the bulk modulus is defined as $B = -\frac{\Delta P}{\Delta V/V_0}$, where ΔP is the change in the pressure and the denominator is the ratio of the change in volume to the initial volume, and $\rho \equiv \frac{m}{V}$ is the mass per unit volume. For example, sound is a mechanical wave that travels through a fluid or a solid. The speed of sound in air with an atmospheric pressure of $1.013 \times 10^5 \text{ Pa}$ and a temperature of 20°C is $v_s \approx 343.00 \text{ m/s}$. Because the density depends on temperature, the speed of sound in air depends on the temperature of the air. This will be discussed in detail in [Sound](#).

Summary

- The speed of a wave on a string depends on the linear density of the string and the tension in the string. The linear density is mass per unit length of the string.
- In general, the speed of a wave depends on the square root of the ratio of the elastic property to the inertial property of the medium.
- The speed of a wave through a fluid is equal to the square root of the ratio of the bulk modulus of the fluid to the density of the fluid.
- The speed of sound through air at $T = 20^\circ \text{C}$ is approximately $v_s = 343.00 \text{ m/s}$.

Conceptual Questions

Exercise:

Problem:

If the tension in a string were increased by a factor of four, by what factor would the wave speed of a wave on the string increase?

Solution:

The wave speed is proportional to the square root of the tension, so the speed is doubled.

Exercise:

Problem:

Does a sound wave move faster in seawater or fresh water, if both the sea water and fresh water are at the same temperature and the sound wave moves near the surface?

$$\left(\rho_w \approx 1000 \frac{\text{kg}}{\text{m}^3}, \rho_s \approx 1030 \frac{\text{kg}}{\text{m}^3}, B_w = 2.15 \times 10^9 \text{ Pa}, B_s = 2.34 \times 10^9 \text{ Pa} \right)$$

Exercise:

Problem:

Guitars have strings of different linear mass density. If the lowest density string and the highest density string are under the same tension, which string would support waves with the higher wave speed?

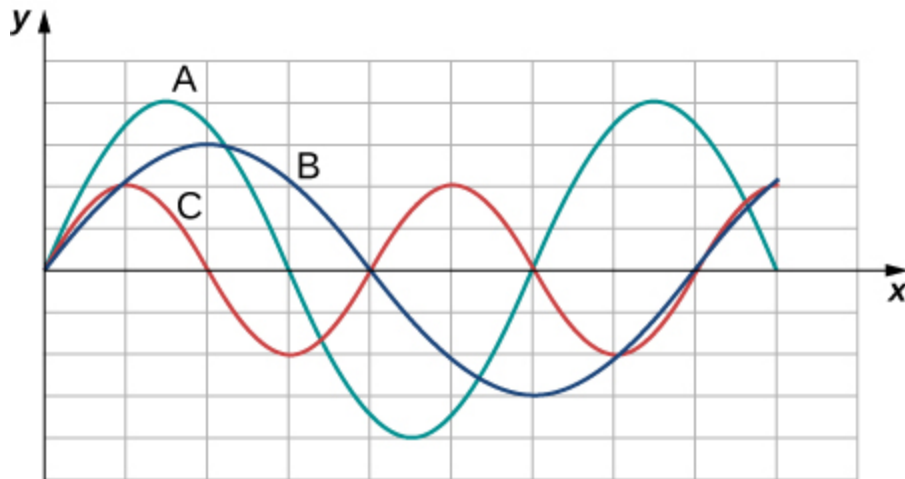
Solution:

Since the speed of a wave on a string is inversely proportional to the square root of the linear mass density, the speed would be higher in the low linear mass density of the string.

Exercise:

Problem:

Shown below are three waves that were sent down a string at different times. The tension in the string remains constant. (a) Rank the waves from the smallest wavelength to the largest wavelength. (b) Rank the waves from the lowest frequency to the highest frequency.

**Exercise:****Problem:**

Electrical power lines connected by two utility poles are sometimes heard to hum when driven into oscillation by the wind. The speed of the waves on the power lines depend on the tension. What provides the tension in the power lines?

Solution:

The tension in the wire is due to the weight of the electrical power cable.

Exercise:

Problem:

Two strings, one with a low mass density and one with a high linear density are spliced together. The higher density end is tied to a lab post and a student holds the free end of the low-mass density string. The student gives the string a flip and sends a pulse down the strings. If the tension is the same in both strings, does the pulse travel at the same wave velocity in both strings? If not, where does it travel faster, in the low density string or the high density string?

Problems**Exercise:****Problem:**

Transverse waves are sent along a 5.00-m-long string with a speed of 30.00 m/s. The string is under a tension of 10.00 N. What is the mass of the string?

Exercise:**Problem:**

A copper wire has a density of $\rho = 8920 \text{ kg/m}^3$, a radius of 1.20 mm, and a length L . The wire is held under a tension of 10.00 N. Transverse waves are sent down the wire. (a) What is the linear mass density of the wire? (b) What is the speed of the waves through the wire?

Solution:

a. $\mu = 0.040 \text{ kg/m}$; b. $v = 15.75 \text{ m/s}$

Exercise:

Problem:

A piano wire has a linear mass density of $\mu = 4.95 \times 10^{-3} \text{ kg/m}$. Under what tension must the string be kept to produce waves with a wave speed of 500.00 m/s?

Exercise:**Problem:**

A string with a linear mass density of $\mu = 0.0060 \text{ kg/m}$ is tied to the ceiling. A 20-kg mass is tied to the free end of the string. The string is plucked, sending a pulse down the string. Estimate the speed of the pulse as it moves down the string.

Solution:

$$v = 180 \text{ m/s}$$

Exercise:**Problem:**

A cord has a linear mass density of $\mu = 0.0075 \text{ kg/m}$ and a length of three meters. The cord is plucked and it takes 0.20 s for the pulse to reach the end of the string. What is the tension of the string?

Exercise:**Problem:**

A string is 3.00 m long with a mass of 5.00 g. The string is held taut with a tension of 500.00 N applied to the string. A pulse is sent down the string. How long does it take the pulse to travel the 3.00 m of the string?

Solution:

$$v = 547.723 \text{ m/s}, \Delta t = 5.48 \text{ ms}$$

Exercise:

Problem:

Two strings are attached to poles, however the first string is twice as long as the second. If both strings have the same tension and μ , what is the ratio of the speed of the pulse of the wave from the first string to the second string?

Exercise:**Problem:**

Two strings are attached to poles, however the first string is twice the linear mass density μ of the second. If both strings have the same tension, what is the ratio of the speed of the pulse of the wave from the first string to the second string?

Solution:

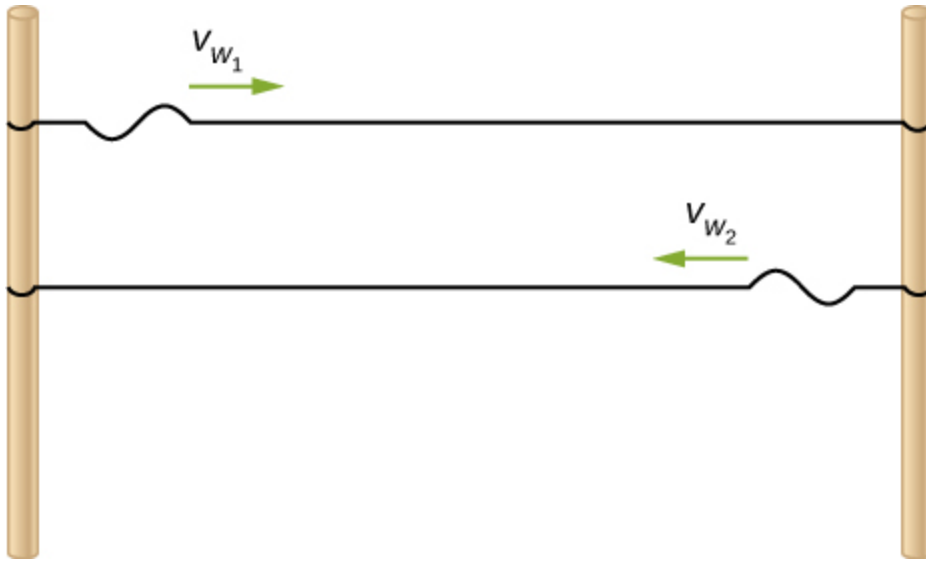
0.707

Exercise:**Problem:**

Transverse waves travel through a string where the tension equals 7.00 N with a speed of 20.00 m/s. What tension would be required for a wave speed of 25.00 m/s?

Exercise:**Problem:**

Two strings are attached between two poles separated by a distance of 2.00 m as shown below, both under the same tension of 600.00 N. String 1 has a linear density of $\mu_1 = 0.0025 \text{ kg/m}$ and string 2 has a linear mass density of $\mu_2 = 0.0035 \text{ kg/m}$. Transverse wave pulses are generated simultaneously at opposite ends of the strings. How much time passes before the pulses pass one another?



Solution:

$$v_1 t + v_2 t = 2.00 \text{ m}, \quad t = 1.69 \text{ ms}$$

Exercise:**Problem:**

Two strings are attached between two poles separated by a distance of 2.00 meters as shown in the preceding figure, both strings have a linear density of $\mu_1 = 0.0025 \text{ kg/m}$, the tension in string 1 is 600.00 N and the tension in string 2 is 700.00 N. Transverse wave pulses are generated simultaneously at opposite ends of the strings. How much time passes before the pulses pass one another?

Exercise:**Problem:**

The note E_4 is played on a piano and has a frequency of $f = 393.88$. If the linear mass density of this string of the piano is $\mu = 0.012 \text{ kg/m}$ and the string is under a tension of 1000.00 N, what is the speed of the wave on the string and the wavelength of the wave?

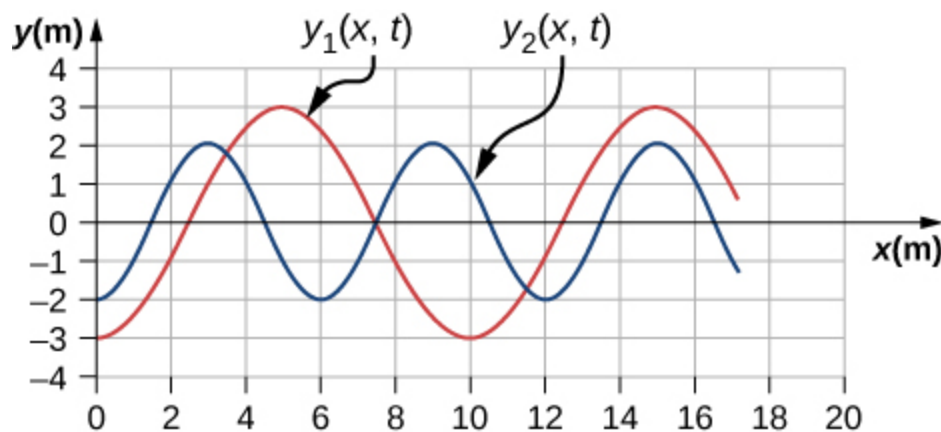
Solution:

$$v = 288.68 \text{ m/s}, \lambda = 0.73 \text{ m}$$

Exercise:

Problem:

Two transverse waves travel through a taut string. The speed of each wave is $v = 30.00 \text{ m/s}$. A plot of the vertical position as a function of the horizontal position is shown below for the time $t = 0.00 \text{ s}$. (a) What is the wavelength of each wave? (b) What is the frequency of each wave? (c) What is the maximum vertical speed of each string?



Exercise:

Problem:

A sinusoidal wave travels down a taut, horizontal string with a linear mass density of $\mu = 0.060 \text{ kg/m}$. The maximum vertical speed of the wave is $v_{y \text{ max}} = 0.30 \text{ cm/s}$. The wave is modeled with the wave equation $y(x, t) = A \sin(6.00 \text{ m}^{-1}x - 24.00 \text{ s}^{-1}t)$. (a) What is the amplitude of the wave? (b) What is the tension in the string?

Solution:

a. $A = 0.0125 \text{ cm}$; b. $F_T = 0.96 \text{ N}$

Exercise:

Problem:

The speed of a transverse wave on a string is $v = 60.00 \text{ m/s}$ and the tension in the string is $F_T = 100.00 \text{ N}$. What must the tension be to increase the speed of the wave to $v = 120.00 \text{ m/s}$?

Energy and Power of a Wave

By the end of this section, you will be able to:

- Explain how energy travels with a pulse or wave
- Describe, using a mathematical expression, how the energy in a wave depends on the amplitude of the wave

All waves carry energy, and sometimes this can be directly observed. Earthquakes can shake whole cities to the ground, performing the work of thousands of wrecking balls ([link](#)). Loud sounds can pulverize nerve cells in the inner ear, causing permanent hearing loss. Ultrasound is used for deep-heat treatment of muscle strains. A laser beam can burn away a malignancy. Water waves chew up beaches.



The destructive effect of an earthquake is observable evidence of the energy carried in these waves. The Richter scale rating of earthquakes is a logarithmic scale related to both their amplitude and the energy they carry.

In this section, we examine the quantitative expression of energy in waves. This will be of fundamental importance in later discussions of waves, from sound to light to quantum mechanics.

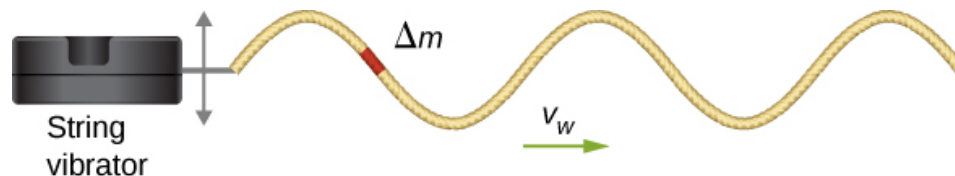
Energy in Waves

The amount of energy in a wave is related to its amplitude and its frequency. Large-amplitude earthquakes produce large ground displacements. Loud sounds have high-pressure amplitudes and come from larger-amplitude source vibrations than soft sounds. Large ocean breakers churn up the shore more than small ones. Consider the example of the seagull and the water wave earlier in the chapter ([link](#)). Work is done on the seagull by the wave as the seagull is moved up, changing its potential energy. The larger the amplitude, the higher the seagull is lifted by the wave and the larger the change in potential energy.

The energy of the wave depends on both the amplitude and the frequency. If the energy of each wavelength is considered to be a discrete packet of energy, a high-frequency wave will deliver more of these packets per unit time than a low-frequency wave. We will see that the average rate of energy transfer in mechanical waves is proportional to both the square of the amplitude and the square of the frequency. If two mechanical waves have equal amplitudes, but one wave has a frequency equal to twice the frequency of the other, the higher-frequency wave will have a rate of energy transfer a factor of four times as great as the rate of energy transfer of the lower-frequency wave. It should be noted that although the rate of energy transport is proportional to both the square of the amplitude and square of the frequency in mechanical waves, the rate of energy transfer in electromagnetic waves is proportional to the square of the amplitude, but independent of the frequency.

Power in Waves

Consider a sinusoidal wave on a string that is produced by a string vibrator, as shown in [link](#). The string vibrator is a device that vibrates a rod up and down. A string of uniform linear mass density is attached to the rod, and the rod oscillates the string, producing a sinusoidal wave. The rod does work on the string, producing energy that propagates along the string. Consider a mass element of the string with a mass Δm , as seen in [link](#). As the energy propagates along the string, each mass element of the string is driven up and down at the same frequency as the wave. Each mass element of the string can be modeled as a simple harmonic oscillator. Since the string has a constant linear density $\mu = \frac{\Delta m}{\Delta x}$, each mass element of the string has the mass $\Delta m = \mu \Delta x$.



A string vibrator is a device that vibrates a rod. A string is attached to the rod, and the rod does work on the string, driving the string up and down. This produces a sinusoidal wave in the string, which moves with a wave velocity v . The wave speed depends on the

tension in the string and the linear mass density of the string. A section of the string with mass Δm oscillates at the same frequency as the wave.

The total mechanical energy of the wave is the sum of its kinetic energy and potential energy. The kinetic energy $K = \frac{1}{2}mv^2$ of each mass element of the string of length Δx is $\Delta K = \frac{1}{2}(\Delta m)v_y^2$, as the mass element oscillates perpendicular to the direction of the motion of the wave. Using the constant linear mass density, the kinetic energy of each mass element of the string with length Δx is

Equation:

$$\Delta K = \frac{1}{2}(\mu \Delta x)v_y^2.$$

A differential equation can be formed by letting the length of the mass element of the string approach zero,

Equation:

$$dK = \lim_{\Delta x \rightarrow 0} \frac{1}{2}(\mu \Delta x)v_y^2 = \frac{1}{2}(\mu dx)v_y^2.$$

Since the wave is a sinusoidal wave with an angular frequency ω , the position of each mass element may be modeled as $y(x, t) = A \sin(kx - \omega t)$. Each mass element of the string oscillates with a velocity $v_y = \frac{\partial y(x, t)}{\partial t} = -A\omega \cos(kx - \omega t)$. The kinetic energy of each mass element of the string becomes

Equation:

$$\begin{aligned} dK &= \frac{1}{2}(\mu dx)(-A\omega \cos(kx - \omega t))^2, \\ &= \frac{1}{2}(\mu dx)A^2\omega^2 \cos^2(kx - \omega t). \end{aligned}$$

The wave can be very long, consisting of many wavelengths. To standardize the energy, consider the kinetic energy associated with a wavelength of the wave. This kinetic energy can be integrated over the wavelength to find the energy associated with each wavelength of the wave:

Equation:

$$\begin{aligned}
dK &= \frac{1}{2}(\mu dx)A^2\omega^2 \cos^2(kx), \\
\int_0^{K_\lambda} dK &= \int_0^\lambda \frac{1}{2}\mu A^2\omega^2 \cos^2(kx)dx = \frac{1}{2}\mu A^2\omega^2 \int_0^\lambda \cos^2(kx)dx, \\
K_\lambda &= \frac{1}{2}\mu A^2\omega^2 \left[\frac{1}{2}x + \frac{1}{4k}\sin(2kx) \right]_0^\lambda = \frac{1}{2}\mu A^2\omega^2 \left[\frac{1}{2}\lambda + \frac{1}{4k}\sin(2k\lambda) - \frac{1}{4k}\sin(0) \right], \\
K_\lambda &= \frac{1}{4}\mu A^2\omega^2 \lambda.
\end{aligned}$$

There is also potential energy associated with the wave. Much like the mass oscillating on a spring, there is a conservative restoring force that, when the mass element is displaced from the equilibrium position, drives the mass element back to the equilibrium position. The potential energy of the mass element can be found by considering the linear restoring force of the string. In [Oscillations](#), we saw that the potential energy stored in a spring with a linear restoring force is equal to $U = \frac{1}{2}k_s x^2$, where the equilibrium position is defined as $x = 0.00$ m. When a mass attached to the spring oscillates in simple harmonic motion, the angular frequency is equal to $\omega = \sqrt{\frac{k_s}{m}}$. As each mass element oscillates in simple harmonic motion, the spring constant is equal to $k_s = \Delta m \omega^2$. The potential energy of the mass element is equal to

Equation:

$$\Delta U = \frac{1}{2}k_s x^2 = \frac{1}{2}\Delta m \omega^2 x^2.$$

Note that k_s is the spring constant and not the wave number $k = \frac{2\pi}{\lambda}$. This equation can be used to find the energy over a wavelength. Integrating over the wavelength, we can compute the potential energy over a wavelength:

Equation:

$$\begin{aligned}
dU &= \frac{1}{2}k_s x^2 = \frac{1}{2}\mu \omega^2 x^2 dx, \\
U_\lambda &= \frac{1}{2}\mu \omega^2 A^2 \int_0^\lambda \sin^2(kx)dx = \frac{1}{4}\mu A^2 \omega^2 \lambda.
\end{aligned}$$

The potential energy associated with a wavelength of the wave is equal to the kinetic energy associated with a wavelength.

The total energy associated with a wavelength is the sum of the potential energy and the kinetic energy:

Equation:

$$E_{\lambda} = U_{\lambda} + K_{\lambda},$$

$$E_{\lambda} = \frac{1}{4}\mu A^2\omega^2\lambda + \frac{1}{4}\mu A^2\omega^2\lambda = \frac{1}{2}\mu A^2\omega^2\lambda.$$

The time-averaged power of a sinusoidal mechanical wave, which is the average rate of energy transfer associated with a wave as it passes a point, can be found by taking the total energy associated with the wave divided by the time it takes to transfer the energy. If the velocity of the sinusoidal wave is constant, the time for one wavelength to pass by a point is equal to the period of the wave, which is also constant. For a sinusoidal mechanical wave, the time-averaged power is therefore the energy associated with a wavelength divided by the period of the wave. The wavelength of the wave divided by the period is equal to the velocity of the wave,

Note:

Equation:

$$P_{\text{ave}} = \frac{E_{\lambda}}{T} = \frac{1}{2}\mu A^2\omega^2 \frac{\lambda}{T} = \frac{1}{2}\mu A^2\omega^2 v.$$

Note that this equation for the time-averaged power of a sinusoidal mechanical wave shows that the power is proportional to the square of the amplitude of the wave and to the square of the angular frequency of the wave. Recall that the angular frequency is equal to $\omega = 2\pi f$, so the power of a mechanical wave is equal to the square of the amplitude and the square of the frequency of the wave.

Example:

Power Supplied by a String Vibrator

Consider a two-meter-long string with a mass of 70.00 g attached to a string vibrator as illustrated in [\[link\]](#). The tension in the string is 90.0 N. When the string vibrator is turned on, it oscillates with a frequency of 60 Hz and produces a sinusoidal wave on the string with an amplitude of 4.00 cm and a constant wave speed. What is the time-averaged power supplied to the wave by the string vibrator?

Strategy

The power supplied to the wave should equal the time-averaged power of the wave on the string. We know the mass of the string (m_s), the length of the string (L_s), and the tension (F_T) in the string. The speed of the wave on the string can be derived from the linear mass density and the tension. The string oscillates with the same frequency as the string vibrator, from which we can find the angular frequency.

Solution

1. Begin with the equation of the time-averaged power of a sinusoidal wave on a string:
Equation:

$$P = \frac{1}{2} \mu A^2 \omega^2 v.$$

The amplitude is given, so we need to calculate the linear mass density of the string, the angular frequency of the wave on the string, and the speed of the wave on the string.

2. We need to calculate the linear density to find the wave speed:

Equation:

$$\mu = \frac{m_s}{L_s} = \frac{0.070 \text{ kg}}{2.00 \text{ m}} = 0.035 \text{ kg/m}.$$

3. The wave speed can be found using the linear mass density and the tension of the string:

Equation:

$$v = \sqrt{\frac{F_T}{\mu}} = \sqrt{\frac{90.00 \text{ N}}{0.035 \text{ kg/m}}} = 50.71 \text{ m/s}.$$

4. The angular frequency can be found from the frequency:

Equation:

$$\omega = 2\pi f = 2\pi (60 \text{ s}^{-1}) = 376.80 \text{ s}^{-1}.$$

5. Calculate the time-averaged power:

Equation:

$$P = \frac{1}{2} \mu A^2 \omega^2 v = \frac{1}{2} \left(0.035 \frac{\text{kg}}{\text{m}} \right) (0.040 \text{ m})^2 (376.80 \text{ s}^{-1})^2 \left(50.71 \frac{\text{m}}{\text{s}} \right) = 201.59 \text{ W}.$$

Significance

The time-averaged power of a sinusoidal wave is proportional to the square of the amplitude of the wave and the square of the angular frequency of the wave. This is true for most mechanical waves. If either the angular frequency or the amplitude of the wave were doubled, the power would increase by a factor of four. The time-averaged power of the wave on a string is also proportional to the speed of the sinusoidal wave on the string. If the speed were doubled, by increasing the tension by a factor of four, the power would also be doubled.

Note:

Exercise:

Problem:

Check Your Understanding Is the time-averaged power of a sinusoidal wave on a string proportional to the linear density of the string?

Solution:

At first glance, the time-averaged power of a sinusoidal wave on a string may look proportional to the linear density of the string because $P = \frac{1}{2}\mu A^2\omega^2 v$; however, the speed of the wave depends on the linear density. Replacing the wave speed with $\sqrt{\frac{F_T}{\mu}}$ shows that the power is proportional to the square root of tension and proportional to the square root of the linear mass density:

$$P = \frac{1}{2}\mu A^2\omega^2 v = \frac{1}{2}\mu A^2\omega^2 \sqrt{\frac{F_T}{\mu}} = \frac{1}{2} A^2\omega^2 \sqrt{\mu F_T}.$$

The equations for the energy of the wave and the time-averaged power were derived for a sinusoidal wave on a string. In general, the energy of a mechanical wave and the power are proportional to the amplitude squared and to the angular frequency squared (and therefore the frequency squared).

Another important characteristic of waves is the intensity of the waves. Waves can also be concentrated or spread out. Waves from an earthquake, for example, spread out over a larger area as they move away from a source, so they do less damage the farther they get from the source. Changing the area the waves cover has important effects. All these pertinent factors are included in the definition of **intensity (*I*)** as power per unit area:

Note:**Equation:**

$$I = \frac{P}{A},$$

where P is the power carried by the wave through area A . The definition of intensity is valid for any energy in transit, including that carried by waves. The SI unit for intensity is watts per square meter (W/m^2). Many waves are spherical waves that move out from a source as a sphere. For example, a sound speaker mounted on a post above the ground may produce sound waves that move away from the source as a spherical wave. Sound waves are discussed in more detail in the next chapter, but in general, the farther you are from the speaker, the less intense the sound you hear. As a spherical wave moves out from a source, the surface area of

the wave increases as the radius increases ($A = 4\pi r^2$). The intensity for a spherical wave is therefore

Note:

Equation:

$$I = \frac{P}{4\pi r^2}.$$

If there are no dissipative forces, the energy will remain constant as the spherical wave moves away from the source, but the intensity will decrease as the surface area increases.

In the case of the two-dimensional circular wave, the wave moves out, increasing the circumference of the wave as the radius of the circle increases. If you toss a pebble in a pond, the surface ripple moves out as a circular wave. As the ripple moves away from the source, the amplitude decreases. The energy of the wave spreads around a larger circumference and the amplitude decreases proportional to $\frac{1}{r}$, which is also the same in the case of a spherical wave, since intensity is proportional to the amplitude squared.

Summary

- The energy and power of a wave are proportional to the square of the amplitude of the wave and the square of the angular frequency of the wave.
- The time-averaged power of a sinusoidal wave on a string is found by $P_{\text{ave}} = \frac{1}{2}\mu A^2\omega^2v$, where μ is the linear mass density of the string, A is the amplitude of the wave, ω is the angular frequency of the wave, and v is the speed of the wave.
- Intensity is defined as the power divided by the area. In a spherical wave, the area is $A = 4\pi r^2$ and the intensity is $I = \frac{P}{4\pi r^2}$. As the wave moves out from a source, the energy is conserved, but the intensity decreases as the area increases.

Conceptual Questions

Exercise:

Problem:

Consider a string with under tension with a constant linear mass density. A sinusoidal wave with an angular frequency and amplitude produced by some external driving force. If the frequency of the driving force is decreased to half of the original frequency, how is the time-averaged power of the wave affected? If the amplitude of the driving force is decreased by half, how is the time-averaged power affected? Explain your answer.

Solution:

The time averaged power is $P = \frac{E\lambda}{T} = \frac{1}{2}\mu A^2 \omega^2 \frac{\lambda}{T} = \frac{1}{2}\mu A^2 \omega^2 v$. If the frequency or amplitude is halved, the power decreases by a factor of 4.

Exercise:**Problem:**

Circular water waves decrease in amplitude as they move away from where a rock is dropped. Explain why.

Exercise:**Problem:**

In a transverse wave on a string, the motion of the string is perpendicular to the motion of the wave. If this is so, how is possible to move energy along the length of the string?

Solution:

As a portion on the string moves vertically, it exerts a force on the neighboring portion of the string, doing work on the portion and transferring the energy.

Exercise:**Problem:**

The energy from the sun warms the portion of the earth facing the sun during the daylight hours. Why are the North and South Poles cold while the equator is quite warm?

Exercise:**Problem:**

The intensity of a spherical waves decreases as the wave moves away from the source. If the intensity of the wave at the source is I_0 , how far from the source will the intensity decrease by a factor of nine?

Solution:

The intensity of a spherical wave is $I = \frac{P}{4\pi r^2}$, if no energy is dissipated the intensity will decrease by a factor of nine at three meters.

Problems**Exercise:**

Problem:

A string of length 5 m and a mass of 90 g is held under a tension of 100 N. A wave travels down the string that is modeled as

$y(x, t) = 0.01 \text{ m} \sin(15.7 \text{ m}^{-1}x - 1170.12 \text{ s}^{-1}t)$. What is the power over one wavelength?

Solution:

$$v = 74.54 \text{ m/s}, P_{\lambda} = 91.85 \text{ W}$$

Exercise:**Problem:**

Ultrasound of intensity $1.50 \times 10^2 \text{ W/m}^2$ is produced by the rectangular head of a medical imaging device measuring 3.00 cm by 5.00 cm. What is its power output?

Exercise:**Problem:**

The low-frequency speaker of a stereo set has a surface area of $A = 0.05 \text{ m}^2$ and produces 1 W of acoustical power. (a) What is the intensity at the speaker? (b) If the speaker projects sound uniformly in all directions, at what distance from the speaker is the intensity 0.1 W/m^2 ?

Solution:

$$\text{a. } I = 20.0 \text{ W/m}^2; \text{ b. } I = \frac{P}{A}, A = 10.0 \text{ m}^2 \\ A = 4\pi r^2, r = 0.892 \text{ m}$$

Exercise:**Problem:**

To increase the intensity of a wave by a factor of 50, by what factor should the amplitude be increased?

Exercise:**Problem:**

A device called an insolation meter is used to measure the intensity of sunlight. It has an area of 100 cm^2 and registers 6.50 W. What is the intensity in W/m^2 ?

Solution:

$$I = 650 \text{ W/m}^2$$

Exercise:**Problem:**

Energy from the Sun arrives at the top of Earth's atmosphere with an intensity of 1400 W/m^2 . How long does it take for $1.80 \times 10^9 \text{ J}$ to arrive on an area of 1.00 m^2 ?

Exercise:**Problem:**

Suppose you have a device that extracts energy from ocean breakers in direct proportion to their intensity. If the device produces 10.0 kW of power on a day when the breakers are 1.20 m high, how much will it produce when they are 0.600 m high?

Solution:

$$P \propto E \propto I \propto X^2 \Rightarrow \frac{P_2}{P_1} = \left(\frac{X_2}{X_1} \right)^2$$

$$P_2 = 2.50 \text{ kW}$$

Exercise:**Problem:**

A photovoltaic array of (solar cells) is 10.0% efficient in gathering solar energy and converting it to electricity. If the average intensity of sunlight on one day is 70.00 W/m^2 , what area should your array have to gather energy at the rate of 100 W ? (b) What is the maximum cost of the array if it must pay for itself in two years of operation averaging 10.0 hours per day? Assume that it earns money at the rate of 9.00 cents per kilowatt-hour.

Exercise:**Problem:**

A microphone receiving a pure sound tone feeds an oscilloscope, producing a wave on its screen. If the sound intensity is originally $2.00 \times 10^{-5} \text{ W/m}^2$, but is turned up until the amplitude increases by 30.0% , what is the new intensity?

Solution:

$$I \propto X^2 \Rightarrow \frac{I_1}{I_2} = \left(\frac{X_1}{X_2} \right)^2 \Rightarrow$$

$$I_2 = 3.38 \times 10^{-5} \text{ W/m}^2$$

Exercise:

Problem:

A string with a mass of 0.30 kg has a length of 4.00 m. If the tension in the string is 50.00 N, and a sinusoidal wave with an amplitude of 2.00 cm is induced on the string, what must the frequency be for an average power of 100.00 W?

Exercise:**Problem:**

The power versus time for a point on a string ($\mu = 0.05 \text{ kg/m}$) in which a sinusoidal traveling wave is induced is shown in the preceding figure. The wave is modeled with the wave equation $y(x, t) = A \sin(20.93 \text{ m}^{-1}x - \omega t)$. What is the frequency and amplitude of the wave?

Solution:

$$f = 100.00 \text{ Hz}, A = 1.10 \text{ cm}$$

Exercise:**Problem:**

A string is under tension F_{T1} . Energy is transmitted by a wave on the string at rate P_1 by a wave of frequency f_1 . What is the ratio of the new energy transmission rate P_2 to P_1 if the tension is doubled?

Exercise:**Problem:**

A 250-Hz tuning fork is struck and the intensity at the source is I_1 at a distance of one meter from the source. (a) What is the intensity at a distance of 4.00 m from the source? (b) How far from the tuning fork is the intensity a tenth of the intensity at the source?

Solution:

$$\begin{aligned} \text{a. } I_2 &= 0.063 I_1; \text{ b. } I_1 4\pi r_1^2 = I_2 4\pi r_2^2 \\ r_2 &= 3.16 \text{ m} \end{aligned}$$

Exercise:**Problem:**

A sound speaker is rated at a voltage of $P = 120.00 \text{ V}$ and a current of $I = 10.00 \text{ A}$. Electrical power consumption is $P = IV$. To test the speaker, a signal of a sine wave is applied to the speaker. Assuming that the sound wave moves as a spherical wave and that all of the energy applied to the speaker is converted to sound energy, how far from the speaker is the intensity equal to 3.82 W/m^2 ?

Exercise:**Problem:**

The energy of a ripple on a pond is proportional to the amplitude squared. If the amplitude of the ripple is 0.1 cm at a distance from the source of 6.00 meters, what was the amplitude at a distance of 2.00 meters from the source?

Solution:

$$2\pi r_1 A_1^2 = 2\pi r_2 A_2^2, A_1 = \left(\frac{r_2}{r_1}\right)^{1/2} A_2 = 0.17 \text{ m}$$

Glossary

intensity (I)
power per unit area

Interference of Waves

By the end of this section, you will be able to:

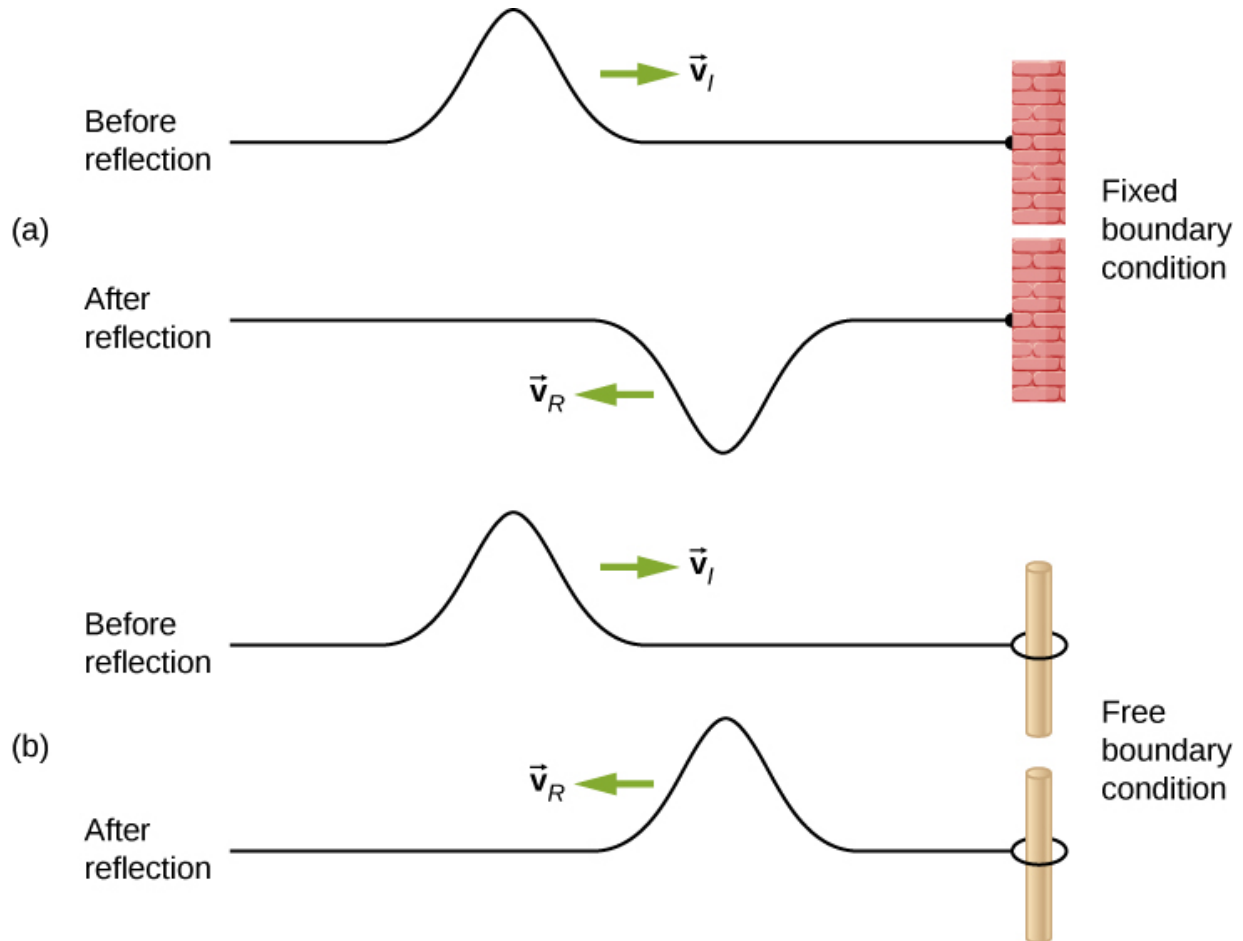
- Explain how mechanical waves are reflected and transmitted at the boundaries of a medium
- Define the terms interference and superposition
- Find the resultant wave of two identical sinusoidal waves that differ only by a phase shift

Up to now, we have been studying mechanical waves that propagate continuously through a medium, but we have not discussed what happens when waves encounter the boundary of the medium or what happens when a wave encounters another wave propagating through the same medium. Waves do interact with boundaries of the medium, and all or part of the wave can be reflected. For example, when you stand some distance from a rigid cliff face and yell, you can hear the sound waves reflect off the rigid surface as an echo. Waves can also interact with other waves propagating in the same medium. If you throw two rocks into a pond some distance from one another, the circular ripples that result from the two stones seem to pass through one another as they propagate out from where the stones entered the water. This phenomenon is known as interference. In this section, we examine what happens to waves encountering a boundary of a medium or another wave propagating in the same medium. We will see that their behavior is quite different from the behavior of particles and rigid bodies. Later, when we study modern physics, we will see that only at the scale of atoms do we see similarities in the properties of waves and particles.

Reflection and Transmission

When a wave propagates through a medium, it reflects when it encounters the boundary of the medium. The wave before hitting the boundary is known as the incident wave. The wave after encountering the boundary is known as the reflected wave. How the wave is reflected at the boundary of the medium depends on the boundary conditions; waves will react differently if the boundary of the medium is fixed in place or free to move ([\[link\]](#)). A **fixed boundary condition** exists when the medium at a boundary is fixed in place

so it cannot move. A **free boundary condition** exists when the medium at the boundary is free to move.



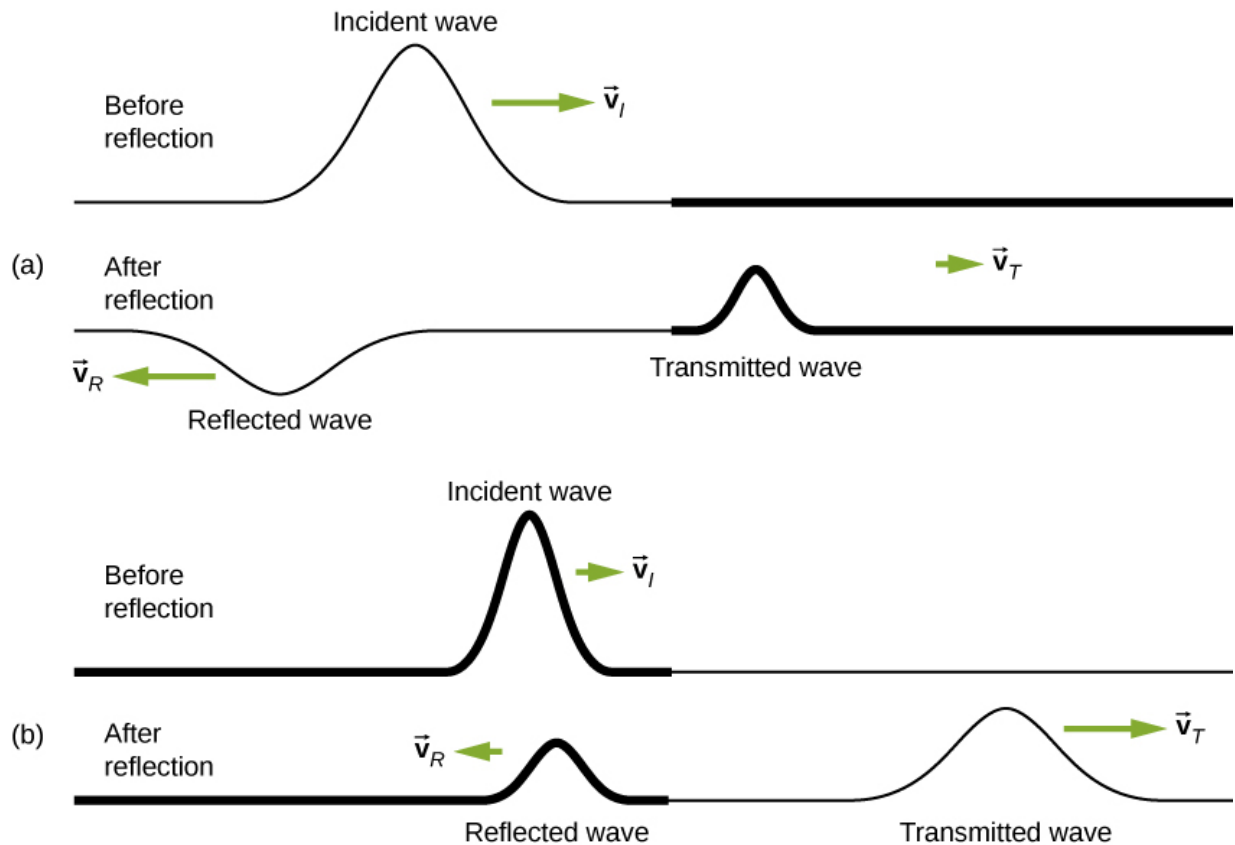
(a) One end of a string is fixed so that it cannot move. A wave propagating on the string, encountering this *fixed boundary condition*, is reflected 180° (π rad) out of phase with respect to the incident wave.

(b) One end of a string is tied to a solid ring of negligible mass on a frictionless lab pole, where the ring is free to move. A wave propagating on the string, encountering this *free boundary condition*, is reflected in phase 0° (0 rad) with respect to the wave.

Part (a) of the [\[link\]](#) shows a fixed boundary condition. Here, one end of the string is fixed to a wall so the end of the string is fixed in place and the medium (the string) at the boundary cannot move. When the wave is reflected, the amplitude of the reflected wave is exactly the same as the amplitude of the incident wave, but the reflected wave is reflected 180° (π rad) out of phase with respect to the incident wave. The phase change can be explained using Newton's third law: Recall that Newton's third law states that when object *A* exerts a force on object *B*, then object *B* exerts an equal and opposite force on object *A*. As the incident wave encounters the wall, the string exerts an upward force on the wall and the wall reacts by exerting an equal and opposite force on the string. The reflection at a fixed boundary is inverted. Note that the figure shows a crest of the incident wave reflected as a trough. If the incident wave were a trough, the reflected wave would be a crest.

Part (b) of the figure shows a free boundary condition. Here, one end of the string is tied to a solid ring of negligible mass on a frictionless pole, so the end of the string is free to move up and down. As the incident wave encounters the boundary of the medium, it is also reflected. In the case of a free boundary condition, the reflected wave is in phase with respect to the incident wave. In this case, the wave encounters the free boundary applying an upward force on the ring, accelerating the ring up. The ring travels up to the maximum height equal to the amplitude of the wave and then accelerates down towards the equilibrium position due to the tension in the string. The figure shows the crest of an incident wave being reflected in phase with respect to the incident wave as a crest. If the incident wave were a trough, the reflected wave would also be a trough. The amplitude of the reflected wave would be equal to the amplitude of the incident wave.

In some situations, the boundary of the medium is neither fixed nor free. Consider [\[link\]](#)(a), where a low-linear mass density string is attached to a string of a higher linear mass density. In this case, the reflected wave is out of phase with respect to the incident wave. There is also a transmitted wave that is in phase with respect to the incident wave. Both the transmitted and the reflected waves have amplitudes less than the amplitude of the incident wave. If the tension is the same in both strings, the wave speed is higher in the string with the lower linear mass density.



Waves traveling along two types of strings: a thick string with a high linear density and a thin string with a low linear density. Both strings are under the same tension, so a wave moves faster on the low-density string than on the high-density string. (a) A wave moving from a low-density to a high-density medium results in a reflected wave that is 180° (π rad) out of phase with respect to the incident pulse (or wave) and a transmitted wave that is in phase with the incident wave. (b) When a wave moves from a high-density medium to a low-density medium, both the reflected and transmitted wave are in phase with respect to the incident wave.

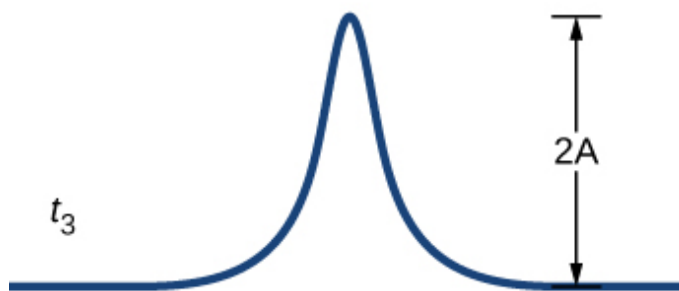
Part (b) of the figure shows a high-linear mass density string is attached to a string of a lower linear density. In this case, the reflected wave is in phase with respect to the incident wave. There is also a transmitted wave that is in phase with respect to the incident wave. Both the incident and the reflected waves have amplitudes less than the amplitude of the incident wave. Here

you may notice that if the tension is the same in both strings, the wave speed is higher in the string with the lower linear mass density.

Superposition and Interference

Most waves do not look very simple. Complex waves are more interesting, even beautiful, but they look formidable. Most interesting mechanical waves consist of a combination of two or more traveling waves propagating in the same medium. The principle of superposition can be used to analyze the combination of waves.

Consider two simple pulses of the same amplitude moving toward one another in the same medium, as shown in [\[link\]](#). Eventually, the waves overlap, producing a wave that has twice the amplitude, and then continue on unaffected by the encounter. The pulses are said to interfere, and this phenomenon is known as **interference**.



Two pulses moving toward one another experience interference. The term interference refers to what happens when two waves overlap.

To analyze the interference of two or more waves, we use the principle of superposition. For mechanical waves, the principle of **superposition** states that if two or more traveling waves combine at the same point, the resulting position of the mass element of the medium, at that point, is the algebraic sum of the position due to the individual waves. This property is exhibited by many waves observed, such as waves on a string, sound waves, and surface water waves. Electromagnetic waves also obey the superposition principle, but the electric and magnetic fields of the combined wave are added instead of the displacement of the medium. Waves that obey the superposition principle are linear waves; waves that do not obey the superposition principle are said to be nonlinear waves. In this chapter, we deal with linear waves, in particular, sinusoidal waves.

The superposition principle can be understood by considering the linear wave equation. In [Mathematics of a Wave](#), we defined a linear wave as a wave whose mathematical representation obeys the linear wave equation. For a transverse wave on a string with an elastic restoring force, the linear wave equation is

Equation:

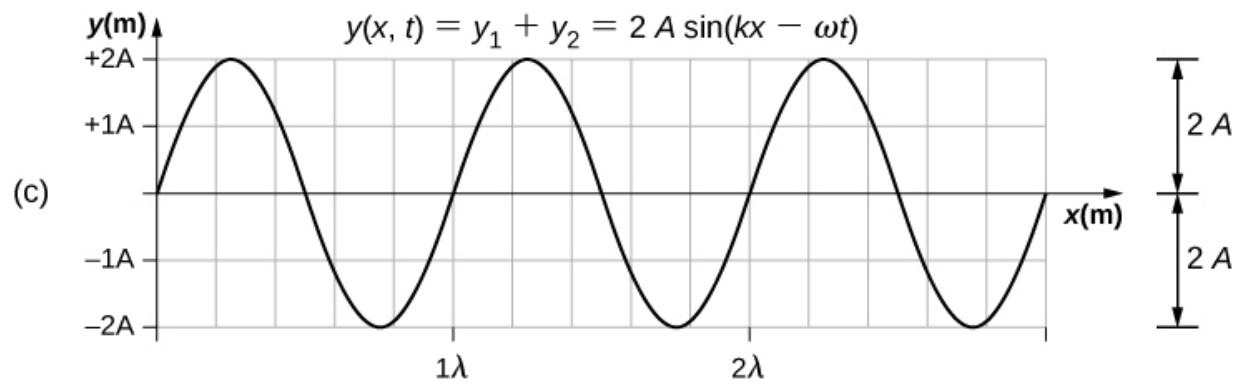
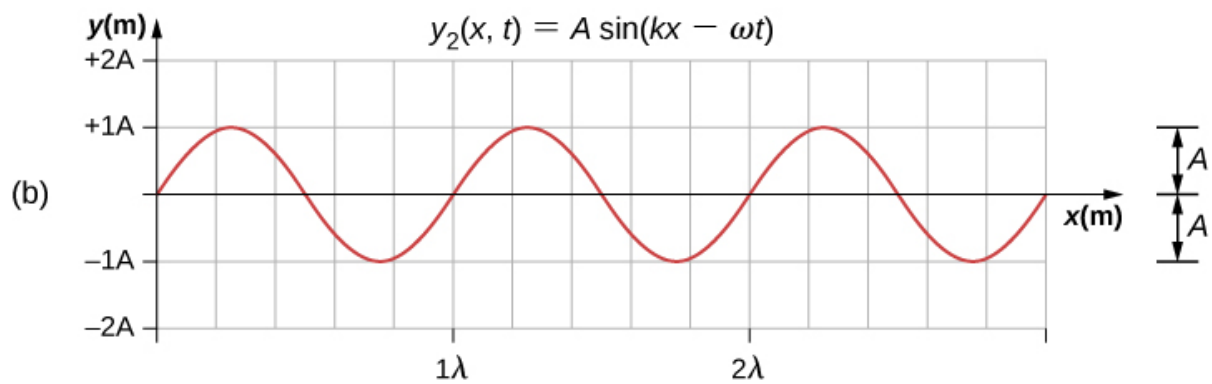
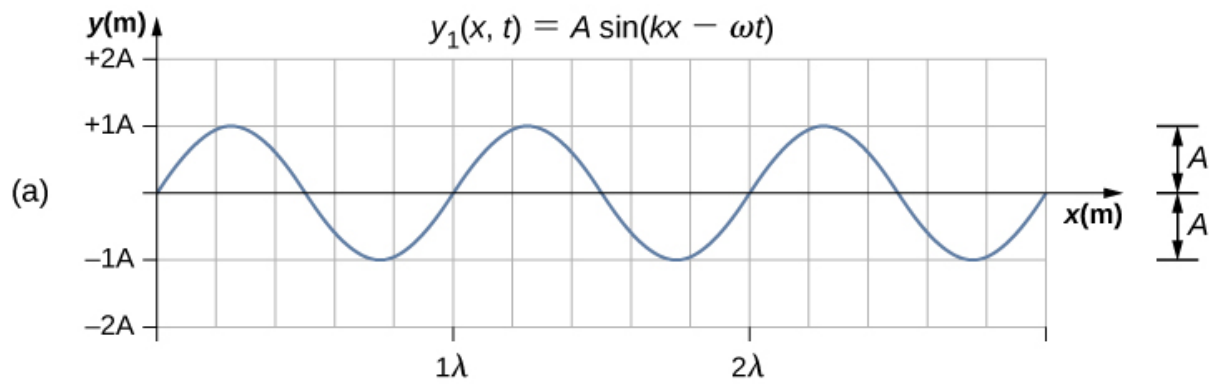
$$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y(x, t)}{\partial t^2}.$$

Any wave function $y(x, t) = y(x \mp vt)$, where the argument of the function is linear ($x \mp vt$) is a solution to the linear wave equation and is a linear wave function. If wave functions $y_1(x, t)$ and $y_2(x, t)$ are solutions to the linear wave equation, the sum of the two functions $y_1(x, t) + y_2(x, t)$ is also a solution to the linear wave equation. Mechanical waves that obey superposition are normally restricted to waves with amplitudes that are small with respect to their wavelengths. If the amplitude is too large, the medium is distorted past the region where the restoring force of the medium is linear.

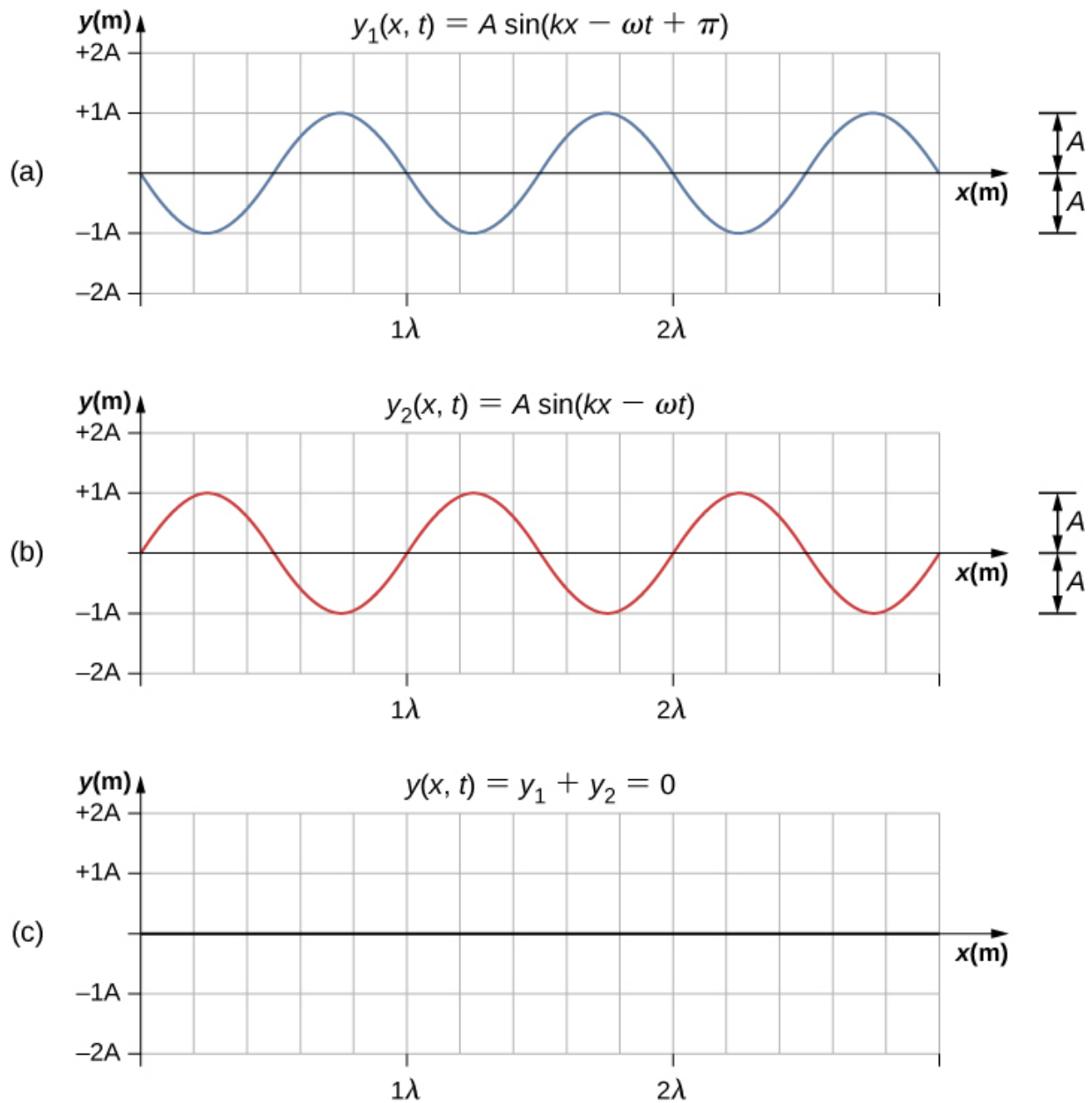
Waves can interfere constructively or destructively. [\[link\]](#) shows two identical sinusoidal waves that arrive at the same point exactly in phase. [\[link\]](#)(a) and (b) show the two individual waves, [\[link\]](#)(c) shows the resultant wave that results from the algebraic sum of the two linear waves. The crests

of the two waves are precisely aligned, as are the troughs. This superposition produces **constructive interference**. Because the disturbances add, constructive interference produces a wave that has twice the amplitude of the individual waves, but has the same wavelength.

[\[link\]](#) shows two identical waves that arrive exactly 180° out of phase, producing **destructive interference**. [\[link\]](#)(a) and (b) show the individual waves, and [\[link\]](#)(c) shows the superposition of the two waves. Because the troughs of one wave add the crest of the other wave, the resulting amplitude is zero for destructive interference—the waves completely cancel.

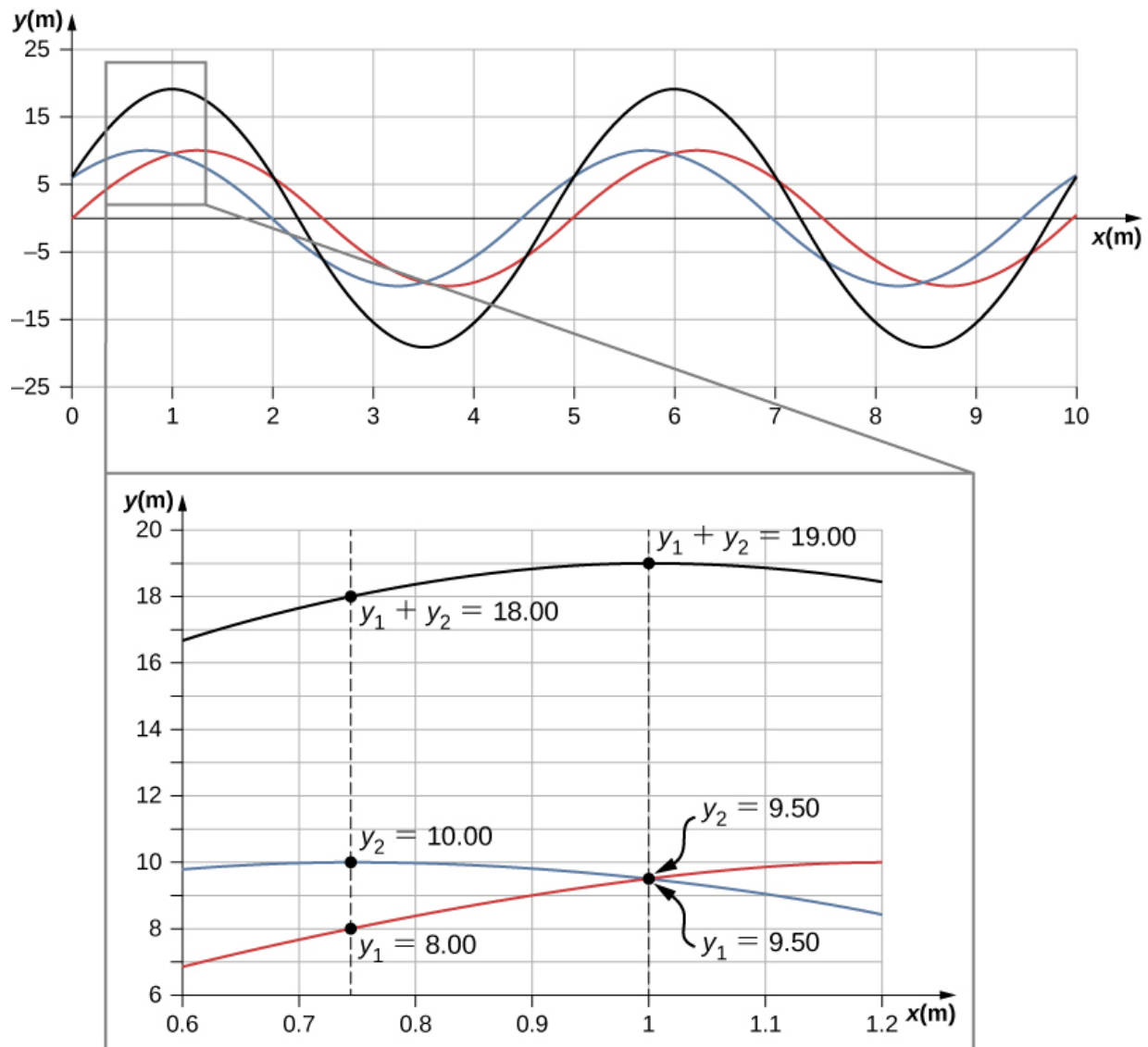


Constructive interference of two identical waves produces a wave with twice the amplitude, but the same wavelength.



Destructive interference of two identical waves, one with a phase shift of 180° (π rad), produces zero amplitude, or complete cancellation.

When linear waves interfere, the resultant wave is just the algebraic sum of the individual waves as stated in the principle of superposition. [\[link\]](#) shows two waves (red and blue) and the resultant wave (black). The resultant wave is the algebraic sum of the two individual waves.

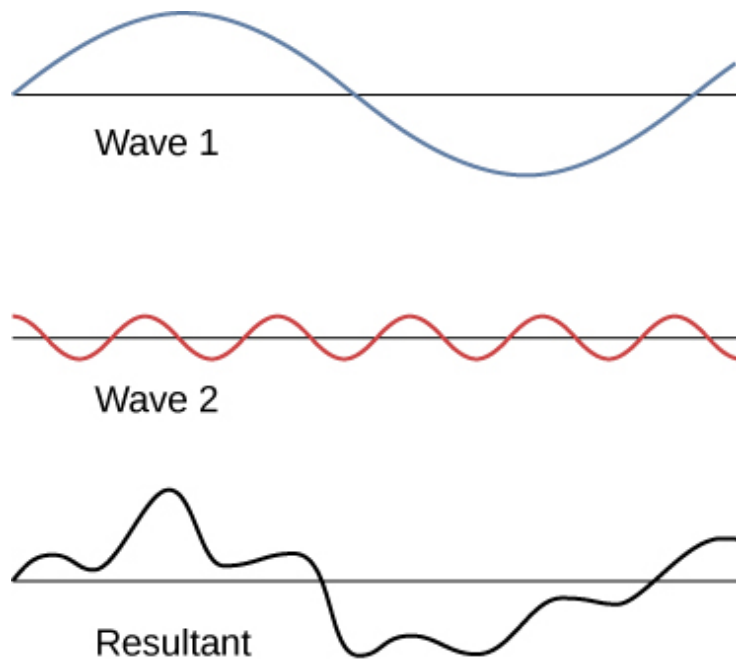


When two linear waves in the same medium interfere, the height of resulting wave is the sum of the heights of the individual waves, taken point by point. This plot shows two waves (red and blue) added together, along with the resulting wave (black). These graphs represent the height of the wave at each point. The waves may be any linear wave, including ripples on a pond, disturbances on a string, sound, or electromagnetic waves.

The superposition of most waves produces a combination of constructive and destructive interference, and can vary from place to place and time to time.

Sound from a stereo, for example, can be loud in one spot and quiet in another. Varying loudness means the sound waves add partially constructively and partially destructively at different locations. A stereo has at least two speakers creating sound waves, and waves can reflect from walls. All these waves interfere, and the resulting wave is the superposition of the waves.

We have shown several examples of the superposition of waves that are similar. [\[link\]](#) illustrates an example of the superposition of two dissimilar waves. Here again, the disturbances add, producing a resultant wave.



Superposition of nonidentical waves
exhibits both constructive and
destructive interference.

At times, when two or more mechanical waves interfere, the pattern produced by the resulting wave can be rich in complexity, some without any readily discernable patterns. For example, plotting the sound wave of your favorite music can look quite complex and is the superposition of the

individual sound waves from many instruments; it is the complexity that makes the music interesting and worth listening to. At other times, waves can interfere and produce interesting phenomena, which are complex in their appearance and yet beautiful in simplicity of the physical principle of superposition, which formed the resulting wave. One example is the phenomenon known as standing waves, produced by two identical waves moving in different directions. We will look more closely at this phenomenon in the next section.

Note:

Try this [simulation](#) to make waves with a dripping faucet, audio speaker, or laser! Add a second source or a pair of slits to create an interference pattern. You can observe one source or two sources. Using two sources, you can observe the interference patterns that result from varying the frequencies and the amplitudes of the sources.

Superposition of Sinusoidal Waves that Differ by a Phase Shift

Many examples in physics consist of two sinusoidal waves that are identical in amplitude, wave number, and angular frequency, but differ by a phase shift:

Equation:

$$\begin{aligned}y_1(x, t) &= A \sin(kx - \omega t + \phi), \\y_2(x, t) &= A \sin(kx - \omega t).\end{aligned}$$

When these two waves exist in the same medium, the resultant wave resulting from the superposition of the two individual waves is the sum of the two individual waves:

Equation:

$$y_R(x, t) = y_1(x, t) + y_2(x, t) = A \sin(kx - \omega t + \phi) + A \sin(kx - \omega t).$$

The resultant wave can be better understood by using the trigonometric identity:

Equation:

$$\sin u + \sin v = 2 \sin \left(\frac{u + v}{2} \right) \cos \left(\frac{u - v}{2} \right),$$

where $u = kx - \omega t + \phi$ and $v = kx - \omega t$. The resulting wave becomes

Equation:

$$\begin{aligned} y_R(x, t) &= y_1(x, t) + y_2(x, t) = A \sin(kx - \omega t + \phi) + A \sin(kx - \omega t) \\ &= 2A \sin \left(\frac{(kx - \omega t + \phi) + (kx - \omega t)}{2} \right) \cos \left(\frac{(kx - \omega t + \phi) - (kx - \omega t)}{2} \right) \\ &= 2A \sin \left(kx - \omega t + \frac{\phi}{2} \right) \cos \left(\frac{\phi}{2} \right). \end{aligned}$$

This equation is usually written as

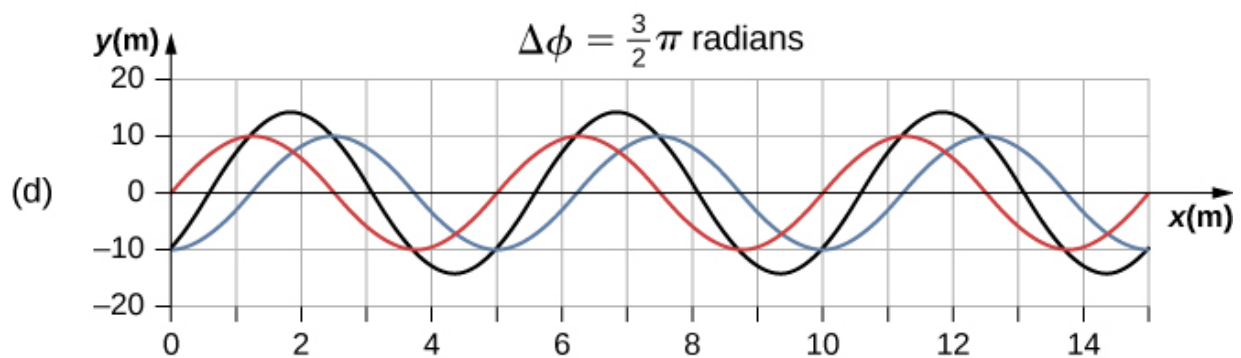
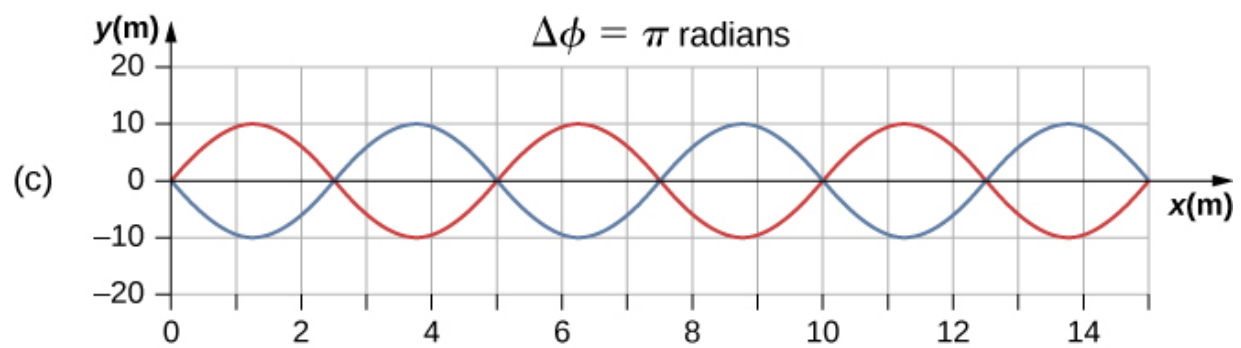
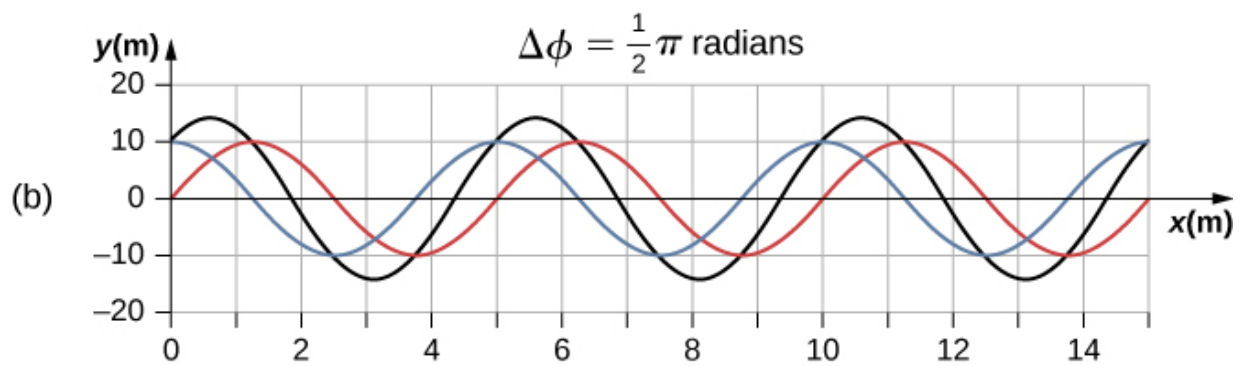
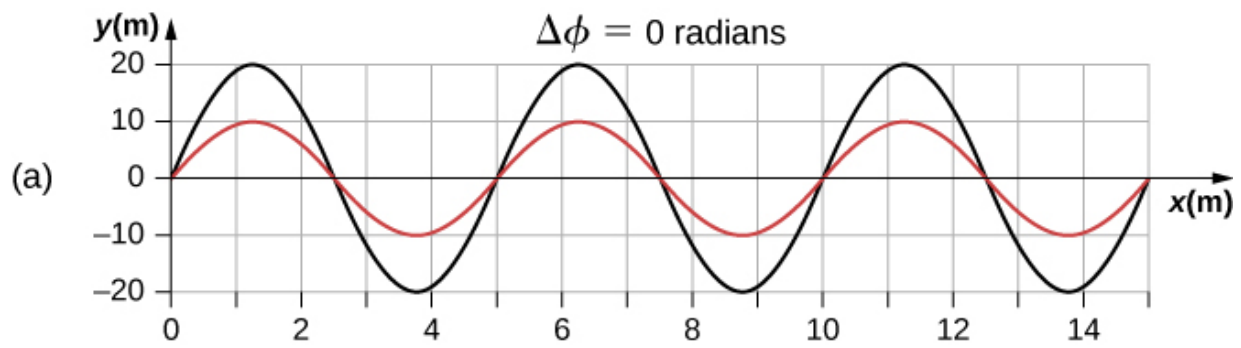
Note:

Equation:

$$y_R(x, t) = \left[2A \cos \left(\frac{\phi}{2} \right) \right] \sin \left(kx - \omega t + \frac{\phi}{2} \right).$$

The resultant wave has the same wave number and angular frequency, an amplitude of $A_R = \left[2A \cos \left(\frac{\phi}{2} \right) \right]$, and a phase shift equal to half the original phase shift. Examples of waves that differ only in a phase shift are shown in [\[link\]](#). The red and blue waves each have the same amplitude, wave number, and angular frequency, and differ only in a phase shift. They therefore have the same period, wavelength, and frequency. The green wave is the result of the superposition of the two waves. When the two waves have

a phase difference of zero, the waves are in phase, and the resultant wave has the same wave number and angular frequency, and an amplitude equal to twice the individual amplitudes (part (a)). This is constructive interference. If the phase difference is 180° , the waves interfere in destructive interference (part (c)). The resultant wave has an amplitude of zero. Any other phase difference results in a wave with the same wave number and angular frequency as the two incident waves but with a phase shift of $\phi/2$ and an amplitude equal to $2A \cos(\phi/2)$. Examples are shown in parts (b) and (d).



Superposition of two waves with identical amplitudes, wavelengths, and frequency, but that differ in a phase shift. The red wave is defined by

the wave function $y_1(x, t) = A \sin(kx - \omega t)$ and the blue wave is defined by the wave function $y_2(x, t) = A \sin(kx - \omega t + \phi)$. The black line shows the result of adding the two waves. The phase difference between the two waves are (a) 0.00 rad, (b) $\pi/2$ rad, (c) π rad, and (d) $3\pi/2$ rad.

Summary

- Superposition is the combination of two waves at the same location.
- Constructive interference occurs from the superposition of two identical waves that are in phase.
- Destructive interference occurs from the superposition of two identical waves that are 180° (π radians) out of phase.
- The wave that results from the superposition of two sine waves that differ only by a phase shift is a wave with an amplitude that depends on the value of the phase difference.

Conceptual Questions

Exercise:

Problem:

An incident sinusoidal wave is sent along a string that is fixed to the wall with a wave speed of v . The wave reflects off the end of the string. Describe the reflected wave.

Exercise:

Problem:

A string of a length of 2.00 m with a linear mass density of $\mu = 0.006 \text{ kg/m}$ is attached to the end of a 2.00-m-long string with a linear mass density of $\mu = 0.012 \text{ kg/m}$. The free end of the higher-density string is fixed to the wall, and a student holds the free end of the low-density string, keeping the tension constant in both strings. The student sends a pulse down the string. Describe what happens at the interface between the two strings.

Solution:

At the interface, the incident pulse produces a reflected pulse and a transmitted pulse. The reflected pulse would be out of phase with respect to the incident pulse, and would move at the same propagation speed as the incident pulse, but would move in the opposite direction. The transmitted pulse would travel in the same direction as the incident pulse, but at half the speed. The transmitted pulse would be in phase with the incident pulse. Both the reflected pulse and the transmitted pulse would have amplitudes less than the amplitude of the incident pulse.

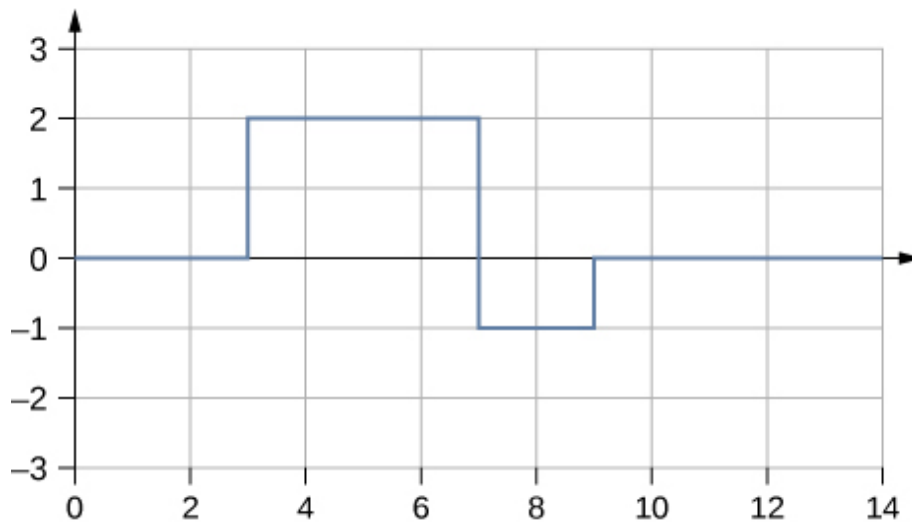
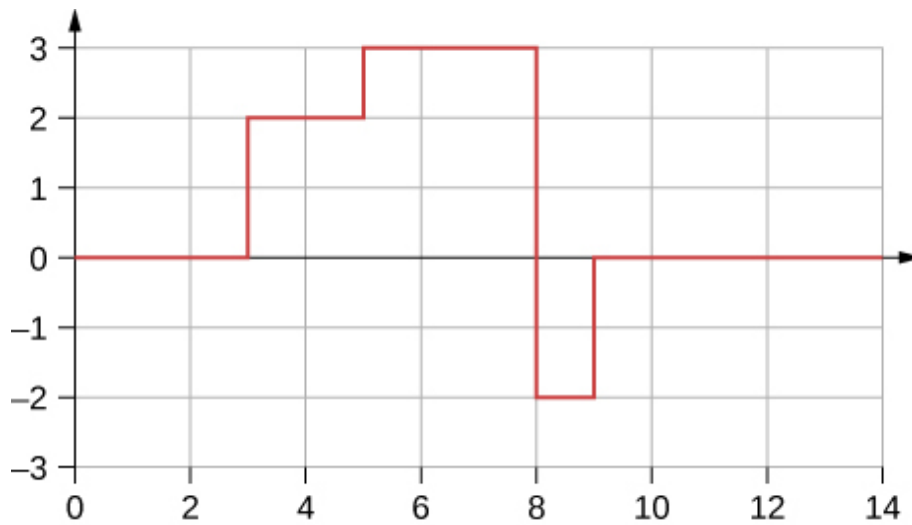
Exercise:**Problem:**

A long, tight spring is held by two students, one student holding each end. Each student gives the end a flip sending one wavelength of a sinusoidal wave down the spring in opposite directions. When the waves meet in the middle, what does the wave look like?

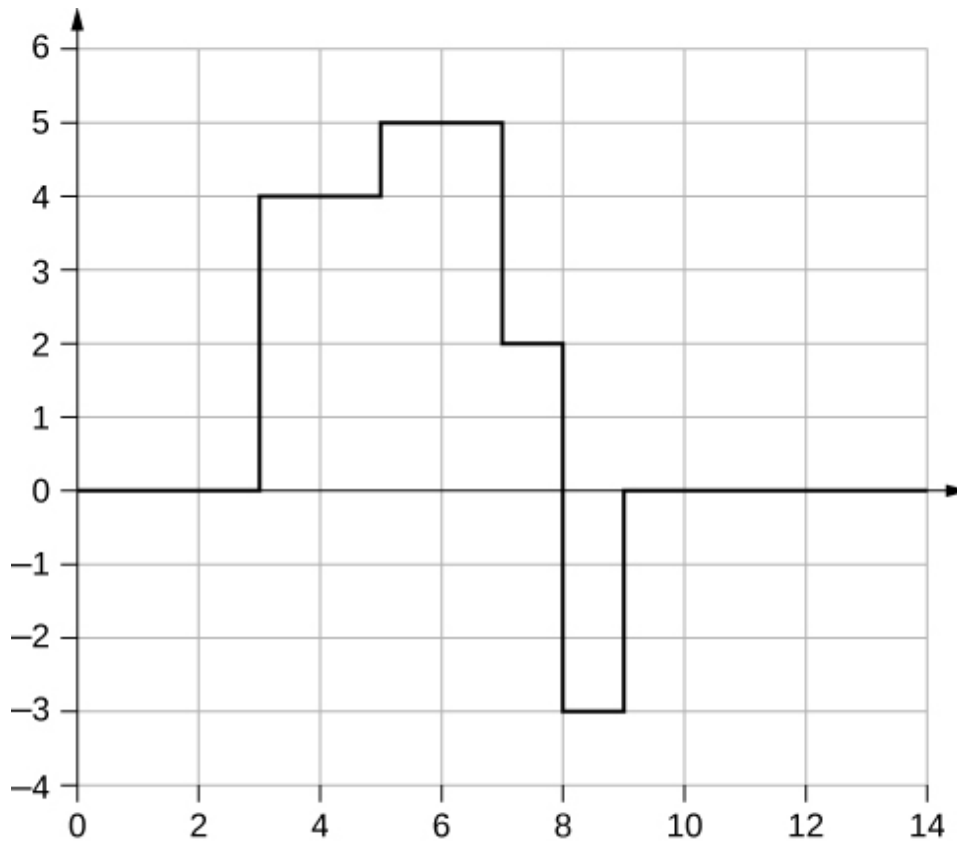
Exercise:

Problem:

Many of the topics discussed in this chapter are useful beyond the topics of mechanical waves. It is hard to conceive of a mechanical wave with sharp corners, but you could encounter such a wave form in your digital electronics class, as shown below. This could be a signal from a device known as an analog to digital converter, in which a continuous voltage signal is converted into a discrete signal or a digital recording of sound. What is the result of the superposition of the two signals?



Solution:



Exercise:

Problem:

A string of a constant linear mass density is held taut by two students, each holding one end. The tension in the string is constant. The students each send waves down the string by wiggling the string. (a) Is it possible for the waves to have different wave speeds? (b) Is it possible for the waves to have different frequencies? (c) Is it possible for the waves to have different wavelengths?

Problems

Exercise:

Problem:

Consider two sinusoidal waves traveling along a string, modeled as $y_1(x, t) = 0.3 \text{ m} \sin(4 \text{ m}^{-1}x + 3 \text{ s}^{-1}t)$ and $y_2(x, t) = 0.6 \text{ m} \sin(8 \text{ m}^{-1}x - 6 \text{ s}^{-1}t)$. What is the height of the resultant wave formed by the interference of the two waves at the position $x = 0.5 \text{ m}$ at time $t = 0.2 \text{ s}$?

Exercise:**Problem:**

Consider two sinusoidal sine waves traveling along a string, modeled as $y_1(x, t) = 0.3 \text{ m} \sin(4 \text{ m}^{-1}x + 3 \text{ s}^{-1}t + \frac{\pi}{3})$ and $y_2(x, t) = 0.6 \text{ m} \sin(8 \text{ m}^{-1}x - 6 \text{ s}^{-1}t)$. What is the height of the resultant wave formed by the interference of the two waves at the position $x = 1.0 \text{ m}$ at time $t = 3.0 \text{ s}$?

Solution:

$$y(x, t) = 0.63 \text{ m}$$

Exercise:**Problem:**

Consider two sinusoidal sine waves traveling along a string, modeled as $y_1(x, t) = 0.3 \text{ m} \sin(4 \text{ m}^{-1}x - 3 \text{ s}^{-1}t)$ and $y_2(x, t) = 0.3 \text{ m} \sin(4 \text{ m}^{-1}x + 3 \text{ s}^{-1}t)$. What is the wave function of the resulting wave? [Hint: Use the trig identity $\sin(u \pm v) = \sin u \cos v \pm \cos u \sin v$]

Exercise:

Problem:

Two sinusoidal waves are moving through a medium in the same direction, both having amplitudes of 3.00 cm, a wavelength of 5.20 m, and a period of 6.52 s, but one has a phase shift of an angle ϕ . What is the phase shift if the resultant wave has an amplitude of 5.00 cm? [Hint: Use the trig identity $\sin u + \sin v = 2 \sin \left(\frac{u+v}{2} \right) \cos \left(\frac{u-v}{2} \right)$]

Solution:

$$A_R = 2A \cos \left(\frac{\phi}{2} \right), \phi = 1.17 \text{ rad}$$

Exercise:**Problem:**

Two sinusoidal waves are moving through a medium in the positive x -direction, both having amplitudes of 6.00 cm, a wavelength of 4.3 m, and a period of 6.00 s, but one has a phase shift of an angle $\phi = 0.50$ rad. What is the height of the resultant wave at a time $t = 3.15$ s and a position $x = 0.45$ m?

Exercise:**Problem:**

Two sinusoidal waves are moving through a medium in the positive x -direction, both having amplitudes of 7.00 cm, a wave number of $k = 3.00 \text{ m}^{-1}$, an angular frequency of $\omega = 2.50 \text{ s}^{-1}$, and a period of 6.00 s, but one has a phase shift of an angle $\phi = \frac{\pi}{12}$ rad. What is the height of the resultant wave at a time $t = 2.00$ s and a position $x = 0.53$ m?

Solution:

$$y_R = 1.90 \text{ cm}$$

Exercise:

Problem:

Consider two waves $y_1(x, t)$ and $y_2(x, t)$ that are identical except for a phase shift propagating in the same medium. (a) What is the phase shift, in radians, if the amplitude of the resulting wave is 1.75 times the amplitude of the individual waves? (b) What is the phase shift in degrees? (c) What is the phase shift as a percentage of the individual wavelength?

Exercise:**Problem:**

Two sinusoidal waves, which are identical except for a phase shift, travel along in the same direction. The wave equation of the resultant wave is $y_R(x, t) = 0.70 \text{ m} \sin(3.00 \text{ m}^{-1}x - 6.28 \text{ s}^{-1}t + \pi/16 \text{ rad})$. What are the angular frequency, wave number, amplitude, and phase shift of the individual waves?

Solution:

$$\begin{aligned}\omega &= 6.28 \text{ s}^{-1}, k = 3.00 \text{ m}^{-1}, \phi = \frac{\pi}{8} \text{ rad}, \\ A_R &= 2A \cos\left(\frac{\phi}{2}\right), A = 0.37 \text{ m}\end{aligned}$$

Exercise:**Problem:**

Two sinusoidal waves, which are identical except for a phase shift, travel along in the same direction. The wave equation of the resultant wave is $y_R(x, t) = 0.35 \text{ cm} \sin(6.28 \text{ m}^{-1}x - 1.57 \text{ s}^{-1}t + \frac{\pi}{4})$. What are the period, wavelength, amplitude, and phase shift of the individual waves?

Exercise:

Problem:

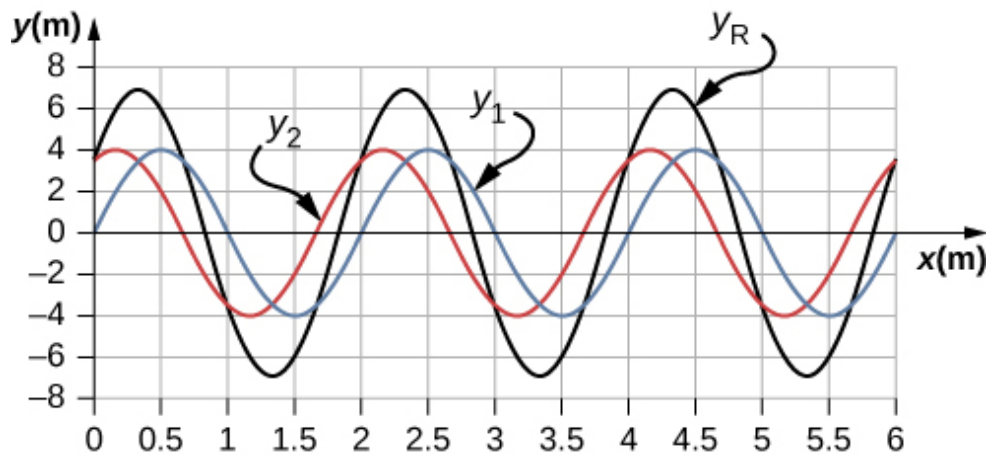
Consider two wave functions,

$$y_1(x, t) = 4.00 \text{ m} \sin(\pi \text{ m}^{-1}x - \pi \text{ s}^{-1}t) \text{ and}$$

$y_2(x, t) = 4.00 \text{ m} \sin(\pi \text{ m}^{-1}x - \pi \text{ s}^{-1}t + \frac{\pi}{3})$. (a) Using a spreadsheet, plot the two wave functions and the wave that results from the superposition of the two wave functions as a function of position ($0.00 \leq x \leq 6.00 \text{ m}$) for the time $t = 0.00 \text{ s}$. (b) What are the wavelength and amplitude of the two original waves? (c) What are the wavelength and amplitude of the resulting wave?

Solution:

a.



;

b. $\lambda = 2.0 \text{ m}$, $A = 4 \text{ m}$; c. $\lambda_R = 2.0 \text{ m}$, $A_R = 6.93 \text{ m}$

Exercise:

Problem:

Consider two wave functions,

$$y_1(x, t) = 2.00 \text{ m} \sin\left(\frac{\pi}{2} \text{ m}^{-1}x - \frac{\pi}{3} \text{ s}^{-1}t\right) \text{ and}$$

$$y_2(x, t) = 2.00 \text{ m} \sin\left(\frac{\pi}{2} \text{ m}^{-1}x - \frac{\pi}{3} \text{ s}^{-1}t + \frac{\pi}{6}\right). \text{ (a) Verify that}$$

$$y_R = 2A \cos\left(\frac{\phi}{2}\right) \sin\left(kx - \omega t + \frac{\phi}{2}\right) \text{ is the solution for the wave that}$$

results from a superposition of the two waves. Make a column for x , y_1 ,

$$y_2, y_1 + y_2, \text{ and } y_R = 2A \cos\left(\frac{\phi}{2}\right) \sin\left(kx - \omega t + \frac{\phi}{2}\right). \text{ Plot four}$$

waves as a function of position where the range of x is from 0 to 12 m.

Exercise:**Problem:**

Consider two wave functions that differ only by a phase shift,

$$y_1(x, t) = A \cos(kx - \omega t) \text{ and } y_2(x, t) = A \cos(kx - \omega t + \phi). \text{ Use}$$

$$\text{the trigonometric identities } \cos u + \cos v = 2 \cos\left(\frac{u+v}{2}\right) \cos\left(\frac{u-v}{2}\right)$$

and $\cos(-\theta) = \cos(\theta)$ to find a wave equation for the wave resulting

from the superposition of the two waves. Does the resulting wave function come as a surprise to you?

Solution:

$$y_R(x, t) = 2A \cos\left(\frac{\phi}{2}\right) \cos\left(kx - \omega t + \frac{\phi}{2}\right); \text{ The result is not}$$

surprising because $\cos(\theta) = \sin\left(\theta + \frac{\pi}{2}\right)$.

Glossary

constructive interference

when two waves arrive at the same point exactly in phase; that is, the crests of the two waves are precisely aligned, as are the troughs

destructive interference

when two identical waves arrive at the same point exactly out of phase; that is, precisely aligned crest to trough

fixed boundary condition

when the medium at a boundary is fixed in place so it cannot move

free boundary condition

exists when the medium at the boundary is free to move

interference

overlap of two or more waves at the same point and time

superposition

phenomenon that occurs when two or more waves arrive at the same point

Standing Waves and Resonance

By the end of this section, you will be able to:

- Describe standing waves and explain how they are produced
- Describe the modes of a standing wave on a string
- Provide examples of standing waves beyond the waves on a string

Throughout this chapter, we have been studying traveling waves, or waves that transport energy from one place to another. Under certain conditions, waves can bounce back and forth through a particular region, effectively becoming stationary. These are called **standing waves**.

Another related effect is known as resonance. In [Oscillations](#), we defined resonance as a phenomenon in which a small-amplitude driving force could produce large-amplitude motion. Think of a child on a swing, which can be modeled as a physical pendulum. Relatively small-amplitude pushes by a parent can produce large-amplitude swings. Sometimes this resonance is good—for example, when producing music with a stringed instrument. At other times, the effects can be devastating, such as the collapse of a building during an earthquake. In the case of standing waves, the relatively large amplitude standing waves are produced by the superposition of smaller amplitude component waves.

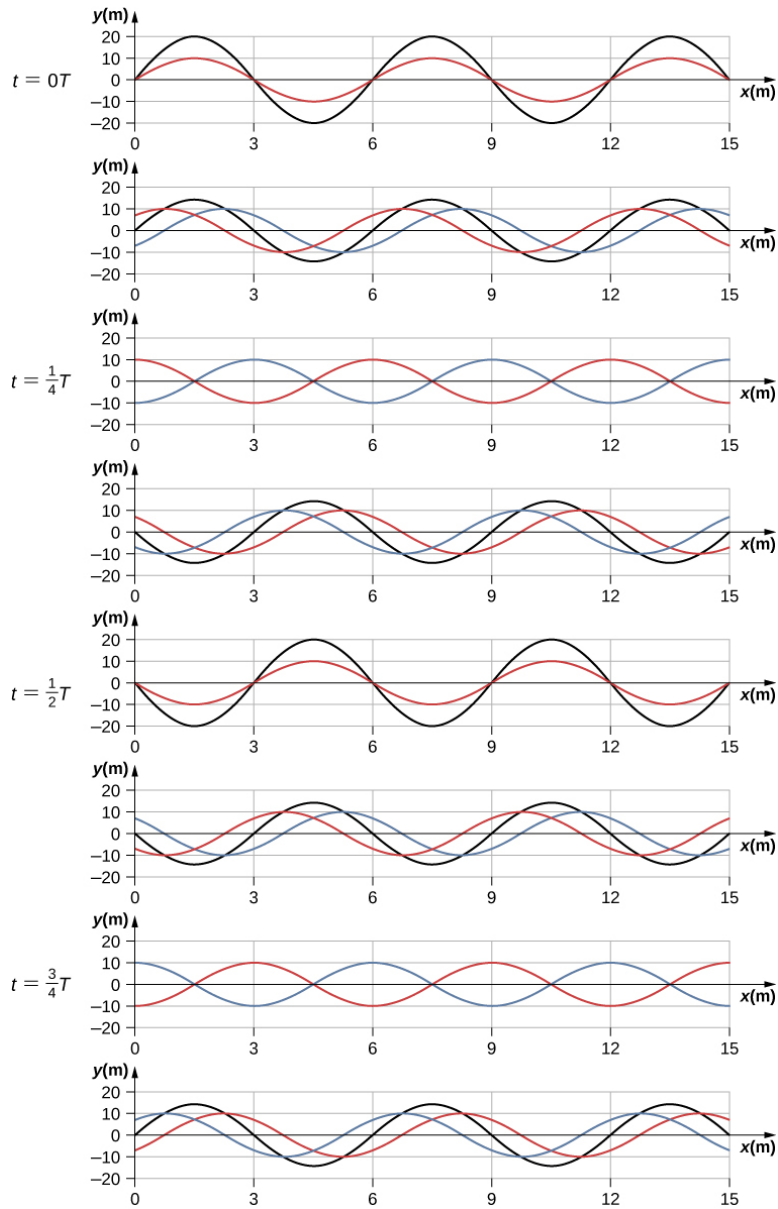
Standing Waves

Sometimes waves do not seem to move; rather, they just vibrate in place. You can see unmoving waves on the surface of a glass of milk in a refrigerator, for example. Vibrations from the refrigerator motor create waves on the milk that oscillate up and down but do not seem to move across the surface. [\[link\]](#) shows an experiment you can try at home. Take a bowl of milk and place it on a common box fan. Vibrations from the fan will produce circular standing waves in the milk. The waves are visible in the photo due to the reflection from a lamp. These waves are formed by the superposition of two or more traveling waves, such as illustrated in [\[link\]](#) for two identical waves moving in opposite directions. The waves move through each other with their disturbances adding as they go by. If the two waves have the same amplitude and wavelength, then they alternate between constructive and destructive interference. The resultant looks like a wave standing in place and, thus, is called a standing wave.



Standing waves are formed on the surface of a bowl of milk sitting on a box fan. The vibrations from the fan causes the surface of the milk to oscillate. The waves

are visible due to the reflection of light from a lamp.
(credit: David Chelton)



Time snapshots of two sine waves. The red wave is moving in the $-x$ -direction and the blue wave is moving in the $+x$ -direction. The resulting wave is shown in black. Consider the resultant wave at the points $x = 0 \text{ m}, 3 \text{ m}, 6 \text{ m}, 9 \text{ m}, 12 \text{ m}, 15 \text{ m}$ and notice that the resultant wave always equals zero at these points, no matter what the time is. These points are known as fixed points (nodes). In between each two nodes is an antinode, a place

where the medium oscillates with an amplitude equal to the sum of the amplitudes of the individual waves.

Consider two identical waves that move in opposite directions. The first wave has a wave function of $y_1(x, t) = A \sin(kx - \omega t)$ and the second wave has a wave function $y_2(x, t) = A \sin(kx + \omega t)$. The waves interfere and form a resultant wave

Equation:

$$\begin{aligned}y(x, t) &= y_1(x, t) + y_2(x, t), \\y(x, t) &= A \sin(kx - \omega t) + A \sin(kx + \omega t).\end{aligned}$$

This can be simplified using the trigonometric identity

Equation:

$$\sin(\alpha \pm \beta) = \sin \alpha \cos \beta \pm \cos \alpha \sin \beta,$$

where $\alpha = kx$ and $\beta = \omega t$, giving us

Equation:

$$y(x, t) = A[\sin(kx)\cos(\omega t) - \cos(kx)\sin(\omega t) + \sin(kx)\cos(\omega t) + \cos(kx)\sin(\omega t)],$$

which simplifies to

Note:

Equation:

$$y(x, t) = [2A \sin(kx)]\cos(\omega t).$$

Notice that the resultant wave is a sine wave that is a function only of position, multiplied by a cosine function that is a function only of time. Graphs of $y(x, t)$ as a function of x for various times are shown in [\[link\]](#). The red wave moves in the negative x -direction, the blue wave moves in the positive x -direction, and the black wave is the sum of the two waves. As the red and blue waves move through each other, they move in and out of constructive interference and destructive interference.

Initially, at time $t = 0$, the two waves are in phase, and the result is a wave that is twice the amplitude of the individual waves. The waves are also in phase at the time $t = \frac{T}{2}$. In fact, the waves are in phase at any integer multiple of half of a period:

Equation:

$$t = n \frac{T}{2} \text{ where } n = 0, 1, 2, 3, \dots \text{ (in phase).}$$

At other times, the two waves are 180° (π radians) out of phase, and the resulting wave is equal to zero. This happens at

Equation:

$$t = \frac{1}{4}T, \frac{3}{4}T, \frac{5}{4}T, \dots, \frac{n}{4}T \text{ where } n = 1, 3, 5, \dots \text{ (out of phase).}$$

Notice that some x -positions of the resultant wave are always zero no matter what the phase relationship is. These positions are called **nodes**. Where do the nodes occur? Consider the solution to the sum of the two waves

Equation:

$$y(x, t) = [2A \sin(kx)] \cos(\omega t).$$

Finding the positions where the sine function equals zero provides the positions of the nodes.

Equation:

$$\begin{aligned} \sin(kx) &= 0 \\ kx &= 0, \pi, 2\pi, 3\pi, \dots \\ \frac{2\pi}{\lambda}x &= 0, \pi, 2\pi, 3\pi, \dots \\ x &= 0, \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}, \dots = n\frac{\lambda}{2} \quad n = 0, 1, 2, 3, \dots \end{aligned}$$

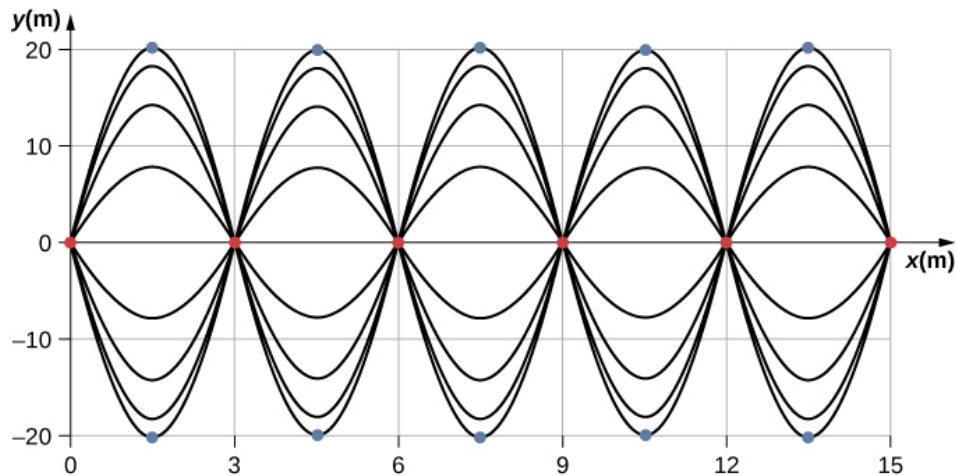
There are also positions where y oscillates between $y = \pm A$. These are the **antinodes**. We can find them by considering which values of x result in $\sin(kx) = \pm 1$.

Equation:

$$\begin{aligned} \sin(kx) &= \pm 1 \\ kx &= \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \dots \\ \frac{2\pi}{\lambda}x &= \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \dots \\ x &= \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \dots = n\frac{\lambda}{4} \quad n = 1, 3, 5, \dots \end{aligned}$$

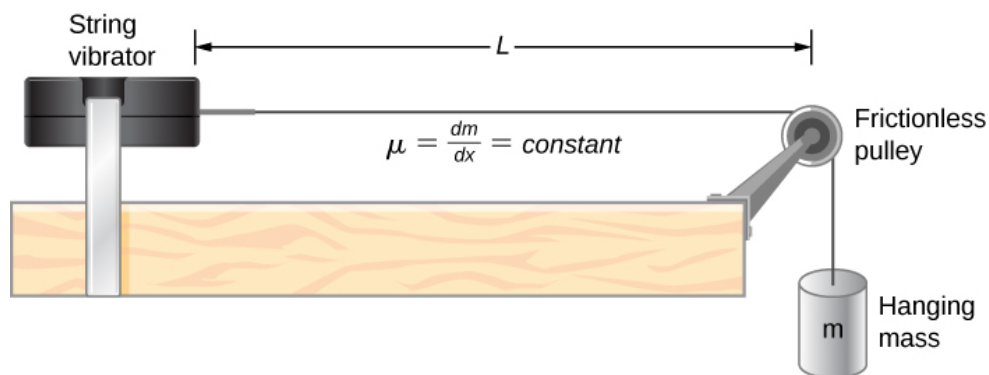
What results is a standing wave as shown in [\[link\]](#), which shows snapshots of the resulting wave of two identical waves moving in opposite directions. The resulting wave appears to be a sine wave with nodes at integer multiples of half wavelengths. The antinodes oscillate between $y = \pm 2A$ due to the cosine term, $\cos(\omega t)$, which oscillates between ± 1 .

The resultant wave appears to be standing still, with no apparent movement in the x -direction, although it is composed of one wave function moving in the positive, whereas the second wave is moving in the negative x -direction. [\[link\]](#) shows various snapshots of the resulting wave. The nodes are marked with red dots while the antinodes are marked with blue dots.



When two identical waves are moving in opposite directions, the resultant wave is a standing wave. Nodes appear at integer multiples of half wavelengths. Antinodes appear at odd multiples of quarter wavelengths, where they oscillate between $y = \pm A$. The nodes are marked with red dots and the antinodes are marked with blue dots.

A common example of standing waves are the waves produced by stringed musical instruments. When the string is plucked, pulses travel along the string in opposite directions. The ends of the strings are fixed in place, so nodes appear at the ends of the strings—the boundary conditions of the system, regulating the resonant frequencies in the strings. The resonance produced on a string instrument can be modeled in a physics lab using the apparatus shown in [\[link\]](#).



A lab setup for creating standing waves on a string. The string has a node on each end and a constant linear density. The length between the fixed boundary conditions is L . The hanging mass provides the tension in the string, and the speed of the waves on the string is proportional to the square root of the tension divided by the linear mass density.

The lab setup shows a string attached to a string vibrator, which oscillates the string with an adjustable frequency f . The other end of the string passes over a frictionless pulley and is tied to a hanging mass. The magnitude of the tension in the string is equal to the weight of the hanging mass. The string has a constant linear density (mass per length) μ and the speed at which a wave travels down the string equals

$v = \sqrt{\frac{F_T}{\mu}} = \sqrt{\frac{mg}{\mu}}$ [\[link\]](#). The symmetrical boundary conditions (a node at each end) dictate the possible frequencies that can excite standing waves. Starting from a frequency of zero and slowly increasing the frequency, the first mode $n = 1$ appears as shown in [\[link\]](#). The first mode, also called the fundamental mode or the first harmonic, shows half of a wavelength has formed, so the wavelength is equal to twice the length between the nodes $\lambda_1 = 2L$. The **fundamental frequency**, or first harmonic frequency, that drives this mode is

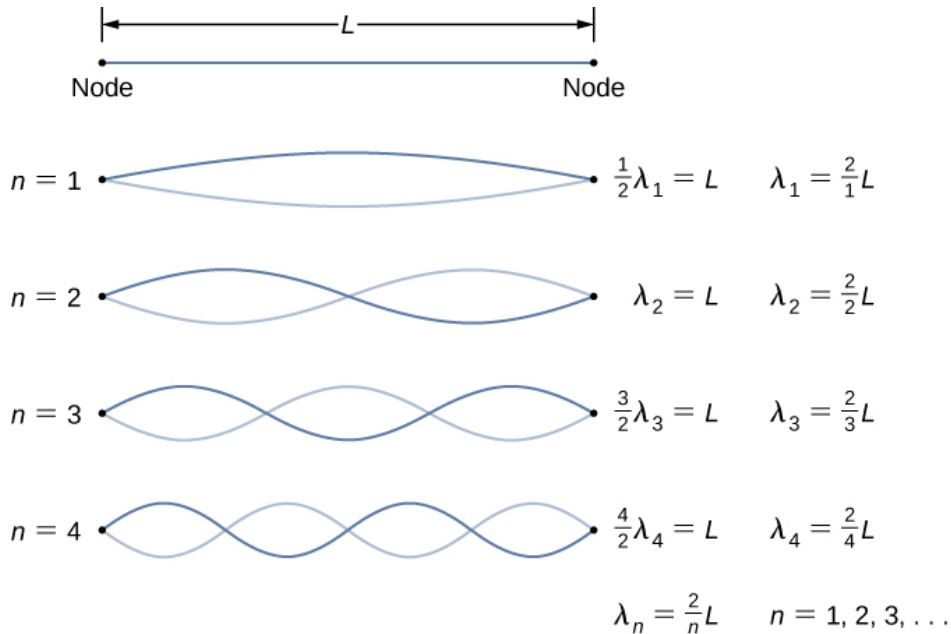
Equation:

$$f_1 = \frac{v}{\lambda_1} = \frac{v}{2L},$$

where the speed of the wave is $v = \sqrt{\frac{F_T}{\mu}}$. Keeping the tension constant and increasing the frequency leads to the second harmonic or the $n = 2$ mode. This mode is a full wavelength $\lambda_2 = L$ and the frequency is twice the fundamental frequency:

Equation:

$$f_2 = \frac{v}{\lambda_2} = \frac{v}{L} = 2f_1.$$



Standing waves created on a string of length L . A node occurs at each end of the string. The nodes are boundary conditions that limit the possible frequencies that excite standing waves. (Note that the amplitudes of the oscillations have been kept constant for visualization. The standing wave patterns possible on the string are known as the normal modes. Conducting

this experiment in the lab would result in a decrease in amplitude as the frequency increases.)

The next two modes, or the third and fourth harmonics, have wavelengths of $\lambda_3 = \frac{2}{3}L$ and $\lambda_4 = \frac{2}{4}L$, driven by frequencies of $f_3 = \frac{3v}{2L} = 3f_1$ and $f_4 = \frac{4v}{2L} = 4f_1$. All frequencies above the frequency f_1 are known as the **overtones**. The equations for the wavelength and the frequency can be summarized as:

Note:

Equation:

$$\lambda_n = \frac{2}{n}L \quad n = 1, 2, 3, 4, 5\ldots$$

Note:

Equation:

$$f_n = n \frac{v}{2L} = nf_1 \quad n = 1, 2, 3, 4, 5\ldots$$

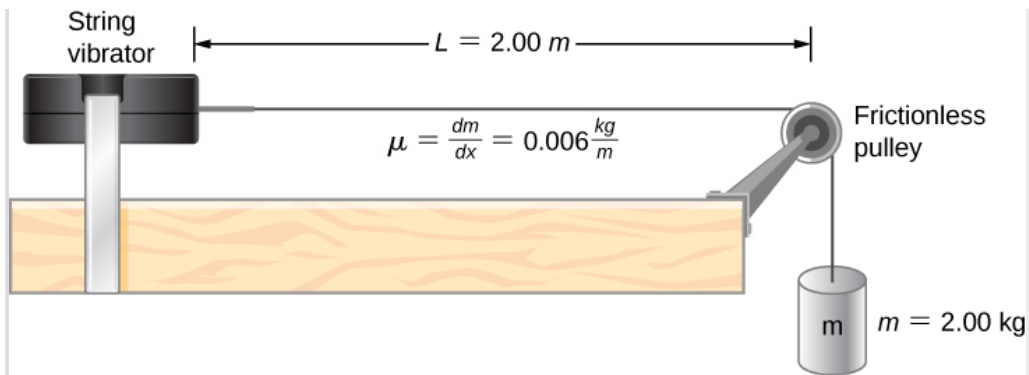
The standing wave patterns that are possible for a string, the first four of which are shown in [\[link\]](#), are known as the **normal modes**, with frequencies known as the normal frequencies. In summary, the first frequency to produce a normal mode is called the fundamental frequency (or first harmonic). Any frequencies above the fundamental frequency are overtones. The second frequency of the $n = 2$ normal mode of the string is the first overtone (or second harmonic). The frequency of the $n = 3$ normal mode is the second overtone (or third harmonic) and so on.

The solutions shown as [\[link\]](#) and [\[link\]](#) are for a string with the boundary condition of a node on each end. When the boundary condition on either side is the same, the system is said to have symmetric boundary conditions. [\[link\]](#) and [\[link\]](#) are good for any symmetric boundary conditions, that is, nodes at both ends or antinodes at both ends.

Example:

Standing Waves on a String

Consider a string of $L = 2.00$ m. attached to an adjustable-frequency string vibrator as shown in [\[link\]](#). The waves produced by the vibrator travel down the string and are reflected by the fixed boundary condition at the pulley. The string, which has a linear mass density of $\mu = 0.006$ kg/m, is passed over a frictionless pulley of a negligible mass, and the tension is provided by a 2.00-kg hanging mass. (a) What is the velocity of the waves on the string? (b) Draw a sketch of the first three normal modes of the standing waves that can be produced on the string and label each with the wavelength. (c) List the frequencies that the string vibrator must be tuned to in order to produce the first three normal modes of the standing waves.



A string attached to an adjustable-frequency string vibrator.

Strategy

- The velocity of the wave can be found using $v = \sqrt{\frac{F_T}{\mu}}$. The tension is provided by the weight of the hanging mass.
- The standing waves will depend on the boundary conditions. There must be a node at each end. The first mode will be one half of a wave. The second can be found by adding a half wavelength. That is the shortest length that will result in a node at the boundaries. For example, adding one quarter of a wavelength will result in an antinode at the boundary and is not a mode which would satisfy the boundary conditions. This is shown in [\[link\]](#).
- Since the wave speed velocity is the wavelength times the frequency, the frequency is wave speed divided by the wavelength.



(a) The figure represents the second mode of the string that satisfies the boundary conditions of a node at each end of the string. (b) This figure could not possibly be a normal mode on the string because it does not satisfy the boundary conditions. There is a node on one end, but an antinode on the other.

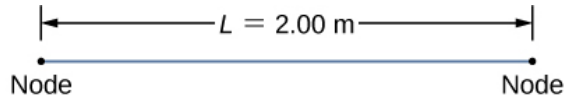
Solution

- Begin with the velocity of a wave on a string. The tension is equal to the weight of the hanging mass. The linear mass density and mass of the hanging mass are given:

Equation:

$$v = \sqrt{\frac{F_T}{\mu}} = \sqrt{\frac{mg}{\mu}} = \sqrt{\frac{2 \text{ kg} (9.8 \frac{\text{m}}{\text{s}^2})}{0.006 \frac{\text{kg}}{\text{m}}}} = 57.15 \text{ m/s}.$$

- The first normal mode that has a node on each end is a half wavelength. The next two modes are found by adding a half of a wavelength.



$$n = 1 \quad \frac{1}{2}\lambda_1 = L \quad \lambda_1 = \frac{2}{1}(2.00 \text{ m}) = 4.00 \text{ m}$$

$$n = 2 \quad \lambda_2 = L \quad \lambda_2 = \frac{2}{2}(2.00 \text{ m}) = 2.00 \text{ m}$$

$$n = 3 \quad \frac{3}{2}\lambda_3 = L \quad \lambda_3 = \frac{2}{3}(2.00 \text{ m}) = 1.33 \text{ m}$$

c. The frequencies of the first three modes are found by using $f = \frac{v_w}{\lambda}$.

Equation:

$$f_1 = \frac{v_w}{\lambda_1} = \frac{57.15 \text{ m/s}}{4.00 \text{ m}} = 14.29 \text{ Hz}$$

$$f_2 = \frac{v_w}{\lambda_2} = \frac{57.15 \text{ m/s}}{2.00 \text{ m}} = 28.58 \text{ Hz}$$

$$f_3 = \frac{v_w}{\lambda_3} = \frac{57.15 \text{ m/s}}{1.333 \text{ m}} = 42.87 \text{ Hz}$$

Significance

The three standing modes in this example were produced by maintaining the tension in the string and adjusting the driving frequency. Keeping the tension in the string constant results in a constant velocity. The same modes could have been produced by keeping the frequency constant and adjusting the speed of the wave in the string (by changing the hanging mass.)

Note:

Visit this [simulation](#) to play with a 1D or 2D system of coupled mass-spring oscillators. Vary the number of masses, set the initial conditions, and watch the system evolve. See the spectrum of normal modes for arbitrary motion. See longitudinal or transverse modes in the 1D system.

Note:

Exercise:

Problem:

Check Your Understanding The equations for the wavelengths and the frequencies of the modes of a wave produced on a string:

Equation:

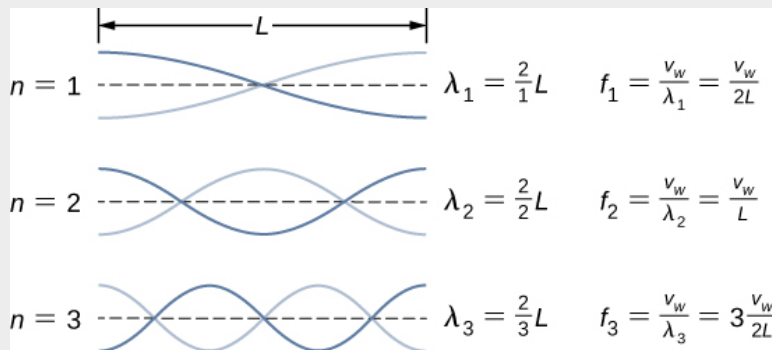
$$\lambda_n = \frac{2}{n}L \quad n = 1, 2, 3, 4, 5\ldots \text{ and}$$

$$f_n = n\frac{v}{2L} = nf_1 \quad n = 1, 2, 3, 4, 5\ldots$$

were derived by considering a wave on a string where there were symmetric boundary conditions of a node at each end. These modes resulted from two sinusoidal waves with identical characteristics except they were moving in opposite directions, confined to a region L with nodes required at both ends. Will the same equations work if there were symmetric boundary conditions with antinodes at each end? What would the normal modes look like for a medium that was free to oscillate on each end? Don't worry for now if you cannot imagine such a medium, just consider two sinusoidal wave functions in a region of length L , with antinodes on each end.

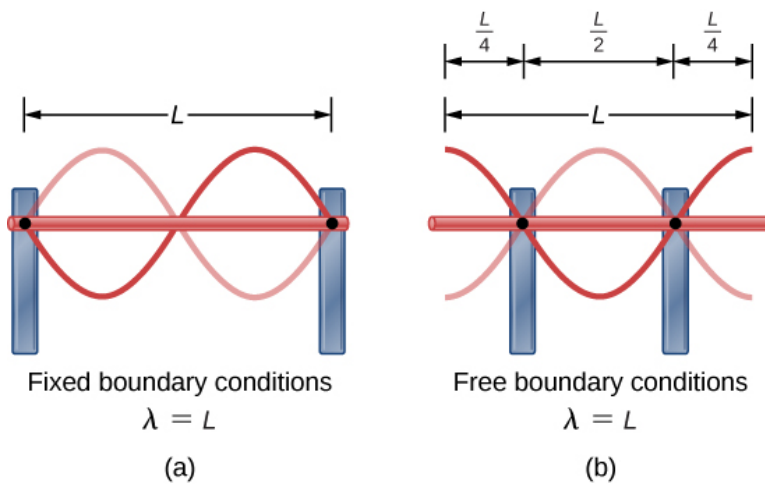
Solution:

Yes, the equations would work equally well for symmetric boundary conditions of a medium free to oscillate on each end where there was an antinode on each end. The normal modes of the first three modes are shown below. The dotted line shows the equilibrium position of the medium.

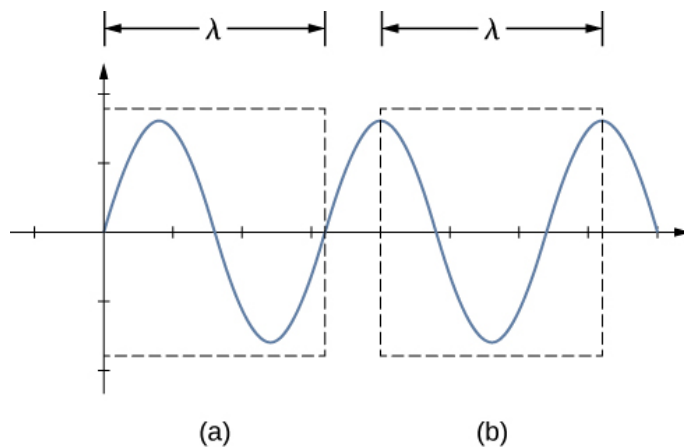


Note that the first mode is two quarters, or one half, of a wavelength. The second mode is one quarter of a wavelength, followed by one half of a wavelength, followed by one quarter of a wavelength, or one full wavelength. The third mode is one and a half wavelengths. These are the same result as the string with a node on each end. The equations for symmetrical boundary conditions work equally well for fixed boundary conditions and free boundary conditions. These results will be revisited in the next chapter when discussing sound wave in an open tube.

The free boundary conditions shown in the last Check Your Understanding may seem hard to visualize. How can there be a system that is free to oscillate on each end? In [\[link\]](#) are shown two possible configuration of a metallic rods (shown in red) attached to two supports (shown in blue). In part (a), the rod is supported at the ends, and there are fixed boundary conditions at both ends. Given the proper frequency, the rod can be driven into resonance with a wavelength equal to length of the rod, with nodes at each end. In part (b), the rod is supported at positions one quarter of the length from each end of the rod, and there are free boundary conditions at both ends. Given the proper frequency, this rod can also be driven into resonance with a wavelength equal to the length of the rod, but there are antinodes at each end. If you are having trouble visualizing the wavelength in this figure, remember that the wavelength may be measured between any two nearest identical points and consider [\[link\]](#).



(a) A metallic rod of length L (red) supported by two supports (blue) on each end. When driven at the proper frequency, the rod can resonate with a wavelength equal to the length of the rod with a node on each end. (b) The same metallic rod of length L (red) supported by two supports (blue) at a position a quarter of the length of the rod from each end. When driven at the proper frequency, the rod can resonate with a wavelength equal to the length of the rod with an antinode on each end.



A wavelength may be measure between the nearest two repeating points. On the wave on a string, this means the same height and slope. (a) The wavelength is measured between the two nearest points where the height is zero and the slope is maximum and positive. (b) The wavelength is measured between two identical points where the height is maximum and the slope is zero.

Note that the study of standing waves can become quite complex. In [\[link\]](#)(a), the $n = 2$ mode of the standing wave is shown, and it results in a wavelength equal to L . In this configuration, the $n = 1$ mode would also have been possible with a standing wave equal to $2L$. Is it possible to get the $n = 1$ mode for the configuration shown in part (b)? The answer is no. In this configuration, there are additional conditions set beyond the boundary conditions. Since the rod is mounted at a point one quarter of the length from each side, a node must exist there, and this limits the possible modes of standing waves that can be created. We leave it as an exercise for the reader to consider if other modes of standing waves are possible. It should be noted that when a system is driven at a frequency that does not cause the system to resonate, vibrations may still occur, but the amplitude of the vibrations will be much smaller than the amplitude at resonance.

A field of mechanical engineering uses the sound produced by the vibrating parts of complex mechanical systems to troubleshoot problems with the systems. Suppose a part in an automobile is resonating at the frequency of the car’s engine, causing unwanted vibrations in the automobile. This may cause the engine to fail prematurely. The engineers use microphones to record the sound produced by the engine, then use a technique called Fourier analysis to find frequencies of sound produced with large amplitudes and then look at the parts list of the automobile to find a part that would resonate at that frequency. The solution may be as simple as changing the composition of the material used or changing the length of the part in question.

There are other numerous examples of resonance in standing waves in the physical world. The air in a tube, such as found in a musical instrument like a flute, can be forced into resonance and produce a pleasant sound, as we discuss in [Sound](#).

At other times, resonance can cause serious problems. A closer look at earthquakes provides evidence for conditions appropriate for resonance, standing waves, and constructive and destructive interference. A building may vibrate for several seconds with a driving frequency matching that of the natural frequency of vibration of the building—producing a resonance resulting in one building collapsing while neighboring buildings do not. Often, buildings of a certain height are devastated while other taller buildings remain intact. The building height matches the condition for setting up a standing wave for that particular height. The span of the roof is also important. Often it is seen that gymnasiums, supermarkets, and churches suffer damage when individual homes suffer far less damage. The roofs with large surface areas supported only at the edges resonate at the frequencies of the earthquakes, causing them to collapse. As the earthquake waves travel along the surface of Earth and reflect off denser rocks, constructive interference occurs at certain points. Often areas closer to the epicenter are not damaged, while areas farther away are damaged.

Summary

- A standing wave is the superposition of two waves which produces a wave that varies in amplitude but does not propagate.
- Nodes are points of no motion in standing waves.
- An antinode is the location of maximum amplitude of a standing wave.
- Normal modes of a wave on a string are the possible standing wave patterns. The lowest frequency that will produce a standing wave is known as the fundamental frequency. The higher frequencies which produce standing waves are called overtones.

Key Equations

Wave speed	$v = \frac{\lambda}{T} = \lambda f$

Linear mass density	$\mu = \frac{\text{mass of the string}}{\text{length of the string}}$
Speed of a wave or pulse on a string under tension	$ v = \sqrt{\frac{F_T}{\mu}}$
Speed of a compression wave in a fluid	$v = \sqrt{\frac{B}{\rho}}$
Resultant wave from superposition of two sinusoidal waves that are identical except for a phase shift	$y_R(x, t) = \left[2A \cos\left(\frac{\phi}{2}\right) \right] \sin\left(kx - \omega t + \frac{\phi}{2}\right)$
Wave number	$k \equiv \frac{2\pi}{\lambda}$
Wave speed	$v = \frac{\omega}{k}$
A periodic wave	$y(x, t) = A \sin(kx \mp \omega t + \phi)$
Phase of a wave	$kx \mp \omega t + \phi$
The linear wave equation	$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v_w^2} \frac{\partial^2 y(x, t)}{\partial t^2}$
Power averaged over a wavelength	$P_{\text{ave}} = \frac{E_{\lambda}}{T} = \frac{1}{2} \mu A^2 \omega^2 \frac{\lambda}{T} = \frac{1}{2} \mu A^2 \omega^2 v$
Intensity	$I = \frac{P}{A}$
Intensity for a spherical wave	$I = \frac{P}{4\pi r^2}$
Equation of a standing wave	$y(x, t) = [2A \sin(kx)] \cos(\omega t)$
Wavelength for symmetric boundary conditions	$\lambda_n = \frac{2}{n} L, \quad n = 1, 2, 3, 4, 5 \dots$
Frequency for symmetric boundary conditions	$f_n = n \frac{v}{2L} = n f_1, \quad n = 1, 2, 3, 4, 5 \dots$

Conceptual Questions

Exercise:

Problem:

A truck manufacturer finds that a strut in the engine is failing prematurely. A sound engineer determines that the strut resonates at the frequency of the engine and suspects that this could be the problem. What are two possible characteristics of the strut can be modified to correct the problem?

Solution:

It may be as easy as changing the length and/or the density a small amount so that the parts do not resonate at the frequency of the motor.

Exercise:

Problem:

Why do roofs of gymnasiums and churches seem to fail more than family homes when an earthquake occurs?

Exercise:**Problem:**

Wine glasses can be set into resonance by moistening your finger and rubbing it around the rim of the glass. Why?

Solution:

Energy is supplied to the glass by the work done by the force of your finger on the glass. When supplied at the right frequency, standing waves form. The glass resonates and the vibrations produce sound.

Exercise:**Problem:**

Air conditioning units are sometimes placed on the roof of homes in the city. Occasionally, the air conditioners cause an undesirable hum throughout the upper floors of the homes. Why does this happen? What can be done to reduce the hum?

Exercise:**Problem:**

Consider a standing wave modeled as $y(x, t) = 4.00 \text{ cm} \sin(3 \text{ m}^{-1}x) \cos(4 \text{ s}^{-1}t)$. Is there a node or an antinode at $x = 0.00 \text{ m}$? What about a standing wave modeled as $y(x, t) = 4.00 \text{ cm} \sin(3 \text{ m}^{-1}x + \frac{\pi}{2}) \cos(4 \text{ s}^{-1}t)$? Is there a node or an antinode at the $x = 0.00 \text{ m}$ position?

Solution:

For the equation $y(x, t) = 4.00 \text{ cm} \sin(3 \text{ m}^{-1}x) \cos(4 \text{ s}^{-1}t)$, there is a node because when $x = 0.00 \text{ m}$, $\sin(3 \text{ m}^{-1}(0.00 \text{ m})) = 0.00$, so $y(0.00 \text{ m}, t) = 0.00 \text{ m}$ for all time. For the equation $y(x, t) = 4.00 \text{ cm} \sin(3 \text{ m}^{-1}x + \frac{\pi}{2}) \cos(4 \text{ s}^{-1}t)$, there is an antinode because when $x = 0.00 \text{ m}$, $\sin(3 \text{ m}^{-1}(0.00 \text{ m}) + \frac{\pi}{2}) = +1.00$, so $y(0.00 \text{ m}, t)$ oscillates between $+A$ and $-A$ as the cosine term oscillates between $+1$ and -1 .

Problems**Exercise:****Problem:**

A wave traveling on a Slinky® that is stretched to 4 m takes 2.4 s to travel the length of the Slinky and back again. (a) What is the speed of the wave? (b) Using the same Slinky stretched to the same length, a standing wave is created which consists of three antinodes and four nodes. At what frequency must the Slinky be oscillating?

Exercise:

Problem:

A 2-m long string is stretched between two supports with a tension that produces a wave speed equal to $v_w = 50.00 \text{ m/s}$. What are the wavelength and frequency of the first three modes that resonate on the string?

Solution:

$$\lambda_n = \frac{2.00}{n}L, \quad f_n = \frac{v}{\lambda_n}$$

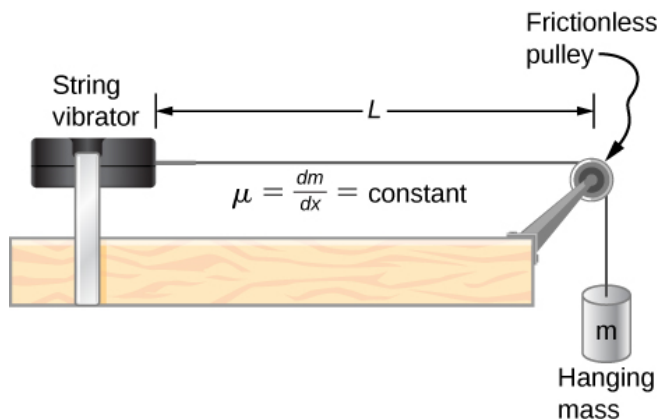
$$\lambda_1 = 4.00 \text{ m}, \quad f_1 = 12.5 \text{ Hz}$$

$$\lambda_2 = 2.00 \text{ m}, \quad f_2 = 25.00 \text{ Hz}$$

$$\lambda_3 = 1.33 \text{ m}, \quad f_3 = 37.59 \text{ Hz}$$

Exercise:**Problem:**

Consider the experimental setup shown below. The length of the string between the string vibrator and the pulley is $L = 1.00 \text{ m}$. The linear density of the string is $\mu = 0.006 \text{ kg/m}$. The string vibrator can oscillate at any frequency. The hanging mass is 2.00 kg . (a) What are the wavelength and frequency of $n = 6$ mode? (b) The string oscillates the air around the string. What is the wavelength of the sound if the speed of the sound is $v_s = 343.00 \text{ m/s}$?

**Exercise:****Problem:**

A cable with a linear density of $\mu = 0.2 \text{ kg/m}$ is hung from telephone poles. The tension in the cable is 500.00 N . The distance between poles is 20 meters . The wind blows across the line, causing the cable to resonate. A standing waves pattern is produced that has 4.5 wavelengths between the two poles. The speed of sound at the current temperature $T = 20^\circ \text{ C}$ is 343.00 m/s . What are the frequency and wavelength of the hum?

Solution:

$$v = 158.11 \text{ m/s}, \quad \lambda = 4.44 \text{ m}, \quad f = 35.61 \text{ Hz}$$

$$\lambda_s = 9.63 \text{ m}$$

Exercise:

Problem:

Consider a rod of length L , mounted in the center to a support. A node must exist where the rod is mounted on a support, as shown below. Draw the first two normal modes of the rod as it is driven into resonance. Label the wavelength and the frequency required to drive the rod into resonance.

$$\leftarrow L = 2.00 \text{ m} \rightarrow$$

**Exercise:****Problem:**

Consider two wave functions $y(x, t) = 0.30 \text{ cm} \sin(3 \text{ m}^{-1}x - 4 \text{ s}^{-1}t)$ and $y(x, t) = 0.30 \text{ cm} \sin(3 \text{ m}^{-1}x + 4 \text{ s}^{-1}t)$. Write a wave function for the resulting standing wave.

Solution:

$$y(x, t) = [0.60 \text{ cm} \sin(3 \text{ m}^{-1}x)] \cos(4 \text{ s}^{-1}t)$$

Exercise:**Problem:**

A 2.40-m wire has a mass of 7.50 g and is under a tension of 160 N. The wire is held rigidly at both ends and set into oscillation. (a) What is the speed of waves on the wire? The string is driven into resonance by a frequency that produces a standing wave with a wavelength equal to 1.20 m. (b) What is the frequency used to drive the string into resonance?

Exercise:**Problem:**

A string with a linear mass density of 0.0062 kg/m and a length of 3.00 m is set into the $n = 100$ mode of resonance. The tension in the string is 20.00 N. What is the wavelength and frequency of the wave?

Solution:

$$\lambda_{100} = 0.06 \text{ m}$$

$$v = 56.8 \text{ m/s}, \quad f_n = n f_1, \quad n = 1, 2, 3, 4, 5 \dots$$

$$f_{100} = 947 \text{ Hz}$$

Exercise:**Problem:**

A string with a linear mass density of 0.0075 kg/m and a length of 6.00 m is set into the $n = 4$ mode of resonance by driving with a frequency of 100.00 Hz. What is the tension in the string?

Exercise:

Problem:

Two sinusoidal waves with identical wavelengths and amplitudes travel in opposite directions along a string producing a standing wave. The linear mass density of the string is $\mu = 0.075 \text{ kg/m}$ and the tension in the string is $F_T = 5.00 \text{ N}$. The time interval between instances of total destructive interference is $\Delta t = 0.13 \text{ s}$. What is the wavelength of the waves?

Solution:

$$T = 2\Delta t, \quad v = \frac{\lambda}{T}, \quad \lambda = 2.12 \text{ m}$$

Exercise:**Problem:**

A string, fixed on both ends, is 5.00 m long and has a mass of 0.15 kg. The tension of the string is 90 N. The string is vibrating to produce a standing wave at the fundamental frequency of the string. (a) What is the speed of the waves on the string? (b) What is the wavelength of the standing wave produced? (c) What is the period of the standing wave?

Exercise:**Problem:**

A string is fixed at both ends. The mass of the string is 0.0090 kg and the length is 3.00 m. The string is under a tension of 200.00 N. The string is driven by a variable frequency source to produce standing waves on the string. Find the wavelengths and frequency of the first four modes of standing waves.

Solution:

$$\begin{aligned} \lambda_1 &= 6.00 \text{ m}, & \lambda_2 &= 3.00 \text{ m}, & \lambda_3 &= 2.00 \text{ m}, & \lambda_4 &= 1.50 \text{ m} \\ v &= 258.20 \text{ m/s} = \lambda f \\ f_1 &= 43.03 \text{ Hz}, & f_2 &= 86.07 \text{ Hz}, & f_3 &= 129.10 \text{ Hz}, & f_4 &= 172.13 \text{ Hz} \end{aligned}$$

Exercise:**Problem:**

The frequencies of two successive modes of standing waves on a string are 258.36 Hz and 301.42 Hz. What is the next frequency above 100.00 Hz that would produce a standing wave?

Exercise:**Problem:**

A string is fixed at both ends to supports 3.50 m apart and has a linear mass density of $\mu = 0.005 \text{ kg/m}$. The string is under a tension of 90.00 N. A standing wave is produced on the string with six nodes and five antinodes. What are the wave speed, wavelength, frequency, and period of the standing wave?

Solution:

$$v = 134.16 \text{ m/s}, \lambda = 1.4 \text{ m}, f = 95.83 \text{ Hz}, T = 0.0104 \text{ s}$$

Exercise:

Problem:

Sine waves are sent down a 1.5-m-long string fixed at both ends. The waves reflect back in the opposite direction. The amplitude of the wave is 4.00 cm. The propagation velocity of the waves is 175 m/s. The $n = 6$ resonance mode of the string is produced. Write an equation for the resulting standing wave.

Additional Problems**Exercise:****Problem:**

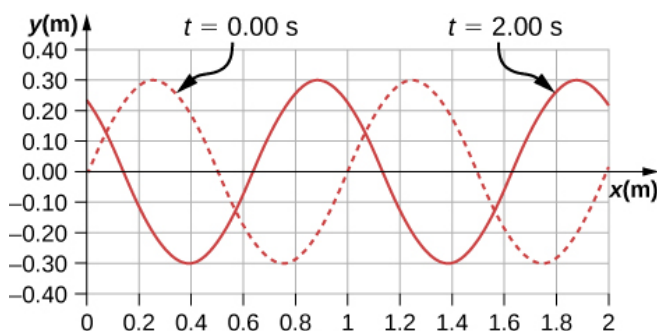
Ultrasound equipment used in the medical profession uses sound waves of a frequency above the range of human hearing. If the frequency of the sound produced by the ultrasound machine is $f = 30$ kHz, what is the wavelength of the ultrasound in bone, if the speed of sound in bone is $v = 3000$ m/s?

Solution:

$$\lambda = 0.10 \text{ m}$$

Exercise:**Problem:**

Shown below is the plot of a wave function that models a wave at time $t = 0.00$ s and $t = 2.00$ s. The dotted line is the wave function at time $t = 0.00$ s and the solid line is the function at time $t = 2.00$ s. Estimate the amplitude, wavelength, velocity, and period of the wave.

**Exercise:****Problem:**

The speed of light in air is approximately $v = 3.00 \times 10^8$ m/s and the speed of light in glass is $v = 2.00 \times 10^8$ m/s. A red laser with a wavelength of $\lambda = 633.00$ nm shines light incident of the glass, and some of the red light is transmitted to the glass. The frequency of the light is the same for the air and the glass. (a) What is the frequency of the light? (b) What is the wavelength of the light in the glass?

Solution:

$$\text{a. } f = 4.74 \times 10^{14} \text{ Hz; b. } \lambda = 422 \text{ nm}$$

Exercise:

Problem:

A radio station broadcasts radio waves at a frequency of 101.7 MHz. The radio waves move through the air at approximately the speed of light in a vacuum. What is the wavelength of the radio waves?

Exercise:**Problem:**

A sunbather stands waist deep in the ocean and observes that six crests of periodic surface waves pass each minute. The crests are 16.00 meters apart. What is the wavelength, frequency, period, and speed of the waves?

Solution:

$$\lambda = 16.00 \text{ m}, \quad f = 0.10 \text{ Hz}, \quad T = 10.00 \text{ s}, \quad v = 1.6 \text{ m/s}$$

Exercise:**Problem:**

A tuning fork vibrates producing sound at a frequency of 512 Hz. The speed of sound in air is $v = 343.00 \text{ m/s}$ if the air is at a temperature of 20.00°C . What is the wavelength of the sound?

Exercise:**Problem:**

A motorboat is traveling across a lake at a speed of $v_b = 15.00 \text{ m/s}$. The boat bounces up and down every 0.50 s as it travels in the same direction as a wave. It bounces up and down every 0.30 s as it travels in a direction opposite the direction of the waves. What is the speed and wavelength of the wave?

Solution:

$$\lambda = (v_b + v)t_b, \quad v = 3.75 \text{ m/s}, \quad \lambda = 3.00 \text{ m}$$

Exercise:**Problem:**

Use the linear wave equation to show that the wave speed of a wave modeled with the wave function $y(x, t) = 0.20 \text{ m} \sin(3.00 \text{ m}^{-1}x + 6.00 \text{ s}^{-1}t)$ is $v = 2.00 \text{ m/s}$. What are the wavelength and the speed of the wave?

Exercise:**Problem:**

Given the wave functions $y_1(x, t) = A \sin(kx - \omega t)$ and $y_2(x, t) = A \sin(kx - \omega t + \phi)$ with $\phi \neq \frac{\pi}{2}$, show that $y_1(x, t) + y_2(x, t)$ is a solution to the linear wave equation with a wave velocity of $v = \frac{\omega}{k}$.

Solution:

$$\frac{\partial^2(y_1+y_2)}{\partial t^2} = -A\omega^2 \sin(kx - \omega t) - A\omega^2 \sin(kx - \omega t + \phi)$$

$$\frac{\partial^2(y_1+y_2)}{\partial x^2} = -Ak^2 \sin(kx - \omega t) - Ak^2 \sin(kx - \omega t + \phi)$$

$$\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y(x,t)}{\partial t^2}$$

$$-A\omega^2 \sin(kx - \omega t) - A\omega^2 \sin(kx - \omega t + \phi) = \left(\frac{1}{v^2}\right) (-Ak^2 \sin(kx - \omega t) - Ak^2 \sin(kx - \omega t + \phi))$$

$$v = \frac{\omega}{k}$$

Exercise:

Problem:

A transverse wave on a string is modeled with the wave function

$y(x, t) = 0.10 \text{ m} \sin(0.15 \text{ m}^{-1}x + 1.50 \text{ s}^{-1}t + 0.20)$. (a) Find the wave velocity. (b) Find the position in the y -direction, the velocity perpendicular to the motion of the wave, and the acceleration perpendicular to the motion of the wave, of a small segment of the string centered at $x = 0.40 \text{ m}$ at time $t = 5.00 \text{ s}$.

Exercise:

Problem:

A sinusoidal wave travels down a taut, horizontal string with a linear mass density of $\mu = 0.060 \text{ kg/m}$. The magnitude of maximum vertical acceleration of the wave is $a_{y\text{max}} = 0.90 \text{ cm/s}^2$ and the amplitude of the wave is 0.40 m . The string is under a tension of $F_T = 600.00 \text{ N}$. The wave moves in the negative x -direction. Write an equation to model the wave.

Solution:

$$y(x, t) = 0.40 \text{ m} \sin(0.015 \text{ m}^{-1}x + 1.5 \text{ s}^{-1}t)$$

Exercise:

Problem:

A transverse wave on a string ($\mu = 0.0030 \text{ kg/m}$) is described with the equation

$y(x, t) = 0.30 \text{ m} \sin\left(\frac{2\pi}{4.00 \text{ m}}(x - 16.00 \frac{\text{m}}{\text{s}}t)\right)$. What is the tension under which the string is held taut?

Exercise:

Problem:

A transverse wave on a horizontal string ($\mu = 0.0060 \text{ kg/m}$) is described with the equation

$y(x, t) = 0.30 \text{ m} \sin\left(\frac{2\pi}{4.00 \text{ m}}(x - v_w t)\right)$. The string is under a tension of 300.00 N . What are the wave speed, wave number, and angular frequency of the wave?

Solution:

$$v = 223.61 \text{ m/s}, k = 1.57 \text{ m}^{-1}, \omega = 142.43 \text{ s}^{-1}$$

Exercise:

Problem:

A student holds an inexpensive sonic range finder and uses the range finder to find the distance to the wall. The sonic range finder emits a sound wave. The sound wave reflects off the wall and returns to the range finder. The round trip takes 0.012 s. The range finder was calibrated for use at room temperature $T = 20^\circ\text{C}$, but the temperature in the room is actually $T = 23^\circ\text{C}$. Assuming that the timing mechanism is perfect, what percentage of error can the student expect due to the calibration?

Exercise:**Problem:**

A wave on a string is driven by a string vibrator, which oscillates at a frequency of 100.00 Hz and an amplitude of 1.00 cm. The string vibrator operates at a voltage of 12.00 V and a current of 0.20 A. The power consumed by the string vibrator is $P = IV$. Assume that the string vibrator is 90% efficient at converting electrical energy into the energy associated with the vibrations of the string. The string is 3.00 m long, and is under a tension of 60.00 N. What is the linear mass density of the string?

Solution:

$$P = \frac{1}{2} A^2 (2\pi f)^2 \sqrt{\mu F_T}$$

$$\mu = 2.00 \times 10^{-4} \text{ kg/m}$$

Exercise:**Problem:**

A traveling wave on a string is modeled by the wave equation $y(x, t) = 3.00 \text{ cm} \sin(8.00 \text{ m}^{-1}x + 100.00 \text{ s}^{-1}t)$. The string is under a tension of 50.00 N and has a linear mass density of $\mu = 0.008 \text{ kg/m}$. What is the average power transferred by the wave on the string?

Exercise:**Problem:**

A transverse wave on a string has a wavelength of 5.0 m, a period of 0.02 s, and an amplitude of 1.5 cm. The average power transferred by the wave is 5.00 W. What is the tension in the string?

Solution:

$$P = \frac{1}{2} \mu A^2 \omega^2 \frac{\lambda}{T}, \mu = 0.0018 \text{ kg/m}$$

Exercise:**Problem:**

(a) What is the intensity of a laser beam used to burn away cancerous tissue that, when 90.0% absorbed, puts 500.0 J of energy into a circular spot 2.00 mm in diameter in 4.00 s? (b) Discuss how this intensity compares to the average intensity of sunlight (about 1 kW/m^2) and the implications that would have if the laser beam entered your eye. Note how your answer depends on the time duration of the exposure.

Exercise:

Problem:

Consider two periodic wave functions, $y_1(x, t) = A \sin(kx - \omega t)$ and $y_2(x, t) = A \sin(kx - \omega t + \phi)$. (a) For what values of ϕ will the wave that results from a superposition of the wave functions have an amplitude of $2A$? (b) For what values of ϕ will the wave that results from a superposition of the wave functions have an amplitude of zero?

Solution:

a. $A_R = 2A \cos\left(\frac{\phi}{2}\right)$, $\cos\left(\frac{\phi}{2}\right) = 1$, $\phi = 0, 2\pi, 4\pi, \dots$; b.
 $A_R = 2A \cos\left(\frac{\phi}{2}\right)$, $\cos\left(\frac{\phi}{2}\right) = 0$, $\phi = 0, \pi, 3\pi, 5\pi, \dots$

Exercise:**Problem:**

Consider two periodic wave functions, $y_1(x, t) = A \sin(kx - \omega t)$ and $y_2(x, t) = A \cos(kx - \omega t + \phi)$. (a) For what values of ϕ will the wave that results from a superposition of the wave functions have an amplitude of $2A$? (b) For what values of ϕ will the wave that results from a superposition of the wave functions have an amplitude of zero?

Exercise:**Problem:**

A trough with dimensions 10.00 meters by 0.10 meters by 0.10 meters is partially filled with water. Small-amplitude surface water waves are produced from both ends of the trough by paddles oscillating in simple harmonic motion. The height of the water waves are modeled with two sinusoidal wave equations, $y_1(x, t) = 0.3 \text{ m} \sin(4 \text{ m}^{-1}x - 3 \text{ s}^{-1}t)$ and $y_2(x, t) = 0.3 \text{ m} \cos(4 \text{ m}^{-1}x + 3 \text{ s}^{-1}t - \frac{\pi}{2})$. What is the wave function of the resulting wave after the waves reach one another and before they reach the end of the trough (i.e., assume that there are only two waves in the trough and ignore reflections)? Use a spreadsheet to check your results. (*Hint:* Use the trig identities $\sin(u \pm v) = \sin u \cos v \pm \cos u \sin v$ and $\cos(u \pm v) = \cos u \cos v \mp \sin u \sin v$)

Solution:

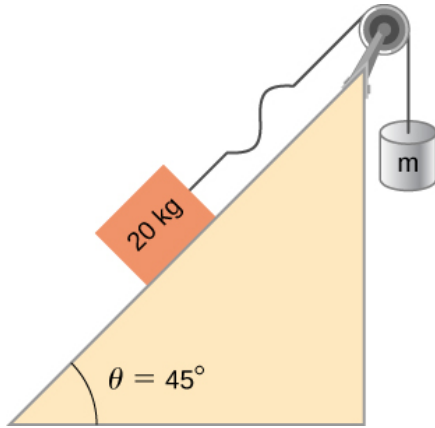
$$y_R(x, t) = 0.6 \text{ m} \sin(4 \text{ m}^{-1}x) \cos(3 \text{ s}^{-1}t)$$

Exercise:**Problem:**

A seismograph records the S- and P-waves from an earthquake 20.00 s apart. If they traveled the same path at constant wave speeds of $v_S = 4.00 \text{ km/s}$ and $v_P = 7.50 \text{ km/s}$, how far away is the epicenter of the earthquake?

Exercise:**Problem:**

Consider what is shown below. A 20.00-kg mass rests on a frictionless ramp inclined at 45° . A string with a linear mass density of $\mu = 0.025 \text{ kg/m}$ is attached to the 20.00-kg mass. The string passes over a frictionless pulley of negligible mass and is attached to a hanging mass (m). The system is in static equilibrium. A wave is induced on the string and travels up the ramp. (a) What is the mass of the hanging mass (m)? (b) At what wave speed does the wave travel up the string?



Solution:

$$\begin{aligned}
 (1) F_T - 20.00 \text{ kg} (9.80 \text{ m/s}^2) \cos 45^\circ &= 0 \\
 \text{a. } (2) m (9.80 \text{ m/s}^2) - F_T &= 0 & \text{; b. } F_T &= 138.57 \text{ N} \\
 m &= 14.14 \text{ kg} & v &= 74.45 \text{ m/s}
 \end{aligned}$$

Exercise:

Problem:

Consider the superposition of three wave functions $y(x, t) = 3.00 \text{ cm} \sin(2 \text{ m}^{-1}x - 3 \text{ s}^{-1}t)$, $y(x, t) = 3.00 \text{ cm} \sin(6 \text{ m}^{-1}x + 3 \text{ s}^{-1}t)$, and $y(x, t) = 3.00 \text{ cm} \sin(2 \text{ m}^{-1}x - 4 \text{ s}^{-1}t)$. What is the height of the resulting wave at position $x = 3.00 \text{ m}$ at time $t = 10.0 \text{ s}$?

Exercise:

Problem:

A string has a mass of 150 g and a length of 3.4 m. One end of the string is fixed to a lab stand and the other is attached to a spring with a spring constant of $k_s = 100 \text{ N/m}$. The free end of the spring is attached to another lab pole. The tension in the string is maintained by the spring. The lab poles are separated by a distance that stretches the spring 2.00 cm. The string is plucked and a pulse travels along the string. What is the propagation speed of the pulse?

Solution:

$$F_T = 2 \text{ N}, v = 6.73 \text{ m/s}$$

Exercise:

Problem:

A standing wave is produced on a string under a tension of 70.0 N by two sinusoidal transverse waves that are identical, but moving in opposite directions. The string is fixed at $x = 0.00 \text{ m}$ and $x = 10.00 \text{ m}$. Nodes appear at $x = 0.00 \text{ m}$, 2.00 m, 4.00 m, 6.00 m, 8.00 m, and 10.00 m. The amplitude of the standing wave is 3.00 cm. It takes 0.10 s for the antinodes to make one complete oscillation. (a) What are the wave functions of the two sine waves that produce the standing wave? (b) What are the maximum velocity and acceleration of the string, perpendicular to the direction of motion of the transverse waves, at the antinodes?

Exercise:**Problem:**

A string with a length of 4 m is held under a constant tension. The string has a linear mass density of $\mu = 0.006 \text{ kg/m}$. Two resonant frequencies of the string are 400 Hz and 480 Hz. There are no resonant frequencies between the two frequencies. (a) What are the wavelengths of the two resonant modes? (b) What is the tension in the string?

Solution:

$$\text{a. } f_n = \frac{nv}{2L}, v = \frac{2Lf_{n+1}}{n+1}, \frac{n+1}{n} = \frac{2Lf_{n+1}}{2Lf_n}, 1 + \frac{1}{n} = 1.2, n = 5; \text{ b. } F_T = 245.76 \text{ N}$$

$$\lambda_n = \frac{2}{n}L, \lambda_5 = 1.6 \text{ m}, \lambda_6 = 1.33 \text{ m}$$

Challenge Problems**Exercise:****Problem:**

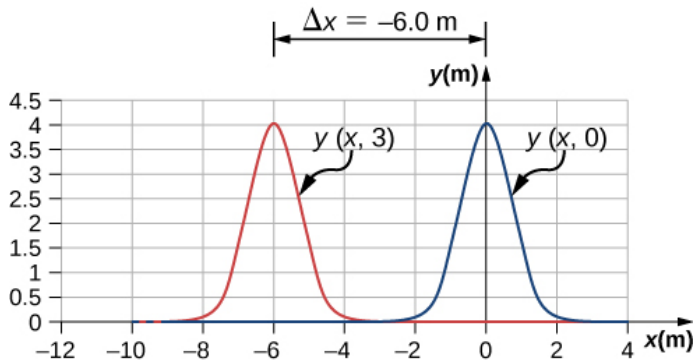
A copper wire has a radius of 200 μm and a length of 5.0 m. The wire is placed under a tension of 3000 N and the wire stretches by a small amount. The wire is plucked and a pulse travels down the wire. What is the propagation speed of the pulse? (Assume the temperature does not change: $(\rho = 8.96 \frac{\text{g}}{\text{cm}^3}, Y = 1.1 \times 10^{11} \frac{\text{N}}{\text{m}}).$)

Exercise:**Problem:**

A pulse moving along the x axis can be modeled as the wave function $y(x, t) = 4.00 \text{ m} e^{-\left(\frac{x+(2.00 \text{ m/s})t}{1.00 \text{ m}}\right)^2}$. (a) What are the direction and propagation speed of the pulse? (b) How far has the wave moved in 3.00 s? (c) Plot the pulse using a spreadsheet at time $t = 0.00 \text{ s}$ and $t = 3.00 \text{ s}$ to verify your answer in part (b).

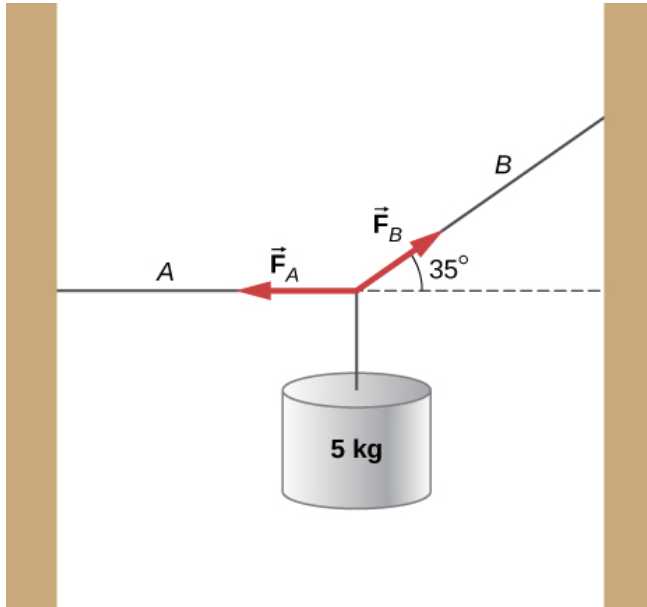
Solution:

a. Moves in the negative x direction at a propagation speed of $v = 2.00 \text{ m/s}$. b. $\Delta x = -6.00 \text{ m}$; c.

Wave Function vs. Time**Exercise:**

Problem:

A string with a linear mass density of $\mu = 0.0085 \text{ kg/m}$ is fixed at both ends. A 5.0-kg mass is hung from the string, as shown below. If a pulse is sent along section A, what is the wave speed in section A and the wave speed in section B?

**Exercise:****Problem:**

Consider two wave functions $y_1(x, t) = A \sin(kx - \omega t)$ and $y_2(x, t) = A \sin(kx + \omega t + \phi)$. What is the wave function resulting from the interference of the two wave? (*Hint:*

$\sin(\alpha \pm \beta) = \sin \alpha \cos \beta \pm \cos \alpha \sin \beta$ and $\phi = \frac{\phi}{2} + \frac{\phi}{2}$.)

Solution:

$$\sin(kx - \omega t) = \sin\left(kx + \frac{\phi}{2}\right) \cos\left(\omega t + \frac{\phi}{2}\right) - \cos\left(kx + \frac{\phi}{2}\right) \sin\left(\omega t + \frac{\phi}{2}\right)$$

$$\sin(kx - \omega t + \phi) = \sin\left(kx + \frac{\phi}{2}\right) \cos\left(\omega t + \frac{\phi}{2}\right) + \cos\left(kx + \frac{\phi}{2}\right) \sin\left(\omega t + \frac{\phi}{2}\right)$$

$$\sin(kx - \omega t) + \sin(kx + \omega t + \phi) = 2 \sin\left(kx + \frac{\phi}{2}\right) \cos\left(\omega t + \frac{\phi}{2}\right)$$

$$y_R = 2 A \sin\left(kx + \frac{\phi}{2}\right) \cos\left(\omega t + \frac{\phi}{2}\right)$$

Exercise:**Problem:**

The wave function that models a standing wave is given as

$y_R(x, t) = 6.00 \text{ cm} \sin(3.00 \text{ m}^{-1}x + 1.20 \text{ rad}) \cos(6.00 \text{ s}^{-1}t + 1.20 \text{ rad})$. What are two wave functions that interfere to form this wave function? Plot the two wave functions and the sum of the sum of the two wave functions at $t = 1.00 \text{ s}$ to verify your answer.

Exercise:

Problem:

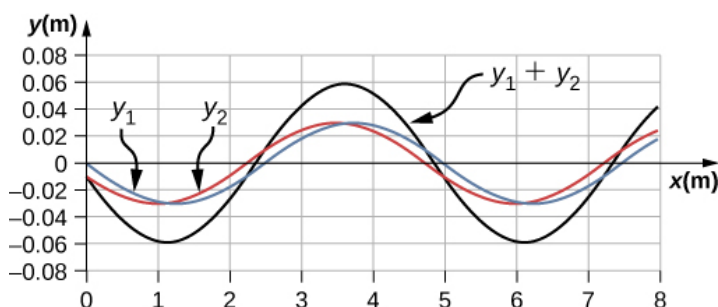
Consider two wave functions $y_1(x, t) = A \sin(kx - \omega t)$ and $y_2(x, t) = A \sin(kx + \omega t + \phi)$. The resultant wave form when you add the two functions is $y_R = 2A \sin\left(kx + \frac{\phi}{2}\right) \cos\left(\omega t + \frac{\phi}{2}\right)$.

Consider the case where $A = 0.03 \text{ m}$, $k = 1.26 \text{ m}^{-1}$, $\omega = \pi \text{ s}^{-1}$, and $\phi = \frac{\pi}{10}$. (a) Where are the first three nodes of the standing wave function starting at zero and moving in the positive x direction? (b) Using a spreadsheet, plot the two wave functions and the resulting function at time $t = 1.00 \text{ s}$ to verify your answer.

Solution:

$$\sin\left(kx + \frac{\phi}{2}\right) = 0, \quad kx + \frac{\phi}{2} = 0, \pi, 2\pi, \quad 1.26 \text{ m}^{-1}x + \frac{\pi}{20} = \pi, 2\pi, 3\pi,$$

$$x = 2.37 \text{ m}, 4.86 \text{ m}, 7.35 \text{ m}$$

**Glossary**

antinode

location of maximum amplitude in standing waves

fundamental frequency

lowest frequency that will produce a standing wave

node

point where the string does not move; more generally, nodes are where the wave disturbance is zero in a standing wave

normal mode

possible standing wave pattern for a standing wave on a string

overtone

frequency that produces standing waves and is higher than the fundamental frequency

standing wave

wave that can bounce back and forth through a particular region, effectively becoming stationary

Introduction

class="introduction"

Hearing is an important human sense that can detect frequencies of sound, ranging between 20 Hz and 20 kHz.

However, other species have very different ranges of hearing. Bats, for example, emit clicks in ultrasound, using frequencies beyond 20 kHz. They can detect nearby insects by hearing the echo of these ultrasonic clicks.

Ultrasound is important in several human applications, including

probing the
interior
structures of
human bodies,
Earth, and the
Sun.

Ultrasound is
also useful in
industry for
nondestructiv
e testing.

(credit:
modification
of work by
Angell
Williams)



Sound is an example of a mechanical wave, specifically, a pressure wave: Sound waves travel through the air and other media as oscillations of molecules. Normal human hearing encompasses an impressive range of frequencies from 20 Hz to 20 kHz. Sounds below 20 Hz are called infrasound, whereas those above 20 kHz are called ultrasound. Some

animals, like the bat shown in [\[link\]](#), can hear sounds in the ultrasonic range.

Many of the concepts covered in [Waves](#) also have applications in the study of sound. For example, when a sound wave encounters an interface between two media with different wave speeds, reflection and transmission of the wave occur.

Ultrasound has many uses in science, engineering, and medicine. Ultrasound is used for nondestructive testing in engineering, such as testing the thickness of coating on metal. In medicine, sound waves are far less destructive than X-rays and can be used to image the fetus in a mother's womb without danger to the fetus or the mother. Later in this chapter, we discuss the Doppler effect, which can be used to determine the velocity of blood in the arteries or wind speed in weather systems.

Sound Waves

By the end of this section, you will be able to:

- Explain the difference between sound and hearing
- Describe sound as a wave
- List the equations used to model sound waves
- Describe compression and rarefactions as they relate to sound

The physical phenomenon of **sound** is a disturbance of matter that is transmitted from its source outward. **Hearing** is the perception of sound, just as seeing is the perception of visible light. On the atomic scale, sound is a disturbance of atoms that is far more ordered than their thermal motions. In many instances, sound is a periodic wave, and the atoms undergo simple harmonic motion. Thus, sound waves can induce oscillations and resonance effects ([\[link\]](#)).



This glass has been shattered by a high-intensity sound wave of the same frequency as the resonant frequency of the glass. (credit: “||read||”/Flickr)

Note:

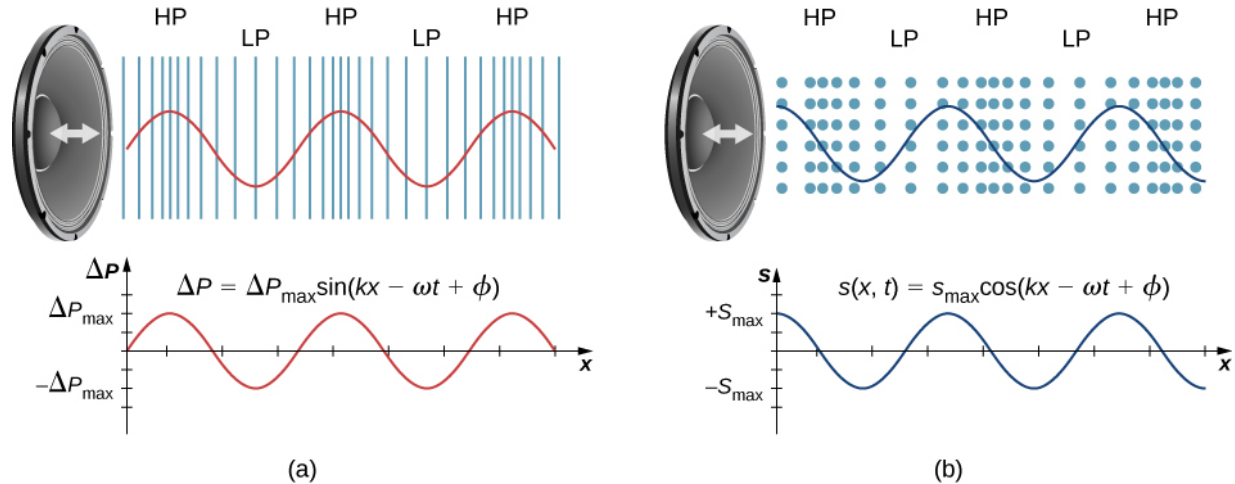
This [video](#) shows waves on the surface of a wine glass, being driven by sound waves from a speaker. As the frequency of the sound wave approaches the resonant frequency of the wine glass, the amplitude and frequency of the waves on the wine glass increase. When the resonant frequency is reached, the glass shatters.

A speaker produces a sound wave by oscillating a cone, causing vibrations of air molecules. In [\[link\]](#), a speaker vibrates at a constant frequency and amplitude, producing vibrations in the surrounding air molecules. As the speaker oscillates back and forth, it transfers energy to the air, mostly as thermal energy. But a small part of the speaker’s energy goes into compressing and expanding the surrounding air, creating slightly higher and lower local pressures. These compressions (high-pressure regions) and rarefactions (low-pressure regions) move out as longitudinal pressure waves having the same frequency as the speaker—they are the disturbance that is a sound wave. (Sound waves in air and most fluids are longitudinal, because fluids have almost no shear strength. In solids, sound waves can be both transverse and longitudinal.)

[\[link\]](#)(a) shows the compressions and rarefactions, and also shows a graph of gauge pressure versus distance from a speaker. As the speaker moves in the positive x -direction, it pushes air molecules, displacing them from their equilibrium positions. As the speaker moves in the negative x -direction, the air molecules move back toward their equilibrium positions due to a restoring force. The air molecules oscillate in simple harmonic motion about their equilibrium positions, as shown in part (b). Note that sound

waves in air are longitudinal, and in the figure, the wave propagates in the positive x -direction and the molecules oscillate parallel to the direction in which the wave propagates.

HP = Compression LP = Rarefaction



(a) A vibrating cone of a speaker, moving in the positive x -direction, compresses the air in front of it and expands the air behind it. As the speaker oscillates, it creates another compression and rarefaction as those on the right move away from the speaker. After many vibrations, a series of compressions and rarefactions moves out from the speaker as a sound wave. The red graph shows the gauge pressure of the air versus the distance from the speaker. Pressures vary only slightly from atmospheric pressure for ordinary sounds. Note that gauge pressure is modeled with a sine function, where the crests of the function line up with the compressions and the troughs line up with the rarefactions. (b)

Sound waves can also be modeled using the displacement of the air molecules. The blue graph shows the displacement of the air molecules versus the position from the speaker and is modeled with a cosine function. Notice that the displacement is zero for the molecules in their equilibrium position and are centered at the compressions and rarefactions. Compressions are formed when molecules on either side of the equilibrium molecules are displaced toward the equilibrium position. Rarefactions are formed when the molecules are displaced away from the equilibrium position.

Note:**Models Describing Sound**

Sound can be modeled as a pressure wave by considering the change in pressure from average pressure,

Equation:

$$\Delta P = \Delta P_{\max} \sin(kx \mp \omega t + \phi).$$

This equation is similar to the periodic wave equations seen in [Waves](#), where ΔP is the change in pressure, ΔP_{\max} is the maximum change in pressure, $k = \frac{2\pi}{\lambda}$ is the wave number, $\omega = \frac{2\pi}{T} = 2\pi f$ is the angular frequency, and ϕ is the initial phase. The wave speed can be determined from $v = \frac{\omega}{k} = \frac{\lambda}{T}$. Sound waves can also be modeled in terms of the displacement of the air molecules. The displacement of the air molecules can be modeled using a cosine function:

Equation:

$$s(x, t) = s_{\max} \cos(kx \mp \omega t + \phi).$$

In this equation, s is the displacement and s_{\max} is the maximum displacement.

Not shown in the figure is the amplitude of a sound wave as it decreases with distance from its source, because the energy of the wave is spread over a larger and larger area. The intensity decreases as it moves away from the speaker, as discussed in [Waves](#). The energy is also absorbed by objects and converted into thermal energy by the viscosity of the air. In addition, during each compression, a little heat transfers to the air; during each rarefaction, even less heat transfers from the air, and these heat transfers reduce the organized disturbance into random thermal motions. Whether the heat transfer from compression to rarefaction is significant depends on how far apart they are—that is, it depends on wavelength. Wavelength, frequency, amplitude, and speed of propagation are important characteristics for sound, as they are for all waves.

Summary

- Sound is a disturbance of matter (a pressure wave) that is transmitted from its source outward. Hearing is the perception of sound.
- Sound can be modeled in terms of pressure or in terms of displacement of molecules.
- The human ear is sensitive to frequencies between 20 Hz and 20 kHz.

Conceptual Questions

Exercise:

Problem: What is the difference between sound and hearing?

Solution:

Sound is a disturbance of matter (a pressure wave) that is transmitted from its source outward. Hearing is the human perception of sound.

Exercise:

Problem:

You will learn that light is an electromagnetic wave that can travel through a vacuum. Can sound waves travel through a vacuum?

Exercise:

Problem:

Sound waves can be modeled as a change in pressure. Why is the change in pressure used and not the actual pressure?

Solution:

Consider a sound wave moving through air. The pressure of the air is the equilibrium condition, it is the change in pressure that produces the sound wave.

Problems

Exercise:

Problem:

Consider a sound wave modeled with the equations $(x, t) = 4.00 \text{ nm} \cos(3.66 \text{ m}^{-1}x - 1256 \text{ s}^{-1}t)$. What is the maximum displacement, the wavelength, the frequency, and the speed of the sound wave?

Solution:

$$s_{\max} = 4.00 \text{ nm}, \quad \lambda = 1.72 \text{ m}, \quad f = 200 \text{ Hz}, \quad v = 343.17 \text{ m/s}$$

Exercise:

Problem:

Consider a sound wave moving through the air modeled with the equations $(x, t) = 6.00 \text{ nm} \cos(54.93 \text{ m}^{-1}x - 18.84 \times 10^3 \text{ s}^{-1}t)$. What is the shortest time required for an air molecule to move between 3.00 nm and -3.00 nm ?

Exercise:

Problem:

Consider a diagnostic ultrasound of frequency 5.00 MHz that is used to examine an irregularity in soft tissue. (a) What is the wavelength in air of such a sound wave if the speed of sound is 343 m/s ? (b) If the speed of sound in tissue is 1800 m/s , what is the wavelength of this wave in tissue?

Solution:

$$\text{a. } \lambda = 68.60 \mu\text{m}; \text{ b. } \lambda = 360.00 \mu\text{m}$$

Exercise:

Problem:

A sound wave is modeled as $\Delta P = 1.80 \text{ Pa} \sin (55.41 \text{ m}^{-1} x - 18,840 \text{ s}^{-1} t)$. What is the maximum change in pressure, the wavelength, the frequency, and the speed of the sound wave?

Exercise:**Problem:**

A sound wave is modeled with the wave function $\Delta P = 1.20 \text{ Pa} \sin (kx - 6.28 \times 10^4 \text{ s}^{-1} t)$ and the sound wave travels in air at a speed of $v = 343.00 \text{ m/s}$. (a) What is the wave number of the sound wave? (b) What is the value for ΔP (3.00 m, 20.00 s)?

Solution:

- a. $k = 183.09 \text{ m}^{-1}$;
- b. $\Delta P = -1.11 \text{ Pa}$

Exercise:**Problem:**

The displacement of the air molecules in sound wave is modeled with the wave function $s(x, t) = 5.00 \text{ nm} \cos (91.54 \text{ m}^{-1} x - 3.14 \times 10^4 \text{ s}^{-1} t)$. (a) What is the wave speed of the sound wave? (b) What is the maximum speed of the air molecules as they oscillate in simple harmonic motion? (c) What is the magnitude of the maximum acceleration of the air molecules as they oscillate in simple harmonic motion?

Exercise:

Problem:

A speaker is placed at the opening of a long horizontal tube. The speaker oscillates at a frequency f , creating a sound wave that moves down the tube. The wave moves through the tube at a speed of $v = 340.00 \text{ m/s}$. The sound wave is modeled with the wave function $s(x, t) = s_{\max} \cos(kx - \omega t + \phi)$. At time $t = 0.00 \text{ s}$, an air molecule at $x = 3.5 \text{ m}$ is at the maximum displacement of 7.00 nm . At the same time, another molecule at $x = 3.7 \text{ m}$ has a displacement of 3.00 nm . What is the frequency at which the speaker is oscillating?

Solution:

$$s_1 = 7.00 \text{ nm}, \quad s_2 = 3.00 \text{ nm}, \quad kx_1 + \phi = 0 \text{ rad}$$

$$kx_2 + \phi = 1.128 \text{ rad}$$

$$k(x_2 - x_1) = 1.128 \text{ rad}, \quad k = 5.64 \text{ m}^{-1}$$

$$\lambda = 1.11 \text{ m}, \quad f = 306.31 \text{ Hz}$$

Exercise:**Problem:**

A 250-Hz tuning fork is struck and begins to vibrate. A sound-level meter is located 34.00 m away. It takes the sound $\Delta t = 0.10 \text{ s}$ to reach the meter. The maximum displacement of the tuning fork is 1.00 mm . Write a wave function for the sound.

Exercise:

Problem:

A sound wave produced by an ultrasonic transducer, moving in air, is modeled with the wave equation

$$s(x, t) = 4.50 \text{ nm} \cos(9.15 \times 10^4 \text{ m}^{-1}x -$$

$2\pi(5.00 \text{ MHz})t)$. The transducer is to be used in nondestructive testing to test for fractures in steel beams. The speed of sound in the steel beam is $v = 5950 \text{ m/s}$. Find the wave function for the sound wave in the steel beam.

Solution:

$$k = 5.28 \times 10^3 \text{ m}^{-1}$$

$$s(x, t) = 4.50 \text{ nm} \cos(5.28 \times 10^3 \text{ m}^{-1}x - 2\pi(5.00 \text{ MHz})t)$$

Exercise:**Problem:**

Porpoises emit sound waves that they use for navigation. If the wavelength of the sound wave emitted is 4.5 cm, and the speed of sound in the water is $v = 1530 \text{ m/s}$, what is the period of the sound?

Exercise:**Problem:**

Bats use sound waves to catch insects. Bats can detect frequencies up to 100 kHz. If the sound waves travel through air at a speed of $v = 343 \text{ m/s}$, what is the wavelength of the sound waves?

Solution:

$$\lambda = 3.43 \text{ mm}$$

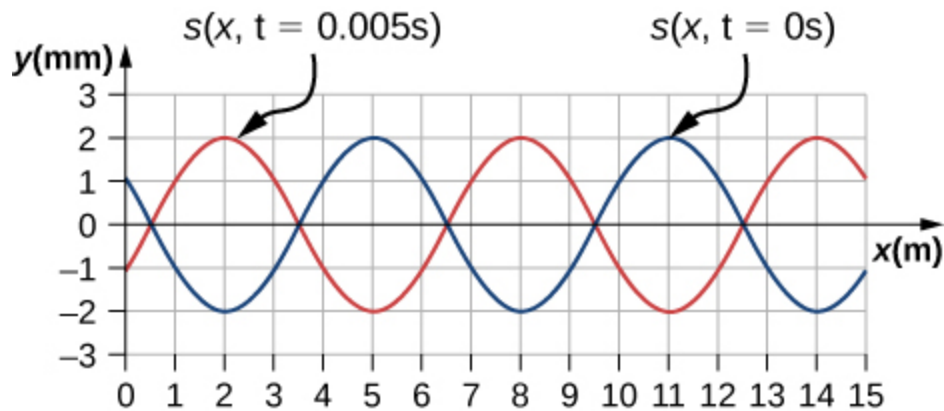
Exercise:

Problem:

A bat sends of a sound wave 100 kHz and the sound waves travel through air at a speed of $v = 343 \text{ m/s}$. (a) If the maximum pressure difference is 1.30 Pa, what is a wave function that would model the sound wave, assuming the wave is sinusoidal? (Assume the phase shift is zero.) (b) What are the period and wavelength of the sound wave?

Exercise:**Problem:**

Consider the graph shown below of a compression wave. Shown are snapshots of the wave function for $t = 0.000 \text{ s}$ (blue) and $t = 0.005 \text{ s}$ (orange). What are the wavelength, maximum displacement, velocity, and period of the compression wave?

**Solution:**

$$\lambda = 6.00 \text{ m}$$

$$s_{\text{max}} = 2.00 \text{ mm}$$

$$v = 600 \text{ m/s}$$

$$T = 0.01 \text{ s}$$

Exercise:

Problem:

Consider the graph in the preceding problem of a compression wave. Shown are snapshots of the wave function for $t = 0.000$ s (blue) and $t = 0.005$ s (orange). Given that the displacement of the molecule at time $t = 0.00$ s and position $x = 0.00$ m is $s(0.00 \text{ m}, 0.00 \text{ s}) = 1.08 \text{ mm}$, derive a wave function to model the compression wave.

Exercise:**Problem:**

A guitar string oscillates at a frequency of 100 Hz and produces a sound wave. (a) What do you think the frequency of the sound wave is that the vibrating string produces? (b) If the speed of the sound wave is $v = 343 \text{ m/s}$, what is the wavelength of the sound wave?

Solution:

(a) $f = 100 \text{ Hz}$, (b) $\lambda = 3.43 \text{ m}$

Glossary

hearing

perception of sound

sound

traveling pressure wave that may be periodic; the wave can be modeled as a pressure wave or as an oscillation of molecules

Speed of Sound

By the end of this section, you will be able to:

- Explain the relationship between wavelength and frequency of sound
- Determine the speed of sound in different media
- Derive the equation for the speed of sound in air
- Determine the speed of sound in air for a given temperature

Sound, like all waves, travels at a certain speed and has the properties of frequency and wavelength. You can observe direct evidence of the speed of sound while watching a fireworks display ([\[link\]](#)). You see the flash of an explosion well before you hear its sound and possibly feel the pressure wave, implying both that sound travels at a finite speed and that it is much slower than light.



When a firework shell explodes, we perceive the light energy before

the sound energy because sound travels more slowly than light does.

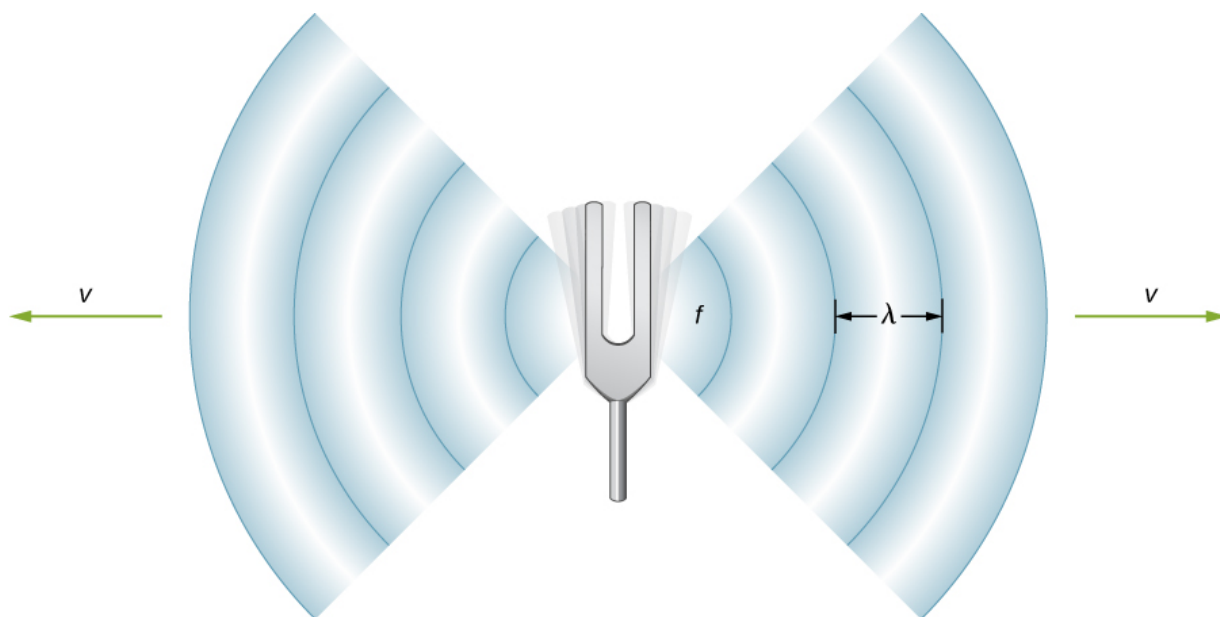
The difference between the speed of light and the speed of sound can also be experienced during an electrical storm. The flash of lighting is often seen before the clap of thunder. You may have heard that if you count the number of seconds between the flash and the sound, you can estimate the distance to the source. Every five seconds converts to about one mile. The velocity of any wave is related to its frequency and wavelength by

Note:

Equation:

$$v = f\lambda,$$

where v is the speed of the wave, f is its frequency, and λ is its wavelength. Recall from [Waves](#) that the wavelength is the length of the wave as measured between sequential identical points. For example, for a surface water wave or sinusoidal wave on a string, the wavelength can be measured between any two convenient sequential points with the same height and slope, such as between two sequential crests or two sequential troughs. Similarly, the wavelength of a sound wave is the distance between sequential identical parts of a wave—for example, between sequential compressions ([link](#)). The frequency is the same as that of the source and is the number of waves that pass a point per unit time.



A sound wave emanates from a source, such as a tuning fork, vibrating at a frequency f . It propagates at speed v and has a wavelength λ .

Speed of Sound in Various Media

[\[link\]](#) shows that the speed of sound varies greatly in different media. The speed of sound in a medium depends on how quickly vibrational energy can be transferred through the medium. For this reason, the derivation of the speed of sound in a medium depends on the medium and on the state of the medium. In general, the equation for the speed of a mechanical wave in a medium depends on the square root of the restoring force, or the elastic property, divided by the inertial property,

Equation:

$$v = \sqrt{\frac{\text{elastic property}}{\text{inertial property}}}.$$

Also, sound waves satisfy the wave equation derived in [Waves](#),

Equation:

$$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y(x, t)}{\partial t^2}.$$

Recall from [Waves](#) that the speed of a wave on a string is equal to $v = \sqrt{\frac{F_T}{\mu}}$, where the restoring force is the tension in the string F_T and the linear density μ is the inertial property. In a fluid, the speed of sound depends on the bulk modulus and the density,

Note:
Equation:

$$v = \sqrt{\frac{B}{\rho}}.$$

The speed of sound in a solid depends on the Young's modulus of the medium and the density,

Note:
Equation:

$$v = \sqrt{\frac{Y}{\rho}}.$$

In an ideal gas (see [The Kinetic Theory of Gases](#)), the equation for the speed of sound is

Note:

Equation:

$$v = \sqrt{\frac{\gamma RT_K}{M}},$$

where γ is the adiabatic index, $R = 8.31 \text{ J/mol} \cdot \text{K}$ is the gas constant, T_K is the absolute temperature in kelvins, and M is the molar mass. In general, the more rigid (or less compressible) the medium, the faster the speed of sound. This observation is analogous to the fact that the frequency of simple harmonic motion is directly proportional to the stiffness of the oscillating object as measured by k , the spring constant. The greater the density of a medium, the slower the speed of sound. This observation is analogous to the fact that the frequency of a simple harmonic motion is inversely proportional to m , the mass of the oscillating object. The speed of sound in air is low, because air is easily compressible. Because liquids and solids are relatively rigid and very difficult to compress, the speed of sound in such media is generally greater than in gases.

Medium	$v \text{ (m/s)}$
<i>Gases at 0° C</i>	
Air	331
Carbon dioxide	259
Oxygen	316

Medium	v (m/s)
Helium	965
Hydrogen	1290
<i>Liquids at 20° C</i>	
Ethanol	1160
Mercury	1450
Water, fresh	1480
Sea Water	1540
Human tissue	1540
<i>Solids (longitudinal or bulk)</i>	
Vulcanized rubber	54
Polyethylene	920
Marble	3810
Glass, Pyrex	5640
Lead	1960
Aluminum	5120
Steel	5960

Speed of Sound in Various Media

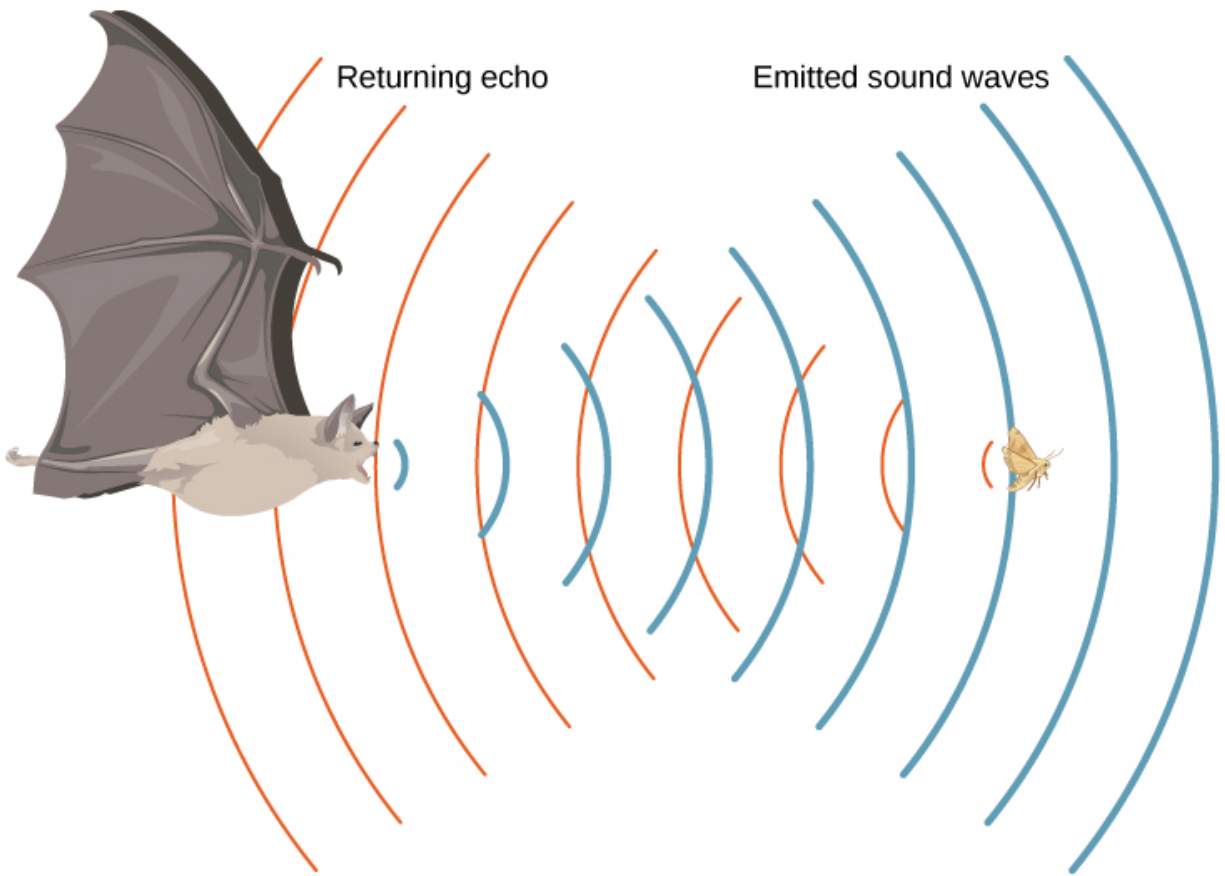
Because the speed of sound depends on the density of the material, and the density depends on the temperature, there is a relationship between the temperature in a given medium and the speed of sound in the medium. For air at sea level, the speed of sound is given by

Note:

Equation:

$$v = 331 \frac{\text{m}}{\text{s}} \sqrt{1 + \frac{T_{\text{C}}}{273^{\circ}\text{C}}} = 331 \frac{\text{m}}{\text{s}} \sqrt{\frac{T_{\text{K}}}{273\text{ K}}}$$

where the temperature in the first equation (denoted as T_{C}) is in degrees Celsius and the temperature in the second equation (denoted as T_{K}) is in kelvins. The speed of sound in gases is related to the average speed of particles in the gas, $v_{\text{rms}} = \sqrt{\frac{3k_{\text{B}}T}{m}}$, where k_{B} is the Boltzmann constant ($1.38 \times 10^{-23} \text{ J/K}$) and m is the mass of each (identical) particle in the gas. Note that v refers to the speed of the coherent propagation of a disturbance (the wave), whereas v_{rms} describes the speeds of particles in random directions. Thus, it is reasonable that the speed of sound in air and other gases should depend on the square root of temperature. While not negligible, this is not a strong dependence. At 0°C , the speed of sound is 331 m/s , whereas at 20.0°C , it is 343 m/s , less than a 4% increase. [\[link\]](#) shows how a bat uses the speed of sound to sense distances.



A bat uses sound echoes to find its way about and to catch prey. The time for the echo to return is directly proportional to the distance.

Derivation of the Speed of Sound in Air

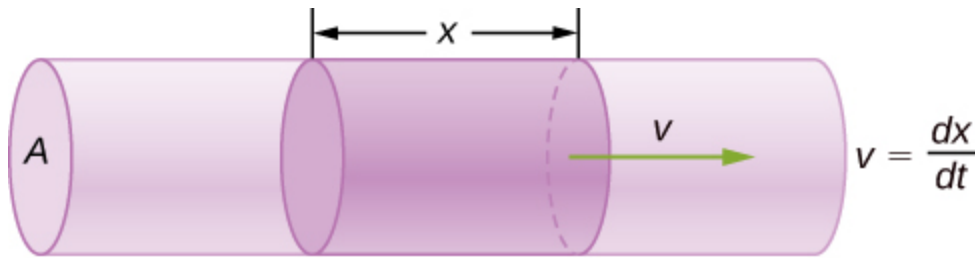
As stated earlier, the speed of sound in a medium depends on the medium and the state of the medium. The derivation of the equation for the speed of sound in air starts with the mass flow rate and continuity equation discussed in [Fluid Mechanics](#).

Consider fluid flow through a pipe with cross-sectional area A ([link](#)). The mass in a small volume of length x of the pipe is equal to the density times the volume, or $m = \rho V = \rho Ax$. The mass flow rate is

Equation:

$$\frac{dm}{dt} = \frac{d}{dt}(\rho V) = \frac{d}{dt}(\rho Ax) = \rho A \frac{dx}{dt} = \rho Av.$$

The continuity equation from [Fluid Mechanics](#) states that the mass flow rate into a volume has to equal the mass flow rate out of the volume, $\rho_{\text{in}} A_{\text{in}} v_{\text{in}} = \rho_{\text{out}} A_{\text{out}} v_{\text{out}}$.

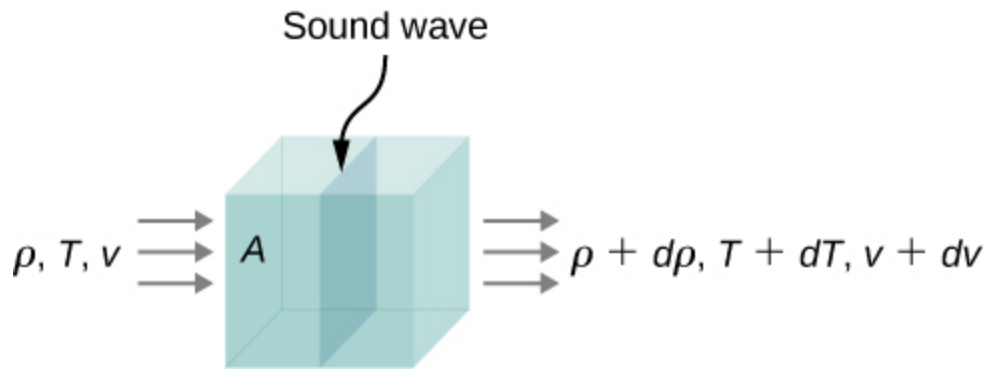


$$m = \rho V = \rho Ax$$

$$\frac{dm}{dt} = \rho A \frac{dx}{dt} = \rho Av$$

The mass of a fluid in a volume is equal to the density times the volume, $m = \rho V = \rho Ax$. The mass flow rate is the time derivative of the mass.

Now consider a sound wave moving through a parcel of air. A parcel of air is a small volume of air with imaginary boundaries ([link](#)). The density, temperature, and velocity on one side of the volume of the fluid are given as ρ, T, v , and on the other side are $\rho + d\rho, T + dT, v + dv$.



A sound wave moves through a volume of fluid. The density, temperature, and velocity of the fluid change from one side to the other.

The continuity equation states that the mass flow rate entering the volume is equal to the mass flow rate leaving the volume, so

Equation:

$$\rho A v = (\rho + d\rho) A (v + dv).$$

This equation can be simplified, noting that the area cancels and considering that the multiplication of two infinitesimals is approximately equal to zero: $d\rho (dv) \approx 0$,

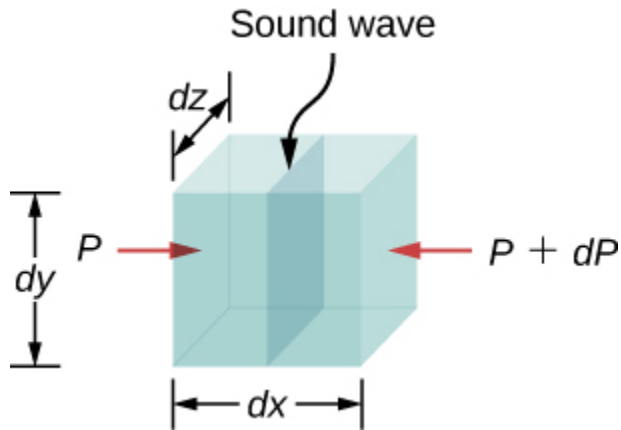
Equation:

$$\begin{aligned} \rho v &= (\rho + d\rho) (v + dv) \\ \rho v &= \rho v + \rho (dv) + (d\rho) v + (d\rho) (dv) \\ 0 &= \rho (dv) + (d\rho) v \\ \rho dv &= -v d\rho. \end{aligned}$$

The net force on the volume of fluid ([link](#)) equals the sum of the forces on the left face and the right face:

Equation:

$$\begin{aligned}
 F_{\text{net}} &= p \, dy \, dz - (p + dp) \, dy \, dz \\
 &= p \, dy \, dz - p \, dy \, dz - dp \, dy \, dz \\
 &= -dp \, dy \, dz \\
 ma &= -dp \, dy \, dz.
 \end{aligned}$$



A sound wave moves through a volume of fluid. The force on each face can be found by the pressure times the area.

The acceleration is the force divided by the mass and the mass is equal to the density times the volume, $m = \rho V = \rho \, dx \, dy \, dz$. We have

Equation:

$$\begin{aligned}
 ma &= -dp \, dy \, dz \\
 a &= -\frac{dp \, dy \, dz}{m} = -\frac{dp \, dy \, dz}{\rho \, dx \, dy \, dz} = -\frac{dp}{(\rho \, dx)} \\
 \frac{dv}{dt} &= -\frac{dp}{(\rho \, dx)} \\
 dv &= -\frac{dp}{(\rho \, dx)} dt = -\frac{dp}{\rho} \frac{1}{v} \\
 \rho v \, dv &= -dp.
 \end{aligned}$$

From the continuity equation $\rho dv = -vd\rho$, we obtain

Equation:

$$\begin{aligned}\rho v dv &= -dp \\ (-vd\rho)v &= -dp \\ v &= \sqrt{\frac{dp}{d\rho}}.\end{aligned}$$

Consider a sound wave moving through air. During the process of compression and expansion of the gas, no heat is added or removed from the system. A process where heat is not added or removed from the system is known as an adiabatic system. Adiabatic processes are covered in detail in [The First Law of Thermodynamics](#), but for now it is sufficient to say that for an adiabatic process, $pV^\gamma = \text{constant}$, where p is the pressure, V is the volume, and gamma (γ) is a constant that depends on the gas. For air, $\gamma = 1.40$. The density equals the number of moles times the molar mass divided by the volume, so the volume is equal to $V = \frac{nM}{\rho}$. The number of moles and the molar mass are constant and can be absorbed into the constant $p\left(\frac{1}{\rho}\right)^\gamma = \text{constant}$. Taking the natural logarithm of both sides yields $\ln p - \gamma \ln \rho = \text{constant}$. Differentiating with respect to the density, the equation becomes

Equation:

$$\begin{aligned}\ln p - \gamma \ln \rho &= \text{constant} \\ \frac{d}{d\rho}(\ln p - \gamma \ln \rho) &= \frac{d}{d\rho}(\text{constant}) \\ \frac{1}{p} \frac{dp}{d\rho} - \frac{\gamma}{\rho} &= 0 \\ \frac{dp}{d\rho} &= \frac{\gamma p}{\rho}.\end{aligned}$$

If the air can be considered an ideal gas, we can use the ideal gas law:

Equation:

$$pV = nRT = \frac{m}{M} RT$$

$$p = \frac{m}{V} \frac{RT}{M} = \rho \frac{RT}{M}.$$

Here M is the molar mass of air:

Equation:

$$\frac{dp}{d\rho} = \frac{\gamma p}{\rho} = \frac{\gamma \left(\rho \frac{RT}{M} \right)}{\rho} = \frac{\gamma RT}{M}.$$

Since the speed of sound is equal to $v = \sqrt{\frac{dp}{d\rho}}$, the speed is equal to

Equation:

$$v = \sqrt{\frac{\gamma RT}{M}}.$$

Note that the velocity is faster at higher temperatures and slower for heavier gases. For air, $\gamma = 1.4$, $M = 0.02897 \frac{\text{kg}}{\text{mol}}$, and $R = 8.31 \frac{\text{J}}{\text{mol} \cdot \text{K}}$. If the temperature is $T_C = 20^\circ \text{C}$ ($T = 293 \text{ K}$), the speed of sound is $v = 343 \text{ m/s}$.

The equation for the speed of sound in air $v = \sqrt{\frac{\gamma RT}{M}}$ can be simplified to give the equation for the speed of sound in air as a function of absolute temperature:

Equation:

$$v = \sqrt{\frac{\gamma RT}{M}}$$

$$= \sqrt{\frac{\gamma RT}{M} \left(\frac{273 \text{ K}}{273 \text{ K}} \right)} = \sqrt{\frac{(273 \text{ K}) \gamma R}{M}} \sqrt{\frac{T}{273 \text{ K}}}$$

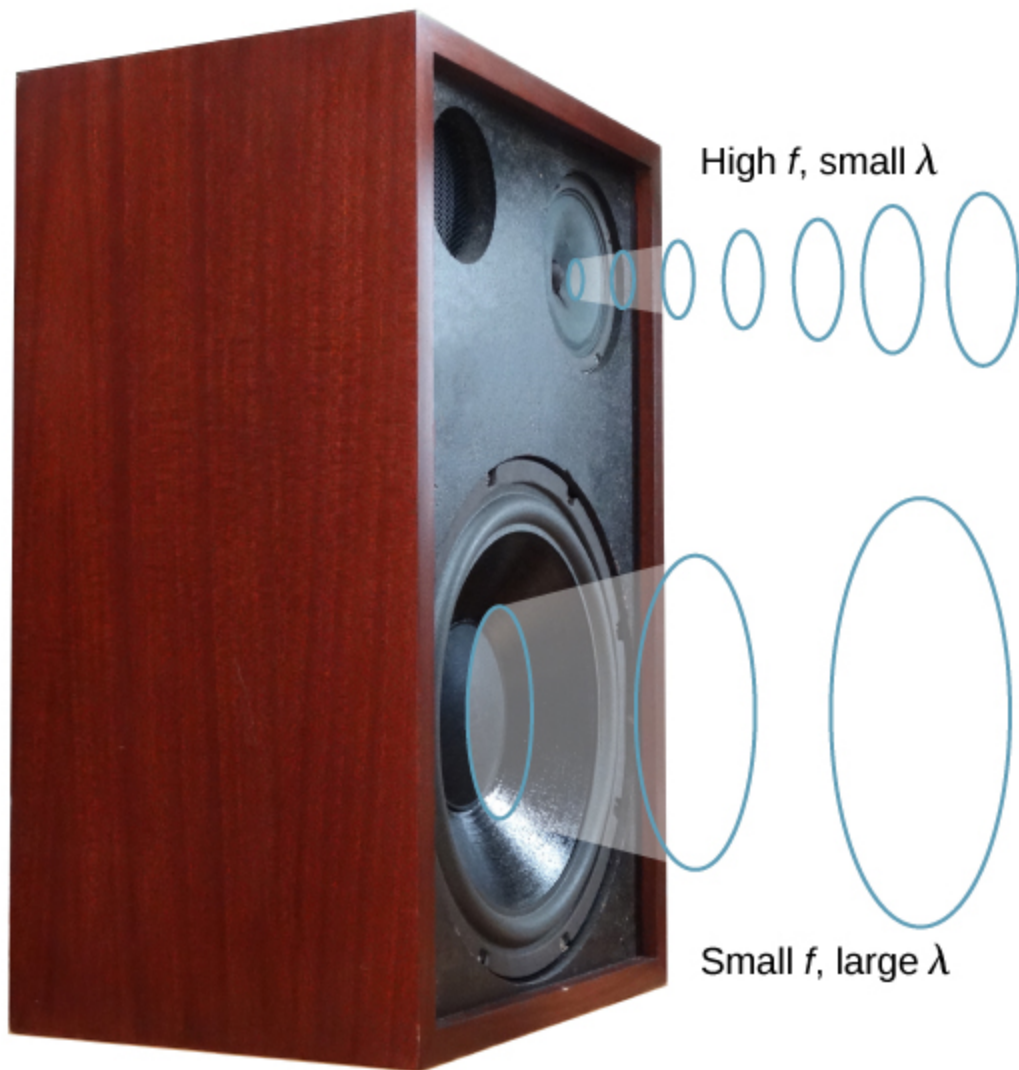
$$\approx 331 \frac{\text{m}}{\text{s}} \sqrt{\frac{T}{273 \text{ K}}}.$$

One of the more important properties of sound is that its speed is nearly independent of the frequency. This independence is certainly true in open air for sounds in the audible range. If this independence were not true, you would certainly notice it for music played by a marching band in a football stadium, for example. Suppose that high-frequency sounds traveled faster—then the farther you were from the band, the more the sound from the low-pitch instruments would lag that from the high-pitch ones. But the music from all instruments arrives in cadence independent of distance, so all frequencies must travel at nearly the same speed. Recall that

Equation:

$$v = f\lambda.$$

In a given medium under fixed conditions, v is constant, so there is a relationship between f and λ ; the higher the frequency, the smaller the wavelength ([\[link\]](#)).



Because they travel at the same speed in a given medium, low-frequency sounds must have a greater wavelength than high-frequency sounds. Here, the lower-frequency sounds are emitted by the large speaker, called a woofer, whereas the higher-frequency sounds are emitted by the small speaker, called a tweeter. (credit: modification of work by Jane Whitney)

Example:**Calculating Wavelengths**

Calculate the wavelengths of sounds at the extremes of the audible range, 20 and 20,000 Hz, in 30.0 °C air. (Assume that the frequency values are accurate to two significant figures.)

Strategy

To find wavelength from frequency, we can use $v = f\lambda$.

Solution

1. Identify knowns. The value for v is given by

Equation:

$$v = (331 \text{ m/s}) \sqrt{\frac{T}{273 \text{ K}}}.$$

2. Convert the temperature into kelvins and then enter the temperature into the equation

Equation:

$$v = (331 \text{ m/s}) \sqrt{\frac{303 \text{ K}}{273 \text{ K}}} = 348.7 \text{ m/s}.$$

3. Solve the relationship between speed and wavelength for λ :

Equation:

$$\lambda = \frac{v}{f}.$$

4. Enter the speed and the minimum frequency to give the maximum wavelength:

Equation:

$$\lambda_{\text{max}} = \frac{348.7 \text{ m/s}}{20 \text{ Hz}} = 17 \text{ m}.$$

5. Enter the speed and the maximum frequency to give the minimum wavelength:

Equation:

$$\lambda_{\min} = \frac{348.7 \text{ m/s}}{20,000 \text{ Hz}} = 0.017 \text{ m} = 1.7 \text{ cm}.$$

Significance

Because the product of f multiplied by λ equals a constant, the smaller f is, the larger λ must be, and vice versa.

The speed of sound can change when sound travels from one medium to another, but the frequency usually remains the same. This is similar to the frequency of a wave on a string being equal to the frequency of the force oscillating the string. If v changes and f remains the same, then the wavelength λ must change. That is, because $v = f\lambda$, the higher the speed of a sound, the greater its wavelength for a given frequency.

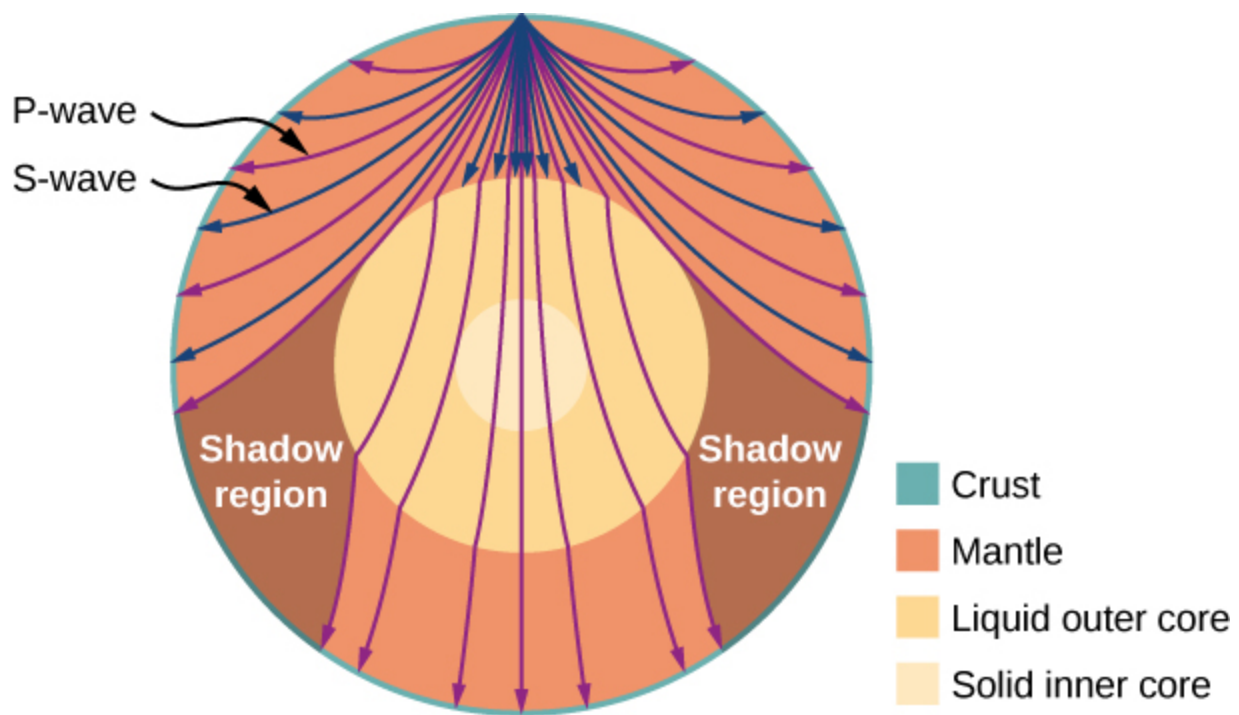
Note:**Exercise:****Problem:**

Check Your Understanding Imagine you observe two firework shells explode. You hear the explosion of one as soon as you see it. However, you see the other shell for several milliseconds before you hear the explosion. Explain why this is so.

Solution:

Sound and light both travel at definite speeds, and the speed of sound is slower than the speed of light. The first shell is probably very close by, so the speed difference is not noticeable. The second shell is farther away, so the light arrives at your eyes noticeably sooner than the sound wave arrives at your ears.

Although sound waves in a fluid are longitudinal, sound waves in a solid travel both as longitudinal waves and transverse waves. Seismic waves, which are essentially sound waves in Earth's crust produced by earthquakes, are an interesting example of how the speed of sound depends on the rigidity of the medium. Earthquakes produce both longitudinal and transverse waves, and these travel at different speeds. The bulk modulus of granite is greater than its shear modulus. For that reason, the speed of longitudinal or pressure waves (P-waves) in earthquakes in granite is significantly higher than the speed of transverse or shear waves (S-waves). Both types of earthquake waves travel slower in less rigid material, such as sediments. P-waves have speeds of 4 to 7 km/s, and S-waves range in speed from 2 to 5 km/s, both being faster in more rigid material. The P-wave gets progressively farther ahead of the S-wave as they travel through Earth's crust. The time between the P- and S-waves is routinely used to determine the distance to their source, the epicenter of the earthquake. Because S-waves do not pass through the liquid core, two shadow regions are produced ([link](#)).



Earthquakes produce both longitudinal waves (P-waves) and

transverse waves (S-waves), and these travel at different speeds. Both waves travel at different speeds in the different regions of Earth, but in general, P-waves travel faster than S-waves. S-waves cannot be supported by the liquid core, producing shadow regions.

As sound waves move away from a speaker, or away from the epicenter of an earthquake, their power per unit area decreases. This is why the sound is very loud near a speaker and becomes less loud as you move away from the speaker. This also explains why there can be an extreme amount of damage at the epicenter of an earthquake but only tremors are felt in areas far from the epicenter. The power per unit area is known as the intensity, and in the next section, we will discuss how the intensity depends on the distance from the source.

Summary

- The speed of sound depends on the medium and the state of the medium.
- In a fluid, because of the absence of shear forces, sound waves are longitudinal. A solid can support both longitudinal and transverse sound waves.
- In air, the speed of sound is related to air temperature T by
$$v = 331 \frac{\text{m}}{\text{s}} \sqrt{\frac{T_{\text{K}}}{273 \text{ K}}} = 331 \frac{\text{m}}{\text{s}} \sqrt{1 + \frac{T_{\text{C}}}{273^{\circ}\text{C}}}.$$
- v is the same for all frequencies and wavelengths of sound in air.

Conceptual Questions

Exercise:

Problem:

How do sound vibrations of atoms differ from thermal motion?

Exercise:

Problem:

When sound passes from one medium to another where its propagation speed is different, does its frequency or wavelength change? Explain your answer briefly.

Solution:

The frequency does not change as the sound wave moves from one medium to another. Since the speed changes and the frequency does not, the wavelength must change. This is similar to the driving force of a harmonic oscillator or a wave on the string.

Exercise:**Problem:**

A popular party trick is to inhale helium and speak in a high-frequency, funny voice. Explain this phenomenon.

Exercise:**Problem:**

You may have used a sonic range finder in lab to measure the distance of an object using a clicking sound from a sound transducer. What is the principle used in this device?

Solution:

The transducer sends out a sound wave, which reflects off the object in question and measures the time it takes for the sound wave to return. Since the speed of sound is constant, the distance to the object can be found by multiplying the velocity of sound by half the time interval measured.

Exercise:

Problem:

The sonic range finder discussed in the preceding question often needs to be calibrated. During the calibration, the software asks for the room temperature. Why do you suppose the room temperature is required?

Problems**Exercise:****Problem:**

When poked by a spear, an operatic soprano lets out a 1200-Hz shriek. What is its wavelength if the speed of sound is 345 m/s?

Exercise:**Problem:**

What frequency sound has a 0.10-m wavelength when the speed of sound is 340 m/s?

Solution:

$$f = 3400 \text{ Hz}$$

Exercise:**Problem:**

Calculate the speed of sound on a day when a 1500-Hz frequency has a wavelength of 0.221 m.

Exercise:**Problem:**

(a) What is the speed of sound in a medium where a 100-kHz frequency produces a 5.96-cm wavelength? (b) Which substance in [\[link\]](#) is this likely to be?

Solution:

a. $v = 5.96 \times 10^3 \text{ m/s}$; b. steel (from value in [\[link\]](#))

Exercise:**Problem:**

Show that the speed of sound in 20.0°C air is 343 m/s , as claimed in the text.

Exercise:**Problem:**

Air temperature in the Sahara Desert can reach 56.0°C (about 134°F). What is the speed of sound in air at that temperature?

Solution:

$$v = 363 \frac{\text{m}}{\text{s}}$$

Exercise:**Problem:**

Dolphins make sounds in air and water. What is the ratio of the wavelength of a sound in air to its wavelength in seawater? Assume air temperature is 20.0°C .

Exercise:**Problem:**

A sonar echo returns to a submarine 1.20 s after being emitted. What is the distance to the object creating the echo? (Assume that the submarine is in the ocean, not in fresh water.)

Solution:

$$\Delta x = 924 \text{ m}$$

Exercise:

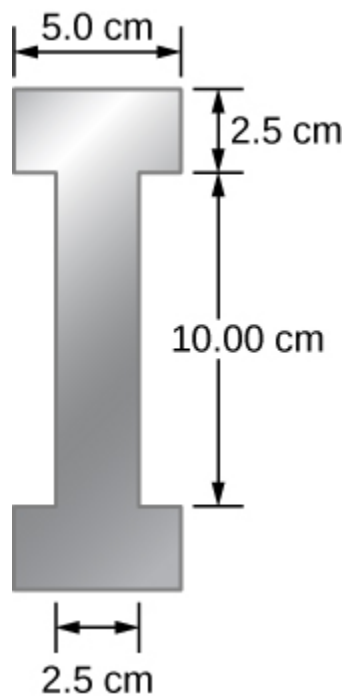
Problem:

(a) If a submarine's sonar can measure echo times with a precision of 0.0100 s, what is the smallest difference in distances it can detect? (Assume that the submarine is in the ocean, not in fresh water.) (b) Discuss the limits this time resolution imposes on the ability of the sonar system to detect the size and shape of the object creating the echo.

Exercise:

Problem:

Ultrasonic sound waves are often used in methods of nondestructive testing. For example, this method can be used to find structural faults in a steel I-beams used in building. Consider a 10.00 meter long, steel I-beam with a cross-section shown below. The weight of the I-beam is 3846.50 N. What would be the speed of sound through in the I-beam? ($Y_{\text{steel}} = 200 \text{ GPa}$, $\beta_{\text{steel}} = 159 \text{ GPa}$).



Solution:

$$V = 0.05 \text{ m}^3$$

$$m = 392.5 \text{ kg}$$

$$\rho = 7850 \text{ kg/m}^3$$

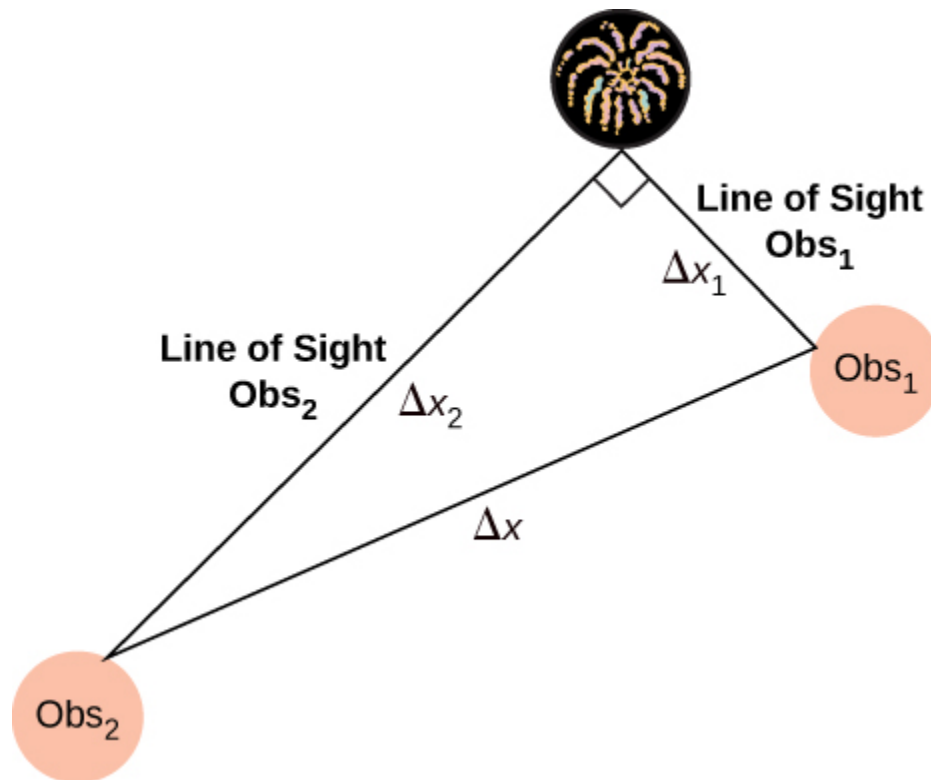
$$v = 5047.54 \text{ m/s}$$

Exercise:**Problem:**

A physicist at a fireworks display times the lag between seeing an explosion and hearing its sound, and finds it to be 0.400 s. (a) How far away is the explosion if air temperature is 24.0°C and if you neglect the time taken for light to reach the physicist? (b) Calculate the distance to the explosion taking the speed of light into account. Note that this distance is negligibly greater.

Exercise:**Problem:**

During a 4th of July celebration, an M80 firework explodes on the ground, producing a bright flash and a loud bang. The air temperature of the night air is $T_F = 90.00^\circ\text{F}$. Two observers see the flash and hear the bang. The first observer notes the time between the flash and the bang as 1.00 second. The second observer notes the difference as 3.00 seconds. The line of sight between the two observers meet at a right angle as shown below. What is the distance Δx between the two observers?



Solution:

$$T_C = 35^\circ \text{C}, v = 351.58 \text{ m/s}$$

$$\Delta x_1 = 35.16 \text{ m}, \Delta x_2 = 52.74 \text{ m}$$

$$\Delta x = 63.39 \text{ m}$$

Exercise:

Problem:

The density of a sample of water is $\rho = 998.00 \text{ kg/m}^3$ and the bulk modulus is $\beta = 2.15 \text{ GPa}$. What is the speed of sound through the sample?

Exercise:

Problem:

Suppose a bat uses sound echoes to locate its insect prey, 3.00 m away. (See [link](#).) (a) Calculate the echo times for temperatures of 5.00°C and 35.0°C . (b) What percent uncertainty does this cause for the bat in locating the insect? (c) Discuss the significance of this uncertainty and whether it could cause difficulties for the bat. (In practice, the bat continues to use sound as it closes in, eliminating most of any difficulties imposed by this and other effects, such as motion of the prey.)

Solution:

a. $t_{5.00^\circ\text{C}} = 0.0180\text{ s}$, $t_{35.0^\circ\text{C}} = 0.0171\text{ s}$; b. % uncertainty = 5.00%; c. This uncertainty could definitely cause difficulties for the bat, if it didn't continue to use sound as it closed in on its prey. A 5% uncertainty could be the difference between catching the prey around the neck or around the chest, which means that it could miss grabbing its prey.

Sound Intensity

By the end of this section, you will be able to:

- Define the term intensity
- Explain the concept of sound intensity level
- Describe how the human ear translates sound

In a quiet forest, you can sometimes hear a single leaf fall to the ground. But when a passing motorist has his stereo turned up, you cannot even hear what the person next to you in your car is saying ([link](#)). We are all very familiar with the loudness of sounds and are aware that loudness is related to how energetically the source is vibrating. High noise exposure is hazardous to hearing, which is why it is important for people working in industrial settings to wear ear protection. The relevant physical quantity is sound intensity, a concept that is valid for all sounds whether or not they are in the audible range.



Noise on crowded roadways, like this one in Delhi, makes it hard to hear others unless they shout. (credit: “Lingaraj G J”/Flickr)

In [Waves](#), we defined intensity as the power per unit area carried by a wave. Power is the rate at which energy is transferred by the wave. In equation form, intensity I is

Equation:

$$I = \frac{P}{A},$$

where P is the power through an area A . The SI unit for I is W/m^2 . If we assume that the sound wave is spherical, and that no energy is lost to thermal processes, the energy of the sound wave is spread over a larger area as distance increases, so the intensity decreases. The area of a sphere is $A = 4\pi r^2$. As the wave spreads out from r_1 to r_2 , the energy also spreads out over a larger area:

Equation:

$$\begin{aligned} P_1 &= P_2 \\ I_1 4\pi r_1^2 &= I_2 4\pi r_2^2; \end{aligned}$$

Note:

Equation:

$$I_2 = I_1 \left(\frac{r_1}{r_2} \right)^2.$$

The intensity decreases as the wave moves out from the source. In an inverse square relationship, such as the intensity, when you double the distance, the intensity decreases to one quarter,

Equation:

$$I_2 = I_1 \left(\frac{r_1}{r_2} \right)^2 = I_1 \left(\frac{r_1}{2r_1} \right)^2 = \frac{1}{4} I_1.$$

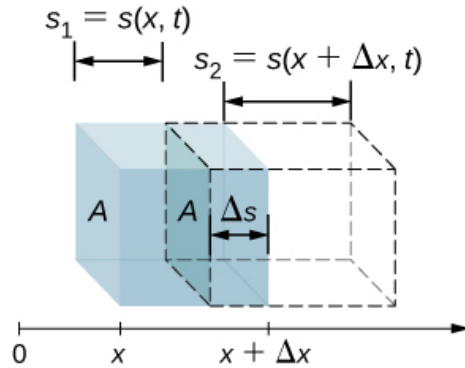
Generally, when considering the intensity of a sound wave, we take the intensity to be the time-averaged value of the power, denoted by $\langle P \rangle$, divided by the area,

Note:

Equation:

$$I = \frac{\langle P \rangle}{A}.$$

The intensity of a sound wave is proportional to the change in the pressure squared and inversely proportional to the density and the speed. Consider a parcel of a medium initially undisturbed and then influenced by a sound wave at time t , as shown in [\[link\]](#).



An undisturbed parcel of a medium with a volume $V = A\Delta x$ shown in blue. A sound wave moves through the medium at time t , and the parcel is displaced and expands, as shown by dotted lines. The change in volume is $\Delta V = A\Delta s = A(s_2 - s_1)$, where s_1 is the displacement of the leading edge of the parcel and s_2 is the displacement of the trailing edge of the parcel. In the figure, $s_2 > s_1$ and the parcel expands, but the parcel can either expand or compress ($s_2 < s_1$), depending on which part of the sound wave (compression or rarefaction) is moving through the parcel.

As the sound wave moves through the parcel, the parcel is displaced and may expand or contract. If $s_2 > s_1$, the volume has increased and the pressure decreases. If $s_2 < s_1$, the volume has decreased and the pressure increases. The change in the volume is

Equation:

$$\Delta V = A\Delta s = A(s_2 - s_1) = A(s(x + \Delta x, t) - s(x, t)).$$

The fractional change in the volume is the change in volume divided by the original volume:

Equation:

$$\frac{dV}{V} = \lim_{\Delta x \rightarrow 0} \frac{A [s(x + \Delta x, t) - s(x, t)]}{A \Delta x} = \frac{\partial s(x, t)}{\partial x}.$$

The fractional change in volume is related to the pressure fluctuation by the bulk modulus $\beta = -\frac{\Delta p(x, t)}{dV/V}$. Recall that the minus sign is required because the volume is *inversely* related to the pressure. (We use lowercase p for pressure to distinguish it from power, denoted by P .) The change in pressure is therefore $\Delta p(x, t) = -\beta \frac{dV}{V} = -\beta \frac{\partial s(x, t)}{\partial x}$. If the sound wave is sinusoidal, then the displacement as shown in [\[link\]](#) is $s(x, t) = s_{\max} \cos(kx \mp \omega t + \phi)$ and the pressure is found to be

Equation:

$$\Delta p(x, t) = -\beta \frac{dV}{V} = -\beta \frac{\partial s(x, t)}{\partial x} = \beta k s_{\max} \sin(kx - \omega t + \phi) = \Delta p_{\max} \sin(kx - \omega t + \phi).$$

The intensity of the sound wave is the power per unit area, and the power is the force times the velocity, $I = \frac{P}{A} = \frac{Fv}{A} = pv$. Here, the velocity is the velocity of the oscillations of the medium, and not the velocity of the sound wave. The velocity of the medium is the time rate of change in the displacement:

Equation:

$$v(x, t) = \frac{\partial}{\partial t} s(x, t) = \frac{\partial}{\partial t} (s_{\max} \cos(kx - \omega t + \phi)) = s_{\max} \omega \sin(kx - \omega t + \phi).$$

Thus, the intensity becomes

Equation:

$$\begin{aligned} I &= \Delta p(x, t) v(x, t) \\ &= \beta k s_{\max} \sin(kx - \omega t + \phi) [s_{\max} \omega \sin(kx - \omega t + \phi)] \\ &= \beta k \omega s_{\max}^2 \sin^2(kx - \omega t + \phi). \end{aligned}$$

To find the time-averaged intensity over one period $T = \frac{2\pi}{\omega}$ for a position x , we integrate over the period, $I = \frac{\beta k \omega s_{\max}^2}{2}$. Using $\Delta p_{\max} = \beta k s_{\max}$, $v = \sqrt{\frac{\beta}{\rho}}$, and $v = \frac{\omega}{k}$, we obtain

Equation:

$$I = \frac{\beta k \omega s_{\max}^2}{2} = \frac{\beta^2 k^2 \omega s_{\max}^2}{2\beta k} = \frac{\omega (\Delta p_{\max})^2}{2(\rho v^2)k} = \frac{v (\Delta p_{\max})^2}{2(\rho v^2)} = \frac{(\Delta p_{\max})^2}{2\rho v}.$$

That is, the intensity of a sound wave is related to its amplitude squared by

Note:

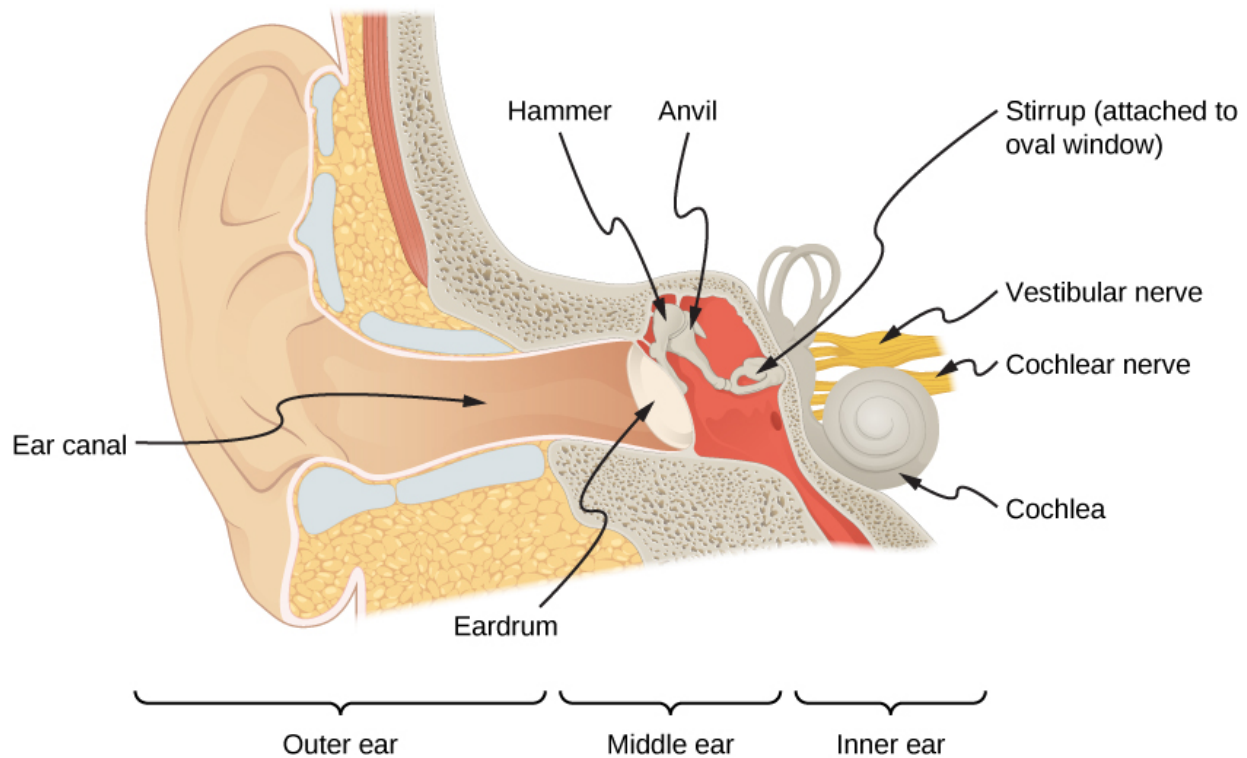
Equation:

$$I = \frac{(\Delta p_{\max})^2}{2\rho v}.$$

Here, Δp_{\max} is the pressure variation or pressure amplitude in units of pascals (Pa) or N/m^2 . The energy (as kinetic energy $\frac{1}{2}mv^2$) of an oscillating element of air due to a traveling sound wave is proportional to its amplitude squared. In this equation, ρ is the density of the material in which the sound wave travels, in units of kg/m^3 , and v is the speed of sound in the medium, in units of m/s. The pressure variation is proportional to the amplitude of the oscillation, so I varies as $(\Delta p)^2$. This relationship is consistent with the fact that the sound wave is produced by some vibration; the greater its pressure amplitude, the more the air is compressed in the sound it creates.

Human Hearing and Sound Intensity Levels

As stated earlier in this chapter, hearing is the perception of sound. The hearing mechanism involves some interesting physics. The sound wave that impinges upon our ear is a pressure wave. The ear is a **transducer** that converts sound waves into electrical nerve impulses in a manner much more sophisticated than, but analogous to, a microphone. [\[link\]](#) shows the anatomy of the ear.



The anatomy of the human ear.

The outer ear, or ear canal, carries sound to the recessed, protected eardrum. The air column in the ear canal resonates and is partially responsible for the sensitivity of the ear to sounds in the 2000–5000-Hz range. The middle ear converts sound into mechanical vibrations and applies these vibrations to the cochlea.

Note:

Watch this [video](#) for a more detailed discussion of the workings of the human ear.

The range of intensities that the human ear can hear depends on the frequency of the sound, but, in general, the range is quite large. The minimum threshold intensity that can be heard is $I_0 = 10^{-12} \text{ W/m}^2$. Pain is experienced at intensities of $I_{\text{pain}} = 1 \text{ W/m}^2$. Measurements of sound intensity (in units of W/m^2) are very cumbersome due to this large range in values. For this reason, as well as for other reasons, the concept of sound intensity level was proposed.

The **sound intensity level** β of a sound, measured in decibels, having an intensity I in watts per meter squared, is defined as

Note:
Equation:

$$\beta(\text{dB}) = 10 \log_{10} \left(\frac{I}{I_0} \right),$$

where $I_0 = 10^{-12} \text{ W/m}^2$ is a reference intensity, corresponding to the threshold intensity of sound that a person with normal hearing can perceive at a frequency of 1.00 kHz. It is more common to consider sound intensity levels in dB than in W/m^2 . How human ears perceive sound can be more accurately described by the logarithm of the intensity rather than directly by the intensity. Because β is defined in terms of a ratio, it is a unitless quantity, telling you the *level* of the sound relative to a fixed standard (10^{-12} W/m^2). The units of decibels (dB) are used to indicate this ratio is multiplied by 10 in its definition. The bel, upon which the decibel is based, is named for Alexander Graham Bell, the inventor of the telephone.

The decibel level of a sound having the threshold intensity of 10^{-12} W/m^2 is $\beta = 0 \text{ dB}$, because $\log_{10} 1 = 0$. [\[link\]](#) gives levels in decibels and intensities in watts per meter squared for some familiar sounds. The ear is sensitive to as little as a trillionth of a watt per meter squared—even more impressive when you realize that the area of the eardrum is only about 1 cm^2 , so that only 10^{-16} W falls on it at the threshold of hearing. Air molecules in a sound wave of this intensity vibrate over a distance of less than one molecular diameter, and the gauge pressures involved are less than 10^{-9} atm .

Sound intensity level β (dB)	Intensity I (W/m^2)	Example/effect
0	1×10^{-12}	Threshold of hearing at 1000 Hz
10	1×10^{-11}	Rustle of leaves
20	1×10^{-10}	Whisper at 1-m distance
30	1×10^{-9}	Quiet home
40	1×10^{-8}	Average home

Sound intensity level β (dB)	Intensity I (W/m^2)	Example/effect
50	1×10^{-7}	Average office, soft music
60	1×10^{-6}	Normal conversation
70	1×10^{-5}	Noisy office, busy traffic
80	1×10^{-4}	Loud radio, classroom lecture
90	1×10^{-3}	Inside a heavy truck; damage from prolonged exposure [footnote] Several government agencies and health-related professional associations recommend that 85 dB not be exceeded for 8-hour daily exposures in the absence of hearing protection.
100	1×10^{-2}	Noisy factory, siren at 30 m; damage from 8 h per day exposure
110	1×10^{-1}	Damage from 30 min per day exposure
120	1	Loud rock concert; pneumatic chipper at 2 m; threshold of pain
140	1×10^2	Jet airplane at 30 m; severe pain, damage in seconds
160	1×10^4	Bursting of eardrums

Sound Intensity Levels and Intensities[\[1\]](#) Several government agencies and health-related professional associations recommend that 85 dB not be exceeded for 8-hour daily exposures in the absence of hearing protection.

An observation readily verified by examining [\[link\]](#) or by using [\[link\]](#) is that each factor of 10 in intensity corresponds to 10 dB. For example, a 90-dB sound compared with a 60-dB sound is 30 dB greater, or three factors of 10 (that is, 10^3 times) as intense. Another example is that if one sound is 10^7 as intense as another, it is 70 dB higher ([\[link\]](#)).

I_2/I_1	$\beta_2 - \beta_1$
2.0	3.0 dB
5.0	7.0 dB
10.0	10.0 dB
100.0	20.0 dB
1000.0	30.0 dB

Ratios of Intensities and Corresponding Differences in Sound Intensity Levels

Example:

Calculating Sound Intensity Levels

Calculate the sound intensity level in decibels for a sound wave traveling in air at 0°C and having a pressure amplitude of 0.656 Pa.

Strategy

We are given Δp , so we can calculate I using the equation $I = \frac{(\Delta p)^2}{2\rho v}$. Using I , we can calculate β straight from its definition in $\beta(\text{dB}) = 10 \log_{10} \left(\frac{I}{I_0} \right)$.

Solution

1. Identify knowns:

Sound travels at 331 m/s in air at 0°C .

Air has a density of 1.29 kg/m^3 at atmospheric pressure and 0°C .

2. Enter these values and the pressure amplitude into $I = \frac{(\Delta p)^2}{2\rho v}$.

Equation:

$$I = \frac{(\Delta p)^2}{2\rho v} = \frac{(0.656 \text{ Pa})^2}{2(1.29 \text{ kg/m}^3)(331 \text{ m/s})} = 5.04 \times 10^{-4} \text{ W/m}^2.$$

3. Enter the value for I and the known value for I_0 into $\beta(\text{dB}) = 10 \log_{10}(I/I_0)$. Calculate to find the sound intensity level in decibels:

Equation:

$$10 \log_{10}(5.04 \times 10^8) = 10(8.70)\text{dB} = 87 \text{ dB}.$$

Significance

This 87-dB sound has an intensity five times as great as an 80-dB sound. So a factor of five in intensity corresponds to a difference of 7 dB in sound intensity level. This value is true for any intensities differing by a factor of five.

Example:**Changing Intensity Levels of a Sound**

Show that if one sound is twice as intense as another, it has a sound level about 3 dB higher.

Strategy

We are given that the ratio of two intensities is 2 to 1, and are then asked to find the difference in their sound levels in decibels. We can solve this problem by using of the properties of logarithms.

Solution

1. Identify knowns:

The ratio of the two intensities is 2 to 1, or

Equation:

$$\frac{I_2}{I_1} = 2.00.$$

We wish to show that the difference in sound levels is about 3 dB. That is, we want to show:

Equation:

$$\beta_2 - \beta_1 = 3 \text{ dB}.$$

Note that

Equation:

$$\log_{10} b - \log_{10} a = \log_{10} \left(\frac{b}{a} \right).$$

2. Use the definition of β to obtain

Equation:

$$\beta_2 - \beta_1 = 10 \log_{10} \left(\frac{I_2}{I_1} \right) = 10 \log_{10} 2.00 = 10(0.301) \text{ dB}.$$

Thus,

Equation:

$$\beta_2 - \beta_1 = 3.01 \text{ dB}.$$

Significance

This means that the two sound intensity levels differ by 3.01 dB, or about 3 dB, as advertised. Note that because only the ratio I_2/I_1 is given (and not the actual intensities), this result is true for any intensities that differ by a factor of two. For example, a 56.0-dB sound is twice as intense as a 53.0-dB sound, a 97.0-dB sound is half as intense as a 100-dB sound, and so on.

Note:**Exercise:****Problem:**

Check Your Understanding Identify common sounds at the levels of 10 dB, 50 dB, and 100 dB.

Solution:

10 dB: rustle of leaves; 50 dB: average office; 100 dB: noisy factory

Another decibel scale is also in use, called the **sound pressure level**, based on the ratio of the pressure amplitude to a reference pressure. This scale is used particularly in applications where sound travels in water. It is beyond the scope of this text to treat this scale because it is not commonly used for sounds in air, but it is important to note that very different decibel levels may be encountered when sound pressure levels are quoted.

Hearing and Pitch

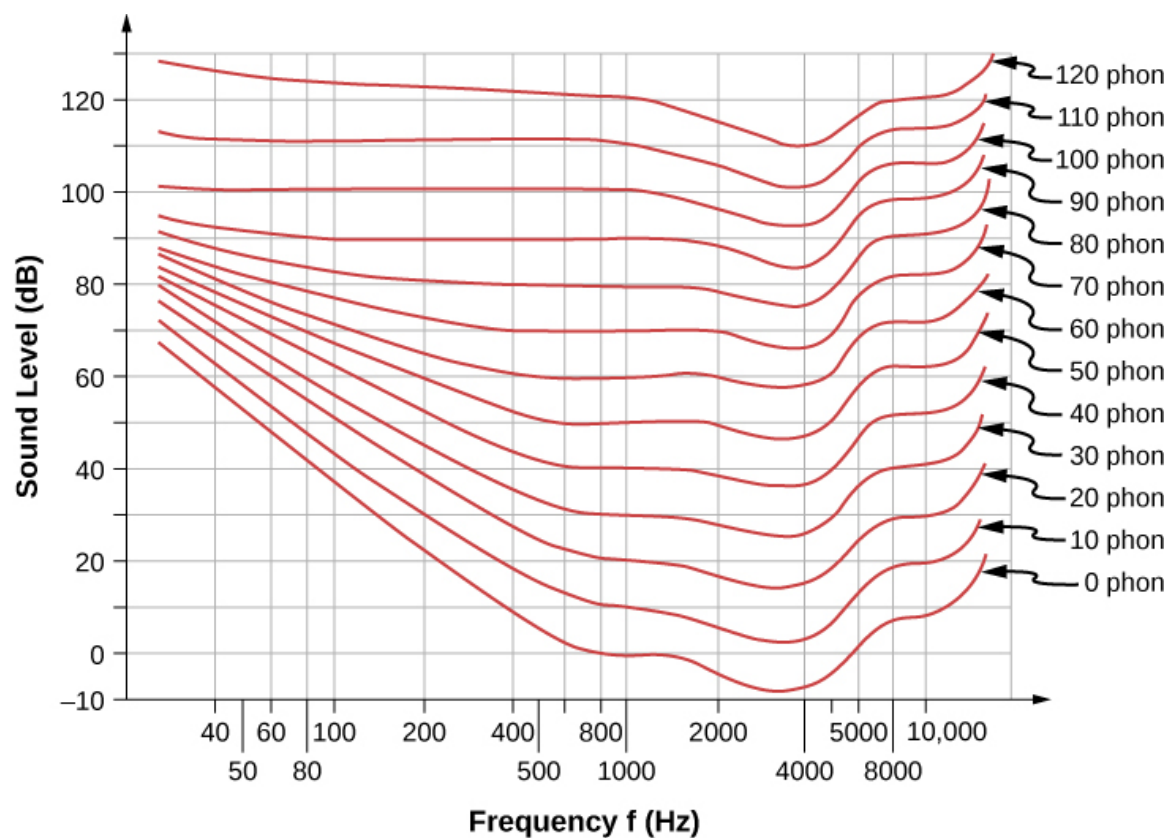
The human ear has a tremendous range and sensitivity. It can give us a wealth of simple information—such as pitch, loudness, and direction.

The perception of frequency is called **pitch**. Typically, humans have excellent relative pitch and can discriminate between two sounds if their frequencies differ by 0.3% or more. For example, 500.0 and 501.5 Hz are noticeably different. Musical **notes** are sounds of a particular frequency that can be produced by most instruments and in Western music have particular names, such as A-sharp, C, or E-flat.

The perception of intensity is called **loudness**. At a given frequency, it is possible to discern differences of about 1 dB, and a change of 3 dB is easily noticed. But loudness is not related to intensity alone. Frequency has a major effect on how loud a sound seems. Sounds near the high- and low-frequency extremes of the hearing range seem even less loud, because the ear is less sensitive at those frequencies. When a violin plays middle C, there is no mistaking it for a piano playing the same note. The reason is that each instrument produces a distinctive set of frequencies and intensities. We call our perception of these combinations of frequencies and intensities tone quality or, more commonly, the **timbre** of the sound. Timbre is the shape of the wave that arises from the many reflections, resonances, and superposition in an instrument.

A unit called a **phon** is used to express loudness numerically. Phons differ from decibels because the phon is a unit of loudness perception, whereas the decibel is a unit of physical intensity. [\[link\]](#) shows the relationship of loudness to intensity (or intensity level) and frequency for persons with normal hearing. The curved lines are equal-loudness curves. Each curve is labeled with its loudness in phons. Any sound along a given curve is perceived as equally loud by the average person. The curves were determined by having large numbers of people compare the

loudness of sounds at different frequencies and sound intensity levels. At a frequency of 1000 Hz, phons are taken to be numerically equal to decibels.



The relationship of loudness in phons to intensity level (in decibels) and intensity (in watts per meter squared) for persons with normal hearing. The curved lines are equal-loudness curves—all sounds on a given curve are perceived as equally loud. Phons and decibels are defined to be the same at 1000 Hz.

Example:

Measuring Loudness

(a) What is the loudness in phons of a 100-Hz sound that has an intensity level of 80 dB? (b) What is the intensity level in decibels of a 4000-Hz sound having a loudness of 70 phons? (c) At what intensity level will an 8000-Hz sound have the same loudness as a 200-Hz sound at 60 dB?

Strategy

The graph in [\[link\]](#) should be referenced to solve this example. To find the loudness of a given sound, you must know its frequency and intensity level, locate that point on the square grid, and

then interpolate between loudness curves to get the loudness in phons. Once that point is located, the intensity level can be determined from the vertical axis.

Solution

1. Identify knowns: The square grid of the graph relating phons and decibels is a plot of intensity level versus frequency—both physical quantities: 100 Hz at 80 dB lies halfway between the curves marked 70 and 80 phons.
Find the loudness: 75 phons.
2. Identify knowns: Values are given to be 4000 Hz at 70 phons.
Follow the 70-phon curve until it reaches 4000 Hz. At that point, it is below the 70 dB line at about 67 dB.
Find the intensity level: 67 dB.
3. Locate the point for a 200 Hz and 60 dB sound.
Find the loudness: This point lies just slightly above the 50-phon curve, and so its loudness is 51 phons.
Look for the 51-phon level is at 8000 Hz: 63 dB.

Significance

These answers, like all information extracted from [\[link\]](#), have uncertainties of several phons or several decibels, partly due to difficulties in interpolation, but mostly related to uncertainties in the equal-loudness curves.

Note:

Exercise:

Problem:

Check Your Understanding Describe how amplitude is related to the loudness of a sound.

Solution:

Amplitude is directly proportional to the experience of loudness. As amplitude increases, loudness increases.

In this section, we discussed the characteristics of sound and how we hear, but how are the sounds we hear produced? Interesting sources of sound are musical instruments and the human voice, and we will discuss these sources. But before we can understand how musical instruments produce sound, we need to look at the basic mechanisms behind these instruments. The theories behind the mechanisms used by musical instruments involve interference, superposition, and standing waves, which we discuss in the next section.

Summary

- Intensity $I = P/A$ is the same for a sound wave as was defined for all waves, where P is the power crossing area A . The SI unit for I is watts per meter squared. The intensity of a sound wave is also related to the pressure amplitude Δp :

Equation:

$$I = \frac{(\Delta p)^2}{2 \rho v},$$

where ρ is the density of the medium in which the sound wave travels and v_w is the speed of sound in the medium.

- Sound intensity level in units of decibels (dB) is

Equation:

$$\beta(\text{dB}) = 10 \log_{10} \left(\frac{I}{I_0} \right),$$

where $I_0 = 10^{-12} \text{ W/m}^2$ is the threshold intensity of hearing.

- The perception of frequency is pitch. The perception of intensity is loudness and loudness has units of phons.

Conceptual Questions

Exercise:

Problem:

Six members of a synchronized swim team wear earplugs to protect themselves against water pressure at depths, but they can still hear the music and perform the combinations in the water perfectly. One day, they were asked to leave the pool so the dive team could practice a few dives, and they tried to practice on a mat, but seemed to have a lot more difficulty. Why might this be?

Solution:

The ear plugs reduce the intensity of the sound both in water and on land, but Navy researchers have found that sound under water is heard through vibrations mastoid, which is the bone behind the ear.

Exercise:

Problem:

A community is concerned about a plan to bring train service to their downtown from the town's outskirts. The current sound intensity level, even though the rail yard is blocks away, is 70 dB downtown. The mayor assures the public that there will be a difference of only 30 dB in sound in the downtown area. Should the townspeople be concerned? Why?

Problems

Exercise:

Problem: What is the intensity in watts per meter squared of a 85.0-dB sound?

Exercise:

Problem:

The warning tag on a lawn mower states that it produces noise at a level of 91.0 dB. What is this in watts per meter squared?

Solution:

$$1.26 \times 10^{-3} \text{ W/m}^2$$

Exercise:

Problem:

A sound wave traveling in air has a pressure amplitude of 0.5 Pa. What is the intensity of the wave?

Exercise:

Problem: What intensity level does the sound in the preceding problem correspond to?

Solution:

85 dB

Exercise:

Problem:

What sound intensity level in dB is produced by earphones that create an intensity of $4.00 \times 10^{-2} \text{ W/m}^2$?

Exercise:

Problem:

What is the decibel level of a sound that is twice as intense as a 90.0-dB sound? (b) What is the decibel level of a sound that is one-fifth as intense as a 90.0-dB sound?

Solution:

a. 93 dB; b. 83 dB

Exercise:

Problem:

What is the intensity of a sound that has a level 7.00 dB lower than a $4.00 \times 10^{-9} \text{ W/m}^2$ sound? (b) What is the intensity of a sound that is 3.00 dB higher than a $4.00 \times 10^{-9} \text{ W/m}^2$ sound?

Exercise:**Problem:**

People with good hearing can perceive sounds as low as -8.00 dB at a frequency of 3000 Hz. What is the intensity of this sound in watts per meter squared?

Solution:

$$1.58 \times 10^{-13} \text{ W/m}^2$$

Exercise:**Problem:**

If a large housefly 3.0 m away from you makes a noise of 40.0 dB, what is the noise level of 1000 flies at that distance, assuming interference has a negligible effect?

Exercise:**Problem:**

Ten cars in a circle at a boom box competition produce a 120-dB sound intensity level at the center of the circle. What is the average sound intensity level produced there by each stereo, assuming interference effects can be neglected?

Solution:

A decrease of a factor of 10 in intensity corresponds to a reduction of 10 dB in sound level:
 $120 \text{ dB} - 10 \text{ dB} = 110 \text{ dB}.$

Exercise:**Problem:**

The amplitude of a sound wave is measured in terms of its maximum gauge pressure. By what factor does the amplitude of a sound wave increase if the sound intensity level goes up by 40.0 dB?

Exercise:**Problem:**

If a sound intensity level of 0 dB at 1000 Hz corresponds to a maximum gauge pressure (sound amplitude) of 10^{-9} atm , what is the maximum gauge pressure in a 60-dB sound? What is the maximum gauge pressure in a 120-dB sound?

Solution:

We know that 60 dB corresponds to a factor of 10^6 increase in intensity. Therefore,

$$I \propto X^2 \Rightarrow \frac{I_2}{I_1} = \left(\frac{X_2}{X_1} \right)^2, \text{ so that } X_2 = 10^{-6} \text{ atm.}$$

$$120 \text{ dB corresponds to a factor of } 10^{12} \text{ increase} \Rightarrow 10^{-9} \text{ atm} (10^{12})^{1/2} = 10^{-3} \text{ atm.}$$

Exercise:**Problem:**

An 8-hour exposure to a sound intensity level of 90.0 dB may cause hearing damage. What energy in joules falls on a 0.800-cm-diameter eardrum so exposed?

Exercise:**Problem:**

Sound is more effectively transmitted into a stethoscope by direct contact rather than through the air, and it is further intensified by being concentrated on the smaller area of the eardrum. It is reasonable to assume that sound is transmitted into a stethoscope 100 times as effectively compared with transmission through the air. What, then, is the gain in decibels produced by a stethoscope that has a sound gathering area of 15.0 cm^2 , and concentrates the sound onto two eardrums with a total area of 0.900 cm^2 with an efficiency of 40.0%?

Solution:

28.2 dB

Exercise:**Problem:**

Loudspeakers can produce intense sounds with surprisingly small energy input in spite of their low efficiencies. Calculate the power input needed to produce a 90.0-dB sound intensity level for a 12.0-cm-diameter speaker that has an efficiency of 1.00%. (This value is the sound intensity level right at the speaker.)

Exercise:**Problem:**

The factor of 10^{-12} in the range of intensities to which the ear can respond, from threshold to that causing damage after brief exposure, is truly remarkable. If you could measure distances over the same range with a single instrument and the smallest distance you could measure was 1 mm, what would the largest be?

Solution:

$$1 \times 10^6 \text{ km}$$

Exercise:

Problem:

What are the closest frequencies to 500 Hz that an average person can clearly distinguish as being different in frequency from 500 Hz? The sounds are not present simultaneously.

Exercise:**Problem:**

Can you tell that your roommate turned up the sound on the TV if its average sound intensity level goes from 70 to 73 dB?

Solution:

$73 \text{ dB} - 70 \text{ dB} = 3 \text{ dB}$; Such a change in sound level is easily noticed.

Exercise:**Problem:**

If a woman needs an amplification of 5.0×10^5 times the threshold intensity to enable her to hear at all frequencies, what is her overall hearing loss in dB? Note that smaller amplification is appropriate for more intense sounds to avoid further damage to her hearing from levels above 90 dB.

Exercise:**Problem:**

A person has a hearing threshold 10 dB above normal at 100 Hz and 50 dB above normal at 4000 Hz. How much more intense must a 100-Hz tone be than a 4000-Hz tone if they are both barely audible to this person?

Solution:

2.5; The 100-Hz tone must be 2.5 times more intense than the 4000-Hz sound to be audible by this person.

Glossary

loudness

perception of sound intensity

notes

basic unit of music with specific names, combined to generate tunes

phon

numerical unit of loudness

pitch

perception of the frequency of a sound

sound intensity level

unitless quantity telling you the level of the sound relative to a fixed standard

sound pressure level

ratio of the pressure amplitude to a reference pressure

timbre

number and relative intensity of multiple sound frequencies

transducer

device that converts energy of a signal into measurable energy form, for example, a microphone converts sound waves into an electrical signal

Normal Modes of a Standing Sound Wave

By the end of this section, you will be able to:

- Explain the mechanism behind sound-reducing headphones
- Describe resonance in a tube closed at one end and open at the other end
- Describe resonance in a tube open at both ends

Interference is the hallmark of waves, all of which exhibit constructive and destructive interference exactly analogous to that seen for water waves. In fact, one way to prove something “is a wave” is to observe interference effects. Since sound is a wave, we expect it to exhibit interference.

Interference of Sound Waves

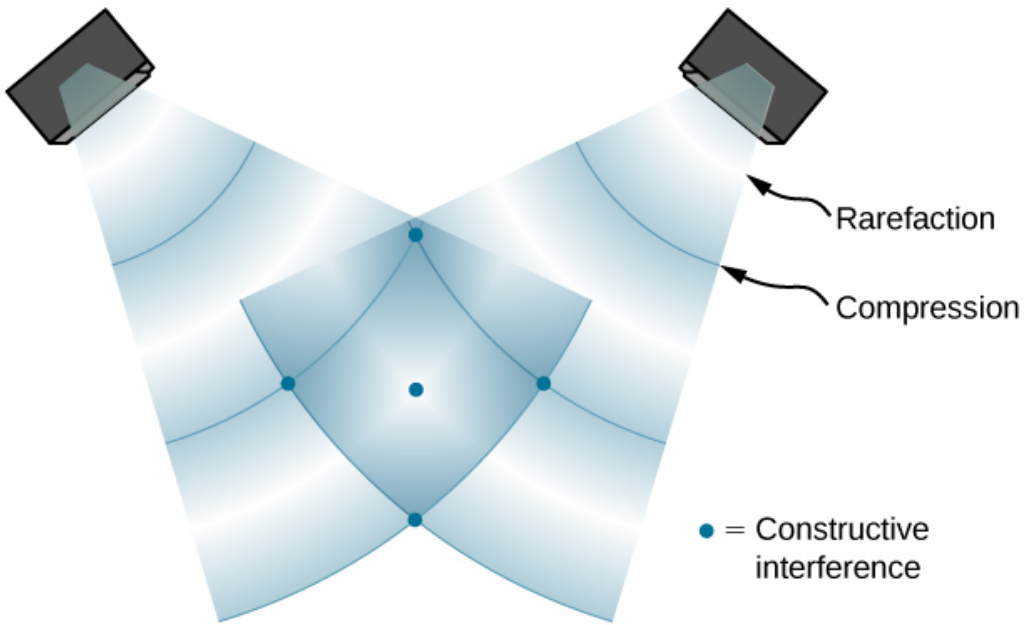
In [Waves](#), we discussed the interference of wave functions that differ only in a phase shift. We found that the wave function resulting from the superposition of $y_1(x, t) = A \sin(kx - \omega t + \phi)$ and $y_2(x, t) = A \sin(kx - \omega t)$ is

Equation:

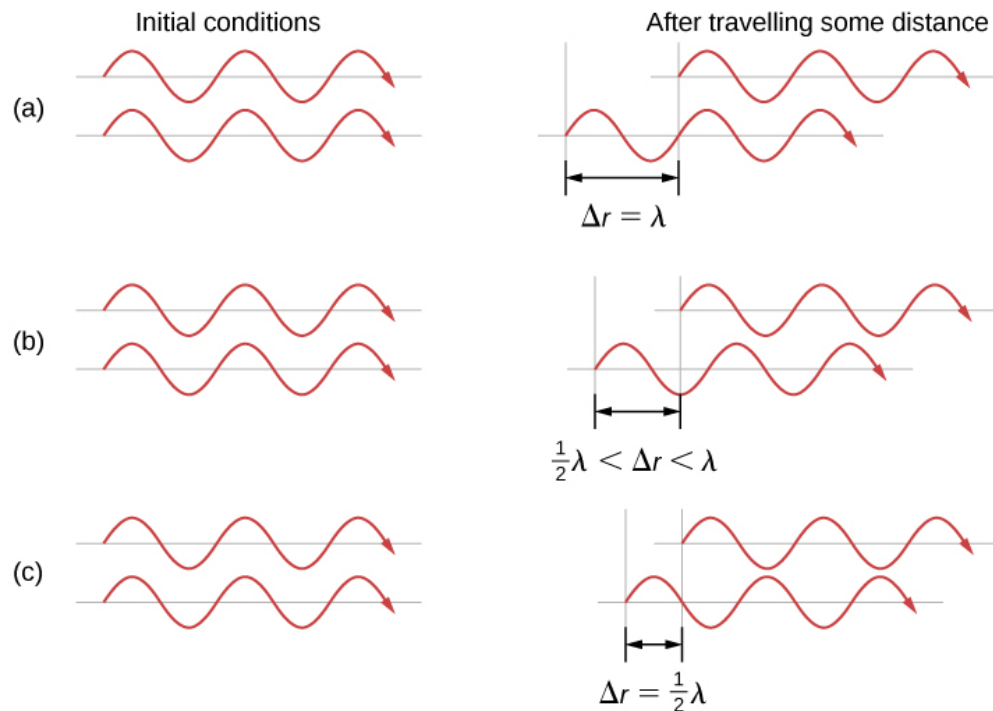
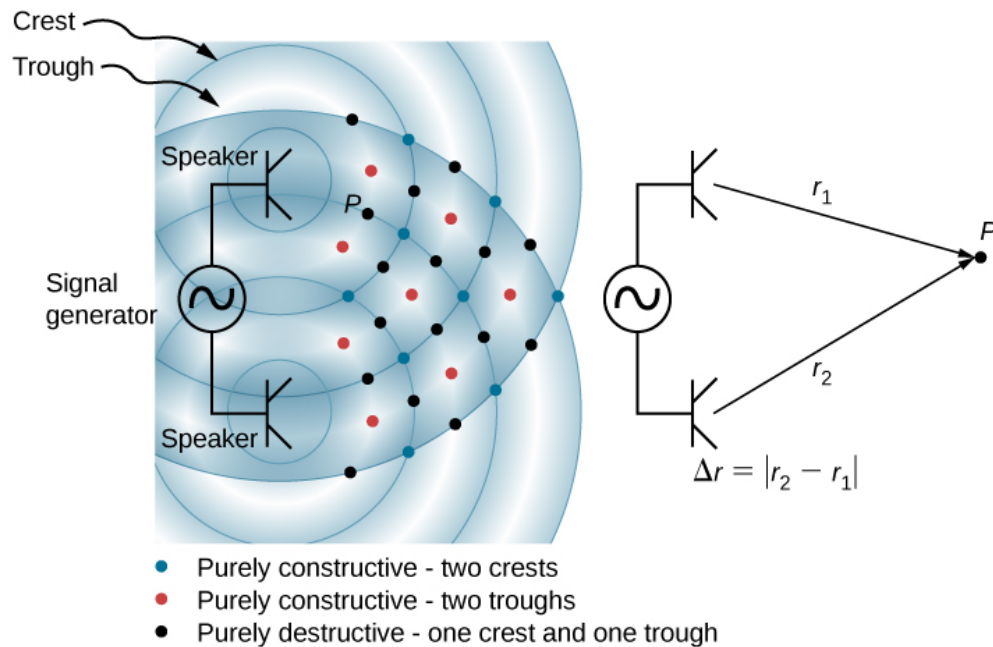
$$y(x, t) = \left[2A \cos\left(\frac{\phi}{2}\right) \right] \sin\left(kx - \omega t + \frac{\phi}{2}\right).$$

One way for two identical waves that are initially in phase to become out of phase with one another is to have the waves travel different distances; that is, they have different path lengths. Sound waves provide an excellent example of a phase shift due to a path difference. As we have discussed, sound waves can basically be modeled as longitudinal waves, where the molecules of the medium oscillate around an equilibrium position, or as pressure waves.

When the waves leave the speakers, they move out as spherical waves ([link](#)). The waves interfere; constructive interference is produced by the combination of two crests or two troughs, as shown. Destructive interference is produced by the combination of a trough and a crest.



When sound waves are produced by a speaker, they travel at the speed of sound and move out as spherical waves. Here, two speakers produce the same steady tone (frequency). The result is points of high-intensity sound (highlighted), which result from two crests (compression) or two troughs (rarefaction) overlapping. Destructive interference results from a crest and trough overlapping. The points where there is constructive interference in the figure occur because the two waves are in phase at those points. Points of destructive interference ([\[link\]](#)) are the result of the two waves being out of phase.



Two speakers being driven by a single signal generator. The sound waves produced by the speakers are in phase and are of a single frequency. The sound waves interfere with each other. When two crests or two troughs coincide, there is constructive interference, marked by the red and blue dots. When a trough and a crest coincide, destructive interference occurs, marked by black dots. The phase

difference is due to the path lengths traveled by the individual waves. Two identical waves travel two different path lengths to a point P . (a)

The difference in the path lengths is one wavelength, resulting in total constructive interference and a resulting amplitude equal to twice the original amplitude. (b) The difference in the path lengths is less than one wavelength but greater than one half a wavelength, resulting in an amplitude greater than zero and less than twice the original amplitude. (c) The difference in the path lengths is one half of a wavelength, resulting in total destructive interference and a resulting amplitude of zero.

The phase difference at each point is due to the different path lengths traveled by each wave. When the difference in the path lengths is an integer multiple of a wavelength,

Equation:

$$\Delta r = |r_2 - r_1| = n\lambda, \text{ where } n = 0, 1, 2, 3, \dots,$$

the waves are in phase and there is constructive interference. When the difference in path lengths is an odd multiple of a half wavelength,

Equation:

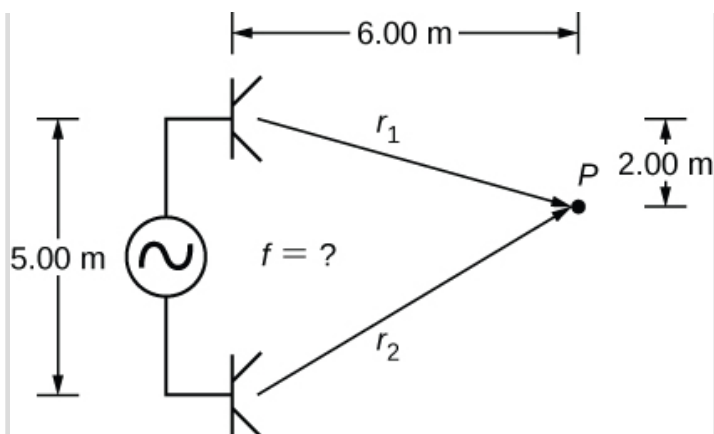
$$\Delta r = |r_2 - r_1| = n\frac{\lambda}{2}, \text{ where } n = 1, 3, 5, \dots,$$

the waves are $180^\circ (\pi \text{ rad})$ out of phase and the result is destructive interference. These points can be located with a sound-level intensity meter.

Example:

Interference of Sound Waves

Two speakers are separated by 5.00 m and are being driven by a signal generator at an unknown frequency. A student with a sound-level meter walks out 6.00 m and down 2.00 m, and finds the first minimum intensity, as shown below. What is the frequency supplied by the signal generator? Assume the wave speed of sound is $v = 343.00 \text{ m/s}$.

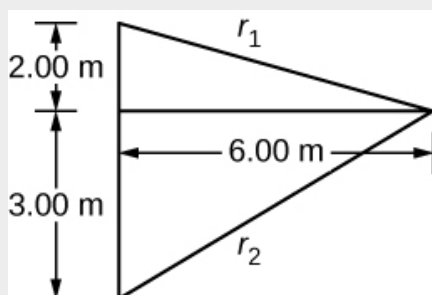


Strategy

The wave velocity is equal to $v = \frac{\lambda}{T} = \lambda f$. The frequency is then $f = \frac{v}{\lambda}$. A minimum intensity indicates destructive interference and the first such point occurs where there is path difference of $\Delta r = \lambda/2$, which can be found from the geometry.

Solution

1. Find the path length to the minimum point from each speaker.



Equation:

$$r_1 = \sqrt{(6.00 \text{ m})^2 + (2.00 \text{ m})^2} = 6.32 \text{ m}, \quad r_2 = \sqrt{(6.00 \text{ m})^2 + (3.00 \text{ m})^2} = 6.71 \text{ m}$$

2. Use the difference in the path length to find the wavelength.

Equation:

$$\Delta r = |r_2 - r_1| = |6.71 \text{ m} - 6.32 \text{ m}| = 0.39 \text{ m}$$

Equation:

$$\lambda = 2\Delta r = 2(0.39 \text{ m}) = 0.78 \text{ m}$$

3. Find the frequency.

Equation:

$$f = \frac{v}{\lambda} = \frac{343.00 \text{ m/s}}{0.78 \text{ m}} = 439.74 \text{ Hz}$$

Significance

If point P were a point of maximum intensity, then the path length would be an integer multiple of the wavelength.

Note:**Exercise:****Problem:**

Check Your Understanding If you walk around two speakers playing music, how come you do not notice places where the music is very loud or very soft, that is, where there is constructive and destructive interference?

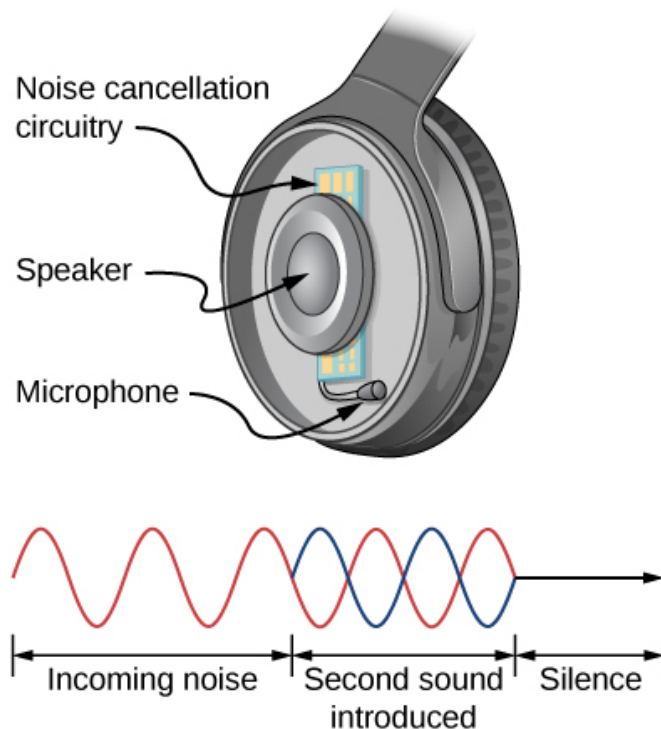
Solution:

In the example, the two speakers were producing sound at a single frequency. Music has various frequencies and wavelengths.

The concept of a phase shift due to a difference in path length is very important. You will use this concept again in [Interference](#) and [Photons and Matter Waves](#), where we discuss how Thomas Young used this method in his famous double-slit experiment to provide evidence that light has wavelike properties.

Noise Reduction through Destructive Interference

[\[link\]](#) shows a clever use of sound interference to cancel noise. Larger-scale applications of active noise reduction by destructive interference have been proposed for entire passenger compartments in commercial aircraft. To obtain destructive interference, a fast electronic analysis is performed, and a second sound is introduced 180° out of phase with the original sound, with its maxima and minima exactly reversed from the incoming noise. Sound waves in fluids are pressure waves and are consistent with Pascal's principle; that is, pressures from two different sources add and subtract like simple numbers. Therefore, positive and negative gauge pressures add to a much smaller pressure, producing a lower-intensity sound. Although completely destructive interference is possible only under the simplest conditions, it is possible to reduce noise levels by 30 dB or more using this technique.



Headphones designed to cancel noise with destructive interference create a sound wave exactly opposite to the incoming sound.

These headphones can be more effective than the simple passive attenuation used in most ear protection. Such headphones were used on the record-setting, around-the-world nonstop flight of the *Voyager* aircraft in 1986 to protect the pilots' hearing from engine noise.

Note:

Exercise:

Problem:

Check Your Understanding Describe how noise-canceling headphones differ from standard headphones used to block outside sounds.

Solution:

Regular headphones only block sound waves with a physical barrier. Noise-canceling headphones use destructive interference to reduce the loudness of outside sounds.

Where else can we observe sound interference? All sound resonances, such as in musical instruments, are due to constructive and destructive interference. Only the resonant frequencies interfere constructively to form standing waves, whereas others interfere destructively and are absent.

Resonance in a Tube Closed at one End

As we discussed in [Waves](#), *standing waves* are formed by two waves moving in opposite directions. When two identical sinusoidal waves move in opposite directions, the waves may be modeled as

Equation:

$$y_1(x, t) = A \sin(kx - \omega t) \text{ and } y_2(x, t) = A \sin(kx + \omega t).$$

When these two waves interfere, the resultant wave is a standing wave:

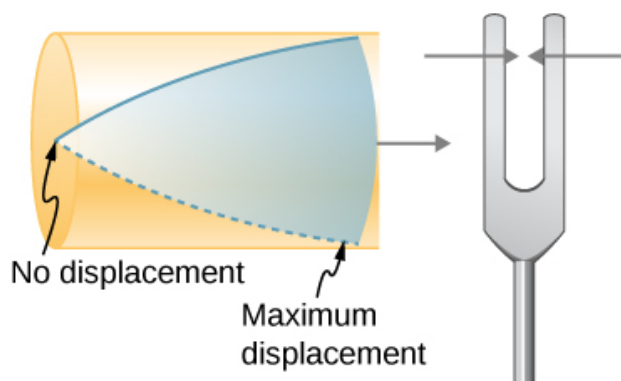
Equation:

$$y_R(x, t) = [2A \sin(kx)] \cos(\omega t).$$

Resonance can be produced due to the boundary conditions imposed on a wave. In [Waves](#), we showed that resonance could be produced in a string under tension that had symmetrical boundary conditions, specifically, a node at each end. We defined a node as a fixed point where the string did not move. We found that the symmetrical boundary conditions resulted in some frequencies resonating and producing standing waves, while other frequencies interfere destructively. Sound waves can resonate in a hollow tube, and the frequencies of the sound waves that resonate depend on the boundary conditions.

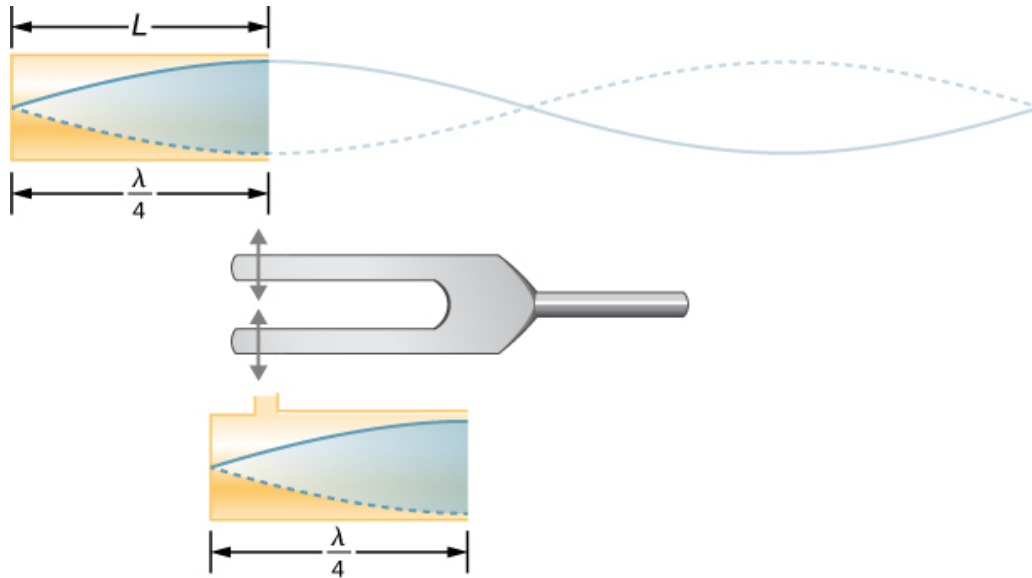
Suppose we have a tube that is closed at one end and open at the other. If we hold a vibrating tuning fork near the open end of the tube, an incident sound wave travels through the tube and reflects off the closed end. The reflected sound has the same frequency and wavelength as the incident sound wave, but is traveling in the opposite direction. At the closed end of the tube, the molecules of air have very little freedom to oscillate, and a node arises. At the open end, the molecules are free to move, and at the right frequency, an antinode occurs. Unlike the symmetrical boundary conditions for the standing waves on the string, the boundary conditions for a tube open at one end and closed at the other end are anti-symmetrical: a node at the closed end and an antinode at the open end.

If the tuning fork has just the right frequency, the air column in the tube resonates loudly, but at most frequencies it vibrates very little. This observation just means that the air column has only certain natural frequencies. Consider the lowest frequency that will cause the tube to resonate, producing a loud sound. There will be a node at the closed end and an antinode at the open end, as shown in [\[link\]](#).



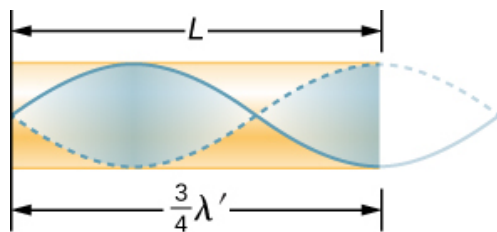
Resonance of air in a tube closed at one end, caused by a tuning fork that vibrates at the lowest frequency that can produce resonance (the fundamental frequency). A node exists at the closed end and an antinode at the open end.

The standing wave formed in the tube has an antinode at the open end and a node at the closed end. The distance from a node to an antinode is one-fourth of a wavelength, and this equals the length of the tube; thus, $\lambda_1 = 4L$. This same resonance can be produced by a vibration introduced at or near the closed end ([\[link\]](#)). It is best to consider this a natural vibration of the air column, independently of how it is induced.



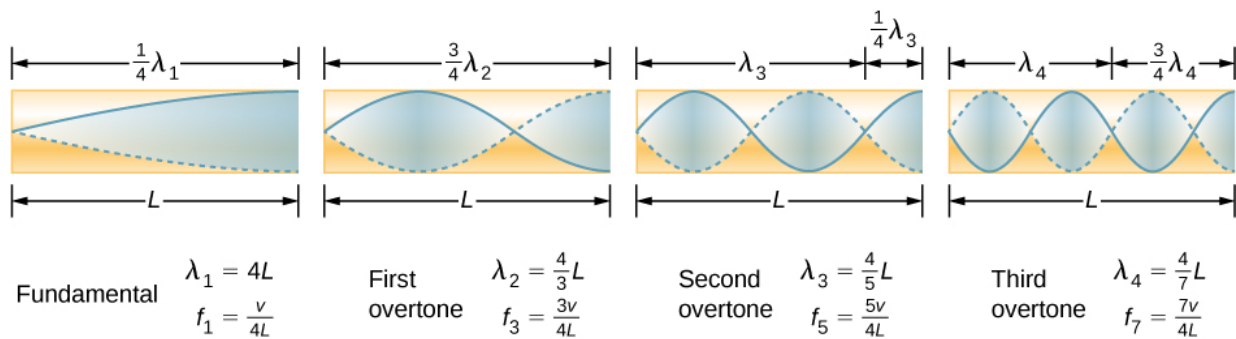
The same standing wave is created in the tube by a vibration introduced near its closed end.

Given that maximum air displacements are possible at the open end and none at the closed end, other shorter wavelengths can resonate in the tube, such as the one shown in [\[link\]](#). Here the standing wave has three-fourths of its wavelength in the tube, or $\frac{3}{4}\lambda_3 = L$, so that $\lambda_3 = \frac{4}{3}L$. Continuing this process reveals a whole series of shorter-wavelength and higher-frequency sounds that resonate in the tube. We use specific terms for the resonances in any system. The lowest resonant frequency is called the **fundamental**, while all higher resonant frequencies are called **overtones**. The resonant frequencies that are integral multiples of the fundamental are collectively called **harmonics**. The fundamental is the first harmonic, the second harmonic is twice the frequency of the first harmonic, and so on. Some of these harmonics may not exist for a given scenario. [\[link\]](#) shows the fundamental and the first three overtones (or the first, third, fifth, and seventh harmonics) in a tube closed at one end.



Another resonance for a tube

closed at one end. This standing wave has maximum air displacement at the open end and none at the closed end. The wavelength is shorter, with three-fourths λ equaling the length of the tube, so that $\lambda = 4L/3$. This higher-frequency vibration is the first overtone.



The fundamental and three lowest overtones for a tube closed at one end. All have maximum air displacements at the open end and none at the closed end.

The relationship for the resonant wavelengths of a tube closed at one end is

Note:

Equation:

$$\lambda_n = \frac{4}{n}L \quad n = 1, 3, 5, \dots$$

Now let us look for a pattern in the resonant frequencies for a simple tube that is closed at one end. The fundamental has $\lambda = 4L$, and frequency is related to wavelength and the speed of sound as given by

$$v = f\lambda.$$

Solving for f in this equation gives

Equation:

$$f = \frac{v}{\lambda} = \frac{v}{4L},$$

where v is the speed of sound in air. Similarly, the first overtone has $\lambda = 4L/3$ (see [\[link\]](#)), so that

Equation:

$$f_3 = 3\frac{v}{4L} = 3f_1.$$

Because $f_3 = 3f_1$, we call the first overtone the third harmonic. Continuing this process, we see a pattern that can be generalized in a single expression. The resonant frequencies of a tube closed at one end are

Note:

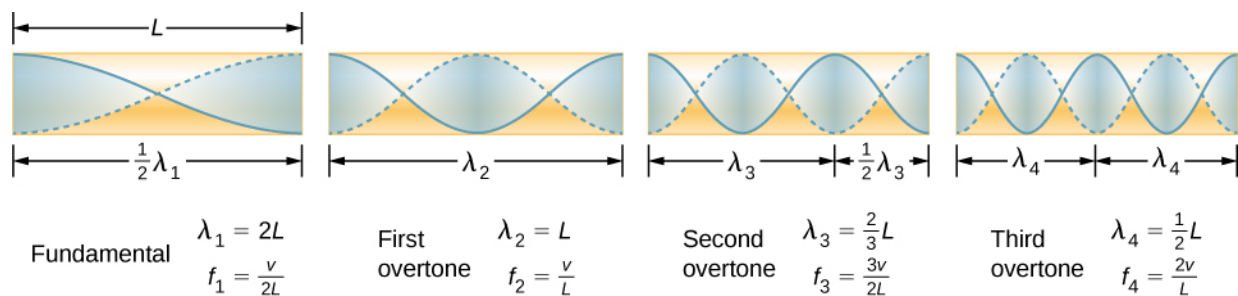
Equation:

$$f_n = n\frac{v}{4L}, \quad n = 1, 3, 5, \dots,$$

where f_1 is the fundamental, f_3 is the first overtone, and so on. It is interesting that the resonant frequencies depend on the speed of sound and, hence, on temperature. This dependence poses a noticeable problem for organs in old unheated cathedrals, and it is also the reason why musicians commonly bring their wind instruments to room temperature before playing them.

Resonance in a Tube Open at Both Ends

Another source of standing waves is a tube that is open at both ends. In this case, the boundary conditions are symmetrical: an antinode at each end. The resonances of tubes open at both ends can be analyzed in a very similar fashion to those for tubes closed at one end. The air columns in tubes open at both ends have maximum air displacements at both ends ([\[link\]](#)). Standing waves form as shown.



The resonant frequencies of a tube open at both ends, including the fundamental and the first three overtones. In all cases, the maximum air displacements occur at both ends of the tube, giving it different natural frequencies than a tube closed at one end.

The relationship for the resonant wavelengths of a tube open at both ends is

Note:

Equation:

$$\lambda_n = \frac{2}{n}L, \quad n = 1, 2, 3, \dots$$

Based on the fact that a tube open at both ends has maximum air displacements at both ends, and using [\[link\]](#) as a guide, we can see that the resonant frequencies of a tube open at both ends are

Note:

Equation:

$$f_n = n \frac{v}{2L}, \quad n = 1, 2, 3, \dots$$

where f_1 is the fundamental, f_2 is the first overtone, f_3 is the second overtone, and so on. Note that a tube open at both ends has a fundamental frequency twice what it would have if closed at one end. It also has a different spectrum of overtones than a tube closed at one end.

Note that a tube open at both ends has symmetrical boundary conditions, similar to the string fixed at both ends discussed in [Waves](#). The relationships for the wavelengths and frequencies of a stringed instrument are the same as given in [\[link\]](#) and [\[link\]](#). The speed of the wave on the string (from [Waves](#)) is $v = \sqrt{\frac{F_T}{\mu}}$. The air around the string vibrates at the same frequency as the string, producing sound of the same frequency. The sound wave moves at the speed of sound and the wavelength can be found using $v = \lambda f$.

Note:

Exercise:

Problem:

Check Your Understanding How is it possible to use a standing wave's node and antinode to determine the length of a closed-end tube?

Solution:

When the tube resonates at its natural frequency, the wave's node is located at the closed end of the tube, and the antinode is located at the open end. The length of the tube is equal to one-fourth of the wavelength of this wave. Thus, if we know the wavelength of the wave, we can determine the length of the tube.

Note:

This [video](#) lets you visualize sound waves.

Note:

Exercise:

Problem:

Check Your Understanding You observe two musical instruments that you cannot identify. One plays high-pitched sounds and the other plays low-pitched sounds. How could you determine which is which without hearing either of them play?

Solution:

Compare their sizes. High-pitch instruments are generally smaller than low-pitch instruments because they generate a smaller wavelength.

Summary

- Unwanted sound can be reduced using destructive interference.
- Sound has the same properties of interference and resonance as defined for all waves.
- In air columns, the lowest-frequency resonance is called the fundamental, whereas all higher resonant frequencies are called overtones. Collectively, they are called harmonics.

Conceptual Questions

Exercise:

Problem:

You are given two wind instruments of identical length. One is open at both ends, whereas the other is closed at one end. Which is able to produce the lowest frequency?

Solution:

The fundamental wavelength of a tube open at each end is $2L$, where the wavelength of a tube open at one end and closed at one end is $4L$. The tube open at one end has the lower fundamental frequency, assuming the speed of sound is the same in both tubes.

Exercise:

Problem:

What is the difference between an overtone and a harmonic? Are all harmonics overtones? Are all overtones harmonics?

Exercise:

Problem:

Two identical columns, open at both ends, are in separate rooms. In room A , the temperature is $T = 20^\circ\text{C}$ and in room B , the temperature is $T = 25^\circ\text{C}$. A speaker is attached to the end of each tube, causing the tubes to resonate at the fundamental frequency. Is the frequency the same for both tubes? Which has the higher frequency?

Solution:

The wavelength in each is twice the length of the tube. The frequency depends on the wavelength and the speed of the sound waves. The frequency in room B is higher because the speed of sound is higher where the temperature is higher.

Problems

Exercise:**Problem:**

(a) What is the fundamental frequency of a 0.672-m-long tube, open at both ends, on a day when the speed of sound is 344 m/s? (b) What is the frequency of its second harmonic?

Exercise:**Problem:**

What is the length of a tube that has a fundamental frequency of 176 Hz and a first overtone of 352 Hz if the speed of sound is 343 m/s?

Solution:

0.974 m

Exercise:**Problem:**

The ear canal resonates like a tube closed at one end. (See [\[link\]](#)Figure 17_03_HumEar[\[link\]](#).) If ear canals range in length from 1.80 to 2.60 cm in an average population, what is the range of fundamental resonant frequencies? Take air temperature to be 37.0°C , which is the same as body temperature.

Exercise:**Problem:**

Calculate the first overtone in an ear canal, which resonates like a 2.40-cm-long tube closed at one end, by taking air temperature to be 37.0°C . Is the ear particularly sensitive to such a frequency? (The resonances of the ear canal are complicated by its nonuniform shape, which we shall ignore.)

Solution:

11.0 kHz; The ear is not particularly sensitive to this frequency, so we don't hear overtones due to the ear canal.

Exercise:

Problem:

A crude approximation of voice production is to consider the breathing passages and mouth to be a resonating tube closed at one end. (a) What is the fundamental frequency if the tube is 0.240 m long, by taking air temperature to be 37.0°C ? (b) What would this frequency become if the person replaced the air with helium? Assume the same temperature dependence for helium as for air.

Exercise:**Problem:**

A 4.0-m-long pipe, open at one end and closed at one end, is in a room where the temperature is $T = 22^\circ\text{C}$. A speaker capable of producing variable frequencies is placed at the open end and is used to cause the tube to resonate. (a) What is the wavelength and the frequency of the fundamental frequency? (b) What is the frequency and wavelength of the first overtone?

Solution:

- a. $v = 344.08\text{ m/s}$, $\lambda_1 = 16.00\text{ m}$, $f_1 = 21.51\text{ Hz}$;
b. $\lambda_3 = 5.33\text{ m}$, $f_3 = 64.56\text{ Hz}$

Exercise:**Problem:**

A 4.0-m-long pipe, open at both ends, is placed in a room where the temperature is $T = 25^\circ\text{C}$. A speaker capable of producing variable frequencies is placed at the open end and is used to cause the tube to resonate. (a) What are the wavelength and the frequency of the fundamental frequency? (b) What are the frequency and wavelength of the first overtone?

Exercise:**Problem:**

A nylon guitar string is fixed between two lab posts 2.00 m apart. The string has a linear mass density of $\mu = 7.20\text{ g/m}$ and is placed under a tension of 160.00 N. The string is placed next to a tube, open at both ends, of length L . The string is plucked and the tube resonates at the $n = 3$ mode. The speed of sound is 343 m/s. What is the length of the tube?

Solution:

$$v_{\text{string}} = 149.07\text{ m/s}, \lambda_3 = 1.33\text{ m}, f_3 = 112.08\text{ Hz}$$
$$\lambda_1 = \frac{v}{f_1}, L = 1.53\text{ m}$$

Exercise:**Problem:**

A 512-Hz tuning fork is struck and placed next to a tube with a movable piston, creating a tube with a variable length. The piston is slid down the pipe and resonance is reached when the piston is 115.50 cm from the open end. The next resonance is reached when the piston is 82.50 cm from the open end. (a) What is the speed of sound in the tube? (b) How far from the open end will the piston cause the next mode of resonance?

Exercise:**Problem:**

Students in a physics lab are asked to find the length of an air column in a tube closed at one end that has a fundamental frequency of 256 Hz. They hold the tube vertically and fill it with water to the top, then lower the water while a 256-Hz tuning fork is rung and listen for the first resonance. (a) What is the air temperature if the resonance occurs for a length of 0.336 m? (b) At what length will they observe the second resonance (first overtone)?

Solution:

a. 22.0°C; b. 1.01 m

Glossary

fundamental

the lowest-frequency resonance

harmonics

the term used to refer collectively to the fundamental and its overtones

overtones

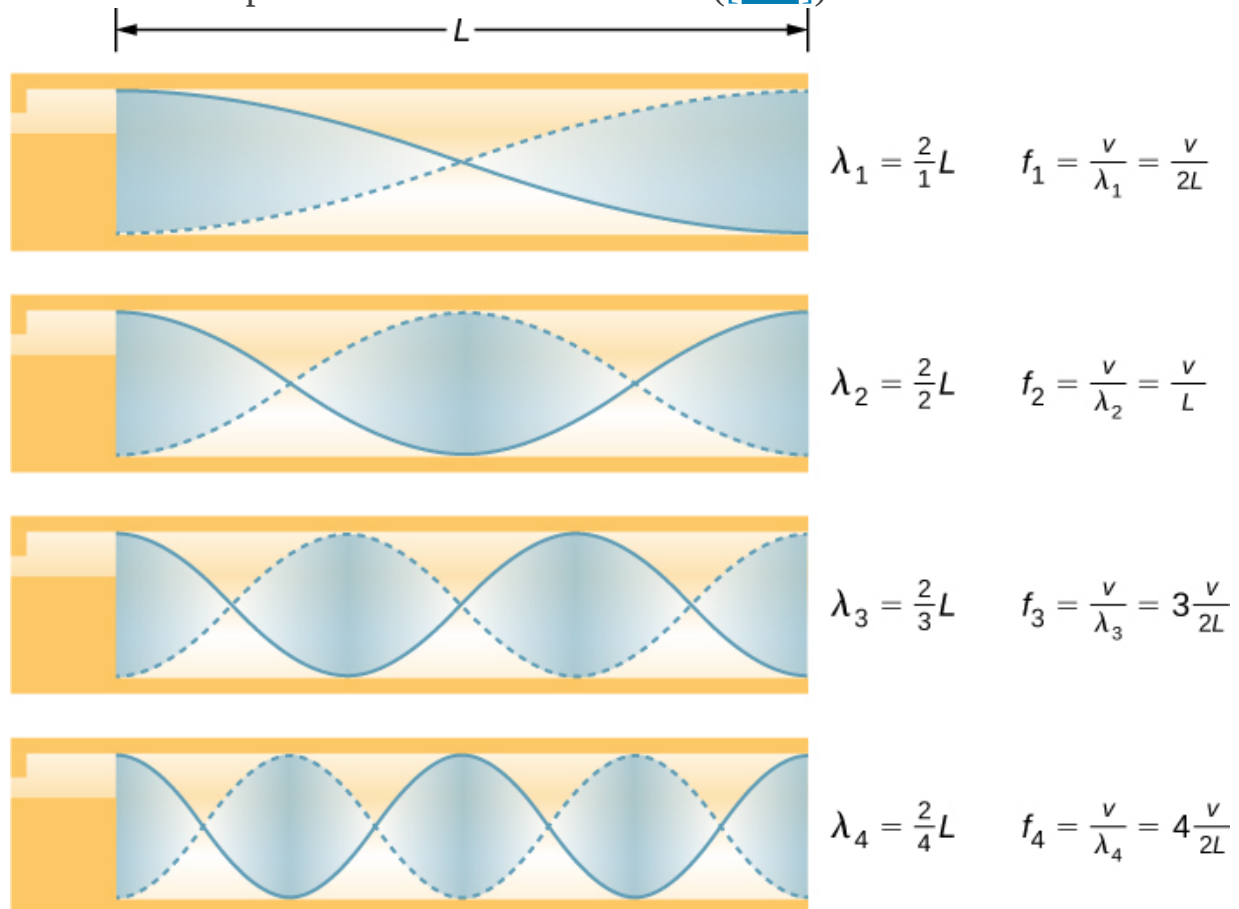
all resonant frequencies higher than the fundamental

Sources of Musical Sound

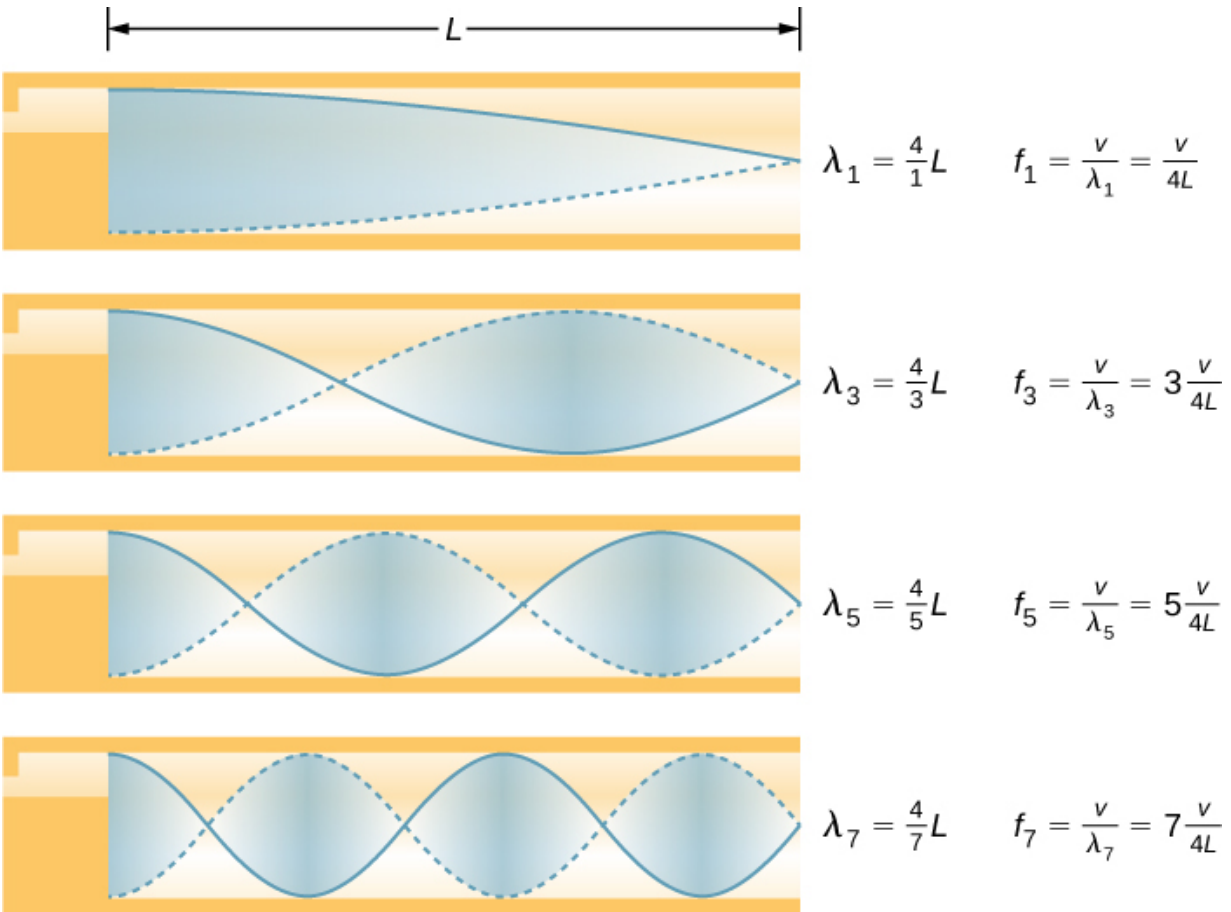
By the end of this section, you will be able to:

- Describe the resonant frequencies in instruments that can be modeled as a tube with symmetrical boundary conditions
- Describe the resonant frequencies in instruments that can be modeled as a tube with anti-symmetrical boundary conditions

Some musical instruments, such as woodwinds, brass, and pipe organs, can be modeled as tubes with symmetrical boundary conditions, that is, either open at both ends or closed at both ends ([\[link\]](#)). Other instruments can be modeled as tubes with anti-symmetrical boundary conditions, such as a tube with one end open and the other end closed ([\[link\]](#)).



Some musical instruments can be modeled as a pipe open at both ends.



Some musical instruments can be modeled as a pipe closed at one end.

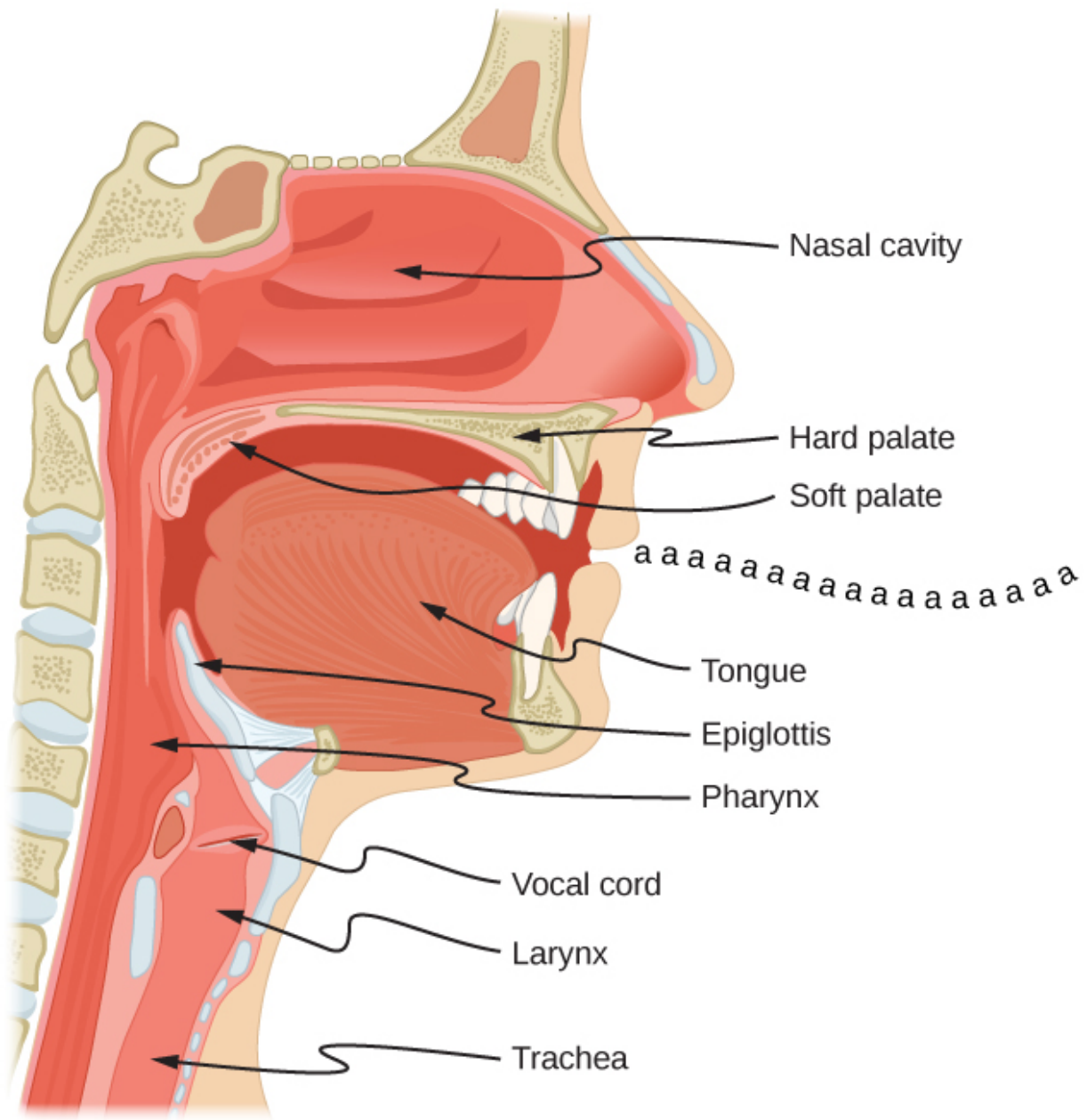
Resonant frequencies are produced by longitudinal waves that travel down the tubes and interfere with the reflected waves traveling in the opposite direction. A pipe organ is manufactured with various tubes of fixed lengths to produce different frequencies. The waves are the result of compressed air allowed to expand in the tubes. Even in open tubes, some reflection occurs due to the constraints of the sides of the tubes and the atmospheric pressure outside the open tube.

The antinodes do not occur at the opening of the tube, but rather depend on the radius of the tube. The waves do not fully expand until they are outside the open end of a tube, and for a thin-walled tube, an *end correction* should

be added. This end correction is approximately 0.6 times the radius of the tube and should be added to the length of the tube.

Players of instruments such as the flute or oboe vary the length of the tube by opening and closing finger holes. On a trombone, you change the tube length by using a sliding tube. Bugles have a fixed length and can produce only a limited range of frequencies.

The fundamental and overtones can be present simultaneously in a variety of combinations. For example, middle C on a trumpet sounds distinctively different from middle C on a clarinet, although both instruments are modified versions of a tube closed at one end. The fundamental frequency is the same (and usually the most intense), but the overtones and their mix of intensities are different and subject to shading by the musician. This mix is what gives various musical instruments (and human voices) their distinctive characteristics, whether they have air columns, strings, sounding boxes, or drumheads. In fact, much of our speech is determined by shaping the cavity formed by the throat and mouth, and positioning the tongue to adjust the fundamental and combination of overtones. For example, simple resonant cavities can be made to resonate with the sound of the vowels ([link](#)). In boys at puberty, the larynx grows and the shape of the resonant cavity changes, giving rise to the difference in predominant frequencies in speech between men and women.



The throat and mouth form an air column closed at one end that resonates in response to vibrations in the voice box. The spectrum of overtones and their intensities vary with mouth shaping and tongue position to form different sounds. The voice box can be replaced with a mechanical vibrator, and understandable speech is still possible. Variations in basic shapes make different voices recognizable.

Example:**Finding the Length of a Tube with a 128-Hz Fundamental**

(a) What length should a tube closed at one end have on a day when the air temperature is

22.0 °C if its fundamental frequency is to be 128 Hz (C below middle C)?

(b) What is the frequency of its fourth overtone?

Strategy

The length L can be found from the relationship $f_n = n \frac{v}{4L}$, but we first need to find the speed of sound v .

Solution

- a. Identify knowns: The fundamental frequency is 128 Hz, and the air temperature is 22.0 °C.

Use $f_n = n \frac{v}{4L}$ to find the fundamental frequency ($n = 1$),

Equation:

$$f_1 = \frac{v}{4L}.$$

Solve this equation for length,

Equation:

$$L = \frac{v}{4f_1}.$$

Find the speed of sound using $v = (331 \text{ m/s}) \sqrt{\frac{T}{273 \text{ K}}}$,

Equation:

$$v = (331 \text{ m/s}) \sqrt{\frac{295 \text{ K}}{273 \text{ K}}} = 344 \text{ m/s}.$$

Enter the values of the speed of sound and frequency into the expression for L .

Equation:

$$L = \frac{v}{4f_1} = \frac{344 \text{ m/s}}{4(128 \text{ Hz})} = 0.672 \text{ m}$$

b. Identify knowns: The first overtone has $n = 3$, the second overtone has $n = 5$, the third overtone has $n = 7$, and the fourth overtone has $n = 9$.

Enter the value for the fourth overtone into $f_n = n \frac{v}{4L}$,

Equation:

$$f_9 = 9 \frac{v}{4L} = 9f_1 = 1.15 \text{ kHz}.$$

Significance

Many wind instruments are modified tubes that have finger holes, valves, and other devices for changing the length of the resonating air column and hence, the frequency of the note played. Horns producing very low frequencies require tubes so long that they are coiled into loops. An example is the tuba. Whether an overtone occurs in a simple tube or a musical instrument depends on how it is stimulated to vibrate and the details of its shape. The trombone, for example, does not produce its fundamental frequency and only makes overtones.

If you have two tubes with the same fundamental frequency, but one is open at both ends and the other is closed at one end, they would sound different when played because they have different overtones. Middle C, for example, would sound richer played on an open tube, because it has even multiples of the fundamental as well as odd. A closed tube has only odd multiples.

Resonance

Resonance occurs in many different systems, including strings, air columns, and atoms. As we discussed in earlier chapters, resonance is the driven or forced oscillation of a system at its natural frequency. At resonance, energy is transferred rapidly to the oscillating system, and the amplitude of its oscillations grows until the system can no longer be described by Hooke's law. An example of this is the distorted sound intentionally produced in certain types of rock music.

Wind instruments use resonance in air columns to amplify tones made by lips or vibrating reeds. Other instruments also use air resonance in clever ways to amplify sound. [\[link\]](#) shows a violin and a guitar, both of which have sounding boxes but with different shapes, resulting in different overtone structures. The vibrating string creates a sound that resonates in the sounding box, greatly amplifying the sound and creating overtones that give the instrument its characteristic timbre. The more complex the shape of the sounding box, the greater its ability to resonate over a wide range of frequencies. The marimba, like the one shown in [\[link\]](#), uses pots or gourds below the wooden slats to amplify their tones. The resonance of the pot can be adjusted by adding water.



(a)



(b)

String instruments such as (a) violins and (b) guitars use resonance in their sounding boxes to amplify and enrich the sound created by their vibrating strings. The bridge and supports couple the string vibrations to the sounding boxes and air within. (credit a: modification of work by Feliciano Guimarães; credit b: modification of work by Steve Snodgrass)



This marimba uses gourds as resonance chambers to amplify its sound.
(credit: “APC Events”/Flickr)

We have emphasized sound applications in our discussions of resonance and standing waves, but these ideas apply to any system that has wave characteristics. Vibrating strings, for example, are actually resonating and have fundamentals and overtones similar to those for air columns. More subtle are the resonances in atoms due to the wave character of their electrons. Their orbitals can be viewed as standing waves, which have a fundamental (ground state) and overtones (excited states). It is fascinating that wave characteristics apply to such a wide range of physical systems.

Summary

- Some musical instruments can be modeled as pipes that have symmetrical boundary conditions: open at both ends or closed at both ends. Other musical instruments can be modeled as pipes that have anti-symmetrical boundary conditions: closed at one end and open at the other.
- Some instruments, such as the pipe organ, have several tubes with different lengths. Instruments such as the flute vary the length of the tube by closing the holes along the tube. The trombone varies the length of the tube using a sliding bar.
- String instruments produce sound using a vibrating string with nodes at each end. The air around the string oscillates at the frequency of the string. The relationship for the frequencies for the string is the same as for the symmetrical boundary conditions of the pipe, with the length of the pipe replaced by the length of the string and the velocity replaced by $v = \sqrt{\frac{F_T}{\mu}}$.

Conceptual Questions

Exercise:

Problem:

How does an unamplified guitar produce sounds so much more intense than those of a plucked string held taut by a simple stick?

Exercise:

Problem:

Consider three pipes of the same length (L). Pipe A is open at both ends, pipe B is closed at both ends, and pipe C has one open end and one closed end. If the velocity of sound is the same in each of the three tubes, in which of the tubes could the lowest fundamental frequency be produced? In which of the tubes could the highest fundamental frequency be produced?

Solution:

When resonating at the fundamental frequency, the wavelength for pipe C is $4L$, and for pipes A and B is $2L$. The frequency is equal to $f = v/\lambda$. Pipe C has the lowest frequency and pipes A and B have equal frequencies, higher than the one in pipe C .

Exercise:**Problem:**

Pipe A has a length L and is open at both ends. Pipe B has a length $L/2$ and has one open end and one closed end. Assume the speed of sound to be the same in both tubes. Which of the harmonics in each tube would be equal?

Exercise:**Problem:**

A string is tied between two lab posts a distance L apart. The tension in the string and the linear mass density is such that the speed of a wave on the string is $v = 343 \text{ m/s}$. A tube with symmetric boundary conditions has a length L and the speed of sound in the tube is $v = 343 \text{ m/s}$. What could be said about the frequencies of the harmonics in the string and the tube? What if the velocity in the string were $v = 686 \text{ m/s}$?

Solution:

Since the boundary conditions are both symmetric, the frequencies are $f_n = \frac{nv}{2L}$. Since the speed is the same in each, the frequencies are the same. If the wave speed were doubled in the string, the frequencies in the string would be twice the frequencies in the tube.

Problems**Exercise:**

Problem:

If a wind instrument, such as a tuba, has a fundamental frequency of 32.0 Hz, what are its first three overtones? It is closed at one end. (The overtones of a real tuba are more complex than this example, because it is a tapered tube.)

Exercise:**Problem:**

What are the first three overtones of a bassoon that has a fundamental frequency of 90.0 Hz? It is open at both ends. (The overtones of a real bassoon are more complex than this example, because its double reed makes it act more like a tube closed at one end.)

Solution:

first overtone = 180 Hz;

second overtone = 270 Hz;

third overtone = 360 Hz

Exercise:**Problem:**

How long must a flute be in order to have a fundamental frequency of 262 Hz (this frequency corresponds to middle C on the evenly tempered chromatic scale) on a day when air temperature is 20.0 °C? It is open at both ends.

Exercise:**Problem:**

What length should an oboe have to produce a fundamental frequency of 110 Hz on a day when the speed of sound is 343 m/s? It is open at both ends.

Solution:

1.56 m

Exercise:**Problem:**

(a) Find the length of an organ pipe closed at one end that produces a fundamental frequency of 256 Hz when air temperature is 18.0°C . (b) What is its fundamental frequency at 25.0°C ?

Exercise:**Problem:**

An organ pipe ($L = 3.00\text{ m}$) is closed at both ends. Compute the wavelengths and frequencies of the first three modes of resonance. Assume the speed of sound is $v = 343.00\text{ m/s}$.

Solution:

The pipe has symmetrical boundary conditions;

$$\lambda_n = \frac{2}{n}L, \quad f_n = \frac{nv}{2L}, \quad n = 1, 2, 3$$

$$\lambda_1 = 6.00\text{ m}, \quad \lambda_2 = 3.00\text{ m}, \quad \lambda_3 = 2.00\text{ m}$$

$$f_1 = 57.17\text{ Hz}, \quad f_2 = 114.33\text{ Hz}, \quad f_3 = 171.50\text{ Hz}$$

Exercise:**Problem:**

An organ pipe ($L = 3.00\text{ m}$) is closed at one end. Compute the wavelengths and frequencies of the first three modes of resonance. Assume the speed of sound is $v = 343.00\text{ m/s}$.

Exercise:

Problem:

A sound wave of a frequency of 2.00 kHz is produced by a string oscillating in the $n = 6$ mode. The linear mass density of the string is $\mu = 0.0065 \text{ kg/m}$ and the length of the string is 1.50 m. What is the tension in the string?

Solution:

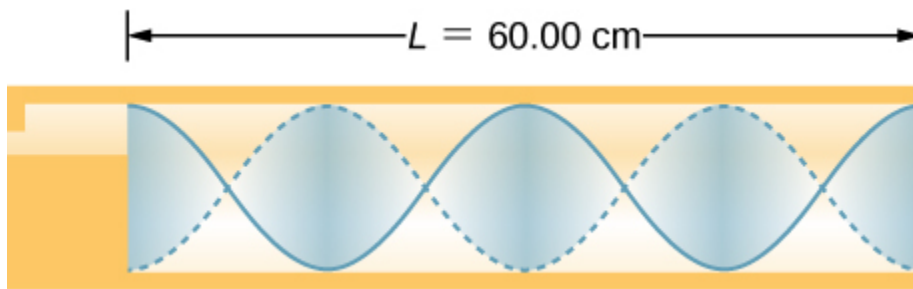
$$\lambda_6 = 0.5 \text{ m}$$

$$v = 1000 \text{ m/s}$$

$$F_T = 6500 \text{ N}$$

Exercise:**Problem:**

Consider the sound created by resonating the tube shown below. The air temperature is $T_C = 30.00^\circ \text{C}$. What are the wavelength, wave speed, and frequency of the sound produced?

**Exercise:****Problem:**

A student holds an 80.00-cm lab pole one quarter of the length from the end of the pole. The lab pole is made of aluminum. The student strikes the lab pole with a hammer. The pole resonates at the lowest possible frequency. What is that frequency?

Solution:

$$f = 6.40 \text{ kHz}$$

Exercise:**Problem:**

A string on the violin has a length of 24.00 cm and a mass of 0.860 g. The fundamental frequency of the string is 1.00 kHz. (a) What is the speed of the wave on the string? (b) What is the tension in the string?

Exercise:**Problem:**

By what fraction will the frequencies produced by a wind instrument change when air temperature goes from 10.0 °C to 30.0 °C? That is, find the ratio of the frequencies at those temperatures.

Solution:

1.03 or 3%

Beats

By the end of this section, you will be able to:

- Determine the beat frequency produced by two sound waves that differ in frequency
- Describe how beats are produced by musical instruments

The study of music provides many examples of the superposition of waves and the constructive and destructive interference that occurs. Very few examples of music being performed consist of a single source playing a single frequency for an extended period of time. You will probably agree that a single frequency of sound for an extended period might be boring to the point of irritation, similar to the unwanted drone of an aircraft engine or a loud fan. Music is pleasant and interesting due to mixing the changing frequencies of various instruments and voices.

An interesting phenomenon that occurs due to the constructive and destructive interference of two or more frequencies of sound is the phenomenon of **beats**. If two sounds differ in frequencies, the sound waves can be modeled as

Equation:

$$y_1 = A \cos(k_1x - 2\pi f_1t) \text{ and } y_2 = A \cos(k_2x - 2\pi f_2t).$$

Using the trigonometric identity $\cos u + \cos v = 2 \cos\left(\frac{u+v}{2}\right) \cos\left(\frac{u-v}{2}\right)$ and considering the point in space as $x = 0.0 \text{ m}$, we find the resulting sound at a point in space, from the superposition of the two sound waves, is equal to [\[link\]](#):

Equation:

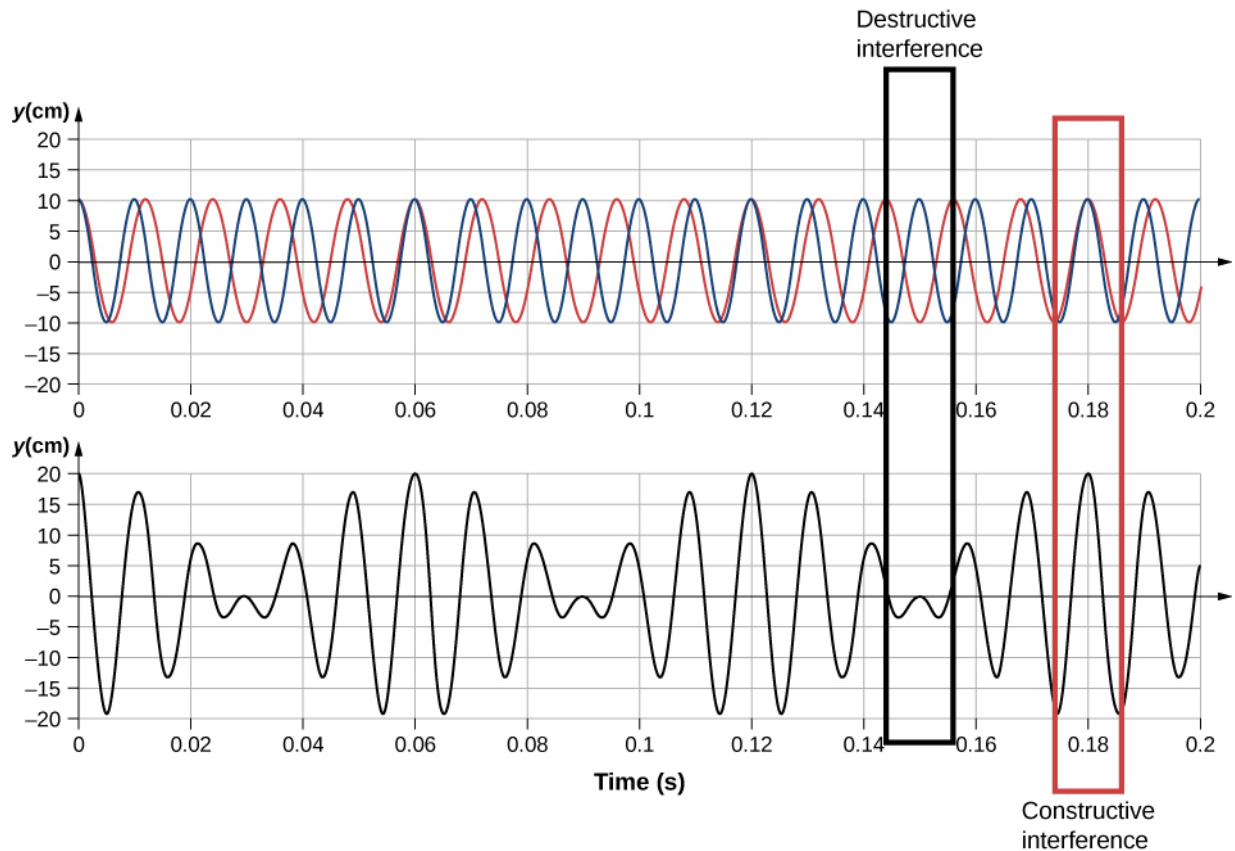
$$y(t) = 2A \cos(2\pi f_{avg}t) \cos\left(2\pi \left(\frac{|f_2 - f_1|}{2}\right)t\right),$$

where the **beat frequency** is

Note:

Equation:

$$f_{\text{beat}} = |f_2 - f_1|.$$



Beats produced by the constructive and destructive interference of two sound waves that differ in frequency.

These beats can be used by piano tuners to tune a piano. A tuning fork is struck and a note is played on the piano. As the piano tuner tunes the string, the beats have a lower frequency as the frequency of the note played approaches the frequency of the tuning fork.

Example:**Find the Beat Frequency Between Two Tuning Forks**

What is the beat frequency produced when a tuning fork of a frequency of 256 Hz and a tuning fork of a frequency of 512 Hz are struck simultaneously?

Strategy

The beat frequency is the difference of the two frequencies.

Solution

We use $f_{\text{beat}} = |f_2 - f_1|$:

Equation:

$$|f_2 - f_1| = (512 - 256) \text{ Hz} = 256 \text{ Hz}.$$

Significance

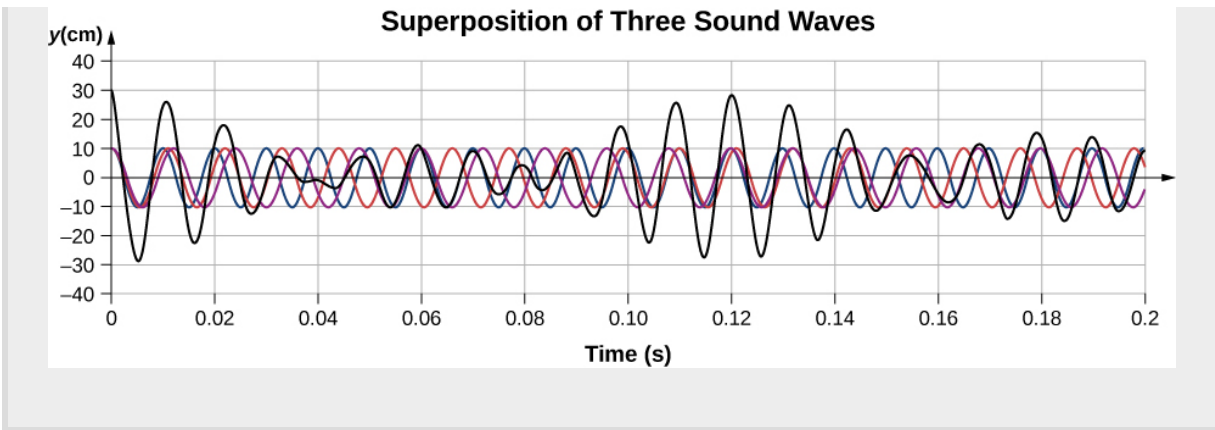
The beat frequency is the absolute value of the difference between the two frequencies. A negative frequency would not make sense.

Note:**Exercise:****Problem:**

Check Your Understanding What would happen if more than two frequencies interacted? Consider three frequencies.

Solution:

An easy way to understand this event is to use a graph, as shown below. It appears that beats are produced, but with a more complex pattern of interference.



The study of the superposition of various waves has many interesting applications beyond the study of sound. In later chapters, we will discuss the wave properties of particles. The particles can be modeled as a “wave packet” that results from the superposition of various waves, where the particle moves at the “group velocity” of the wave packet.

Summary

- When two sound waves that differ in frequency interfere, beats are created with a beat frequency that is equal to the absolute value of the difference in the frequencies.

Conceptual Questions

Exercise:

Problem:

Two speakers are attached to variable-frequency signal generator. Speaker *A* produces a constant-frequency sound wave of 1.00 kHz, and speaker *B* produces a tone of 1.10 kHz. The beat frequency is 0.10 kHz. If the frequency of each speaker is doubled, what is the beat frequency produced?

Exercise:

Problem:

The label has been scratched off a tuning fork and you need to know its frequency. From its size, you suspect that it is somewhere around 250 Hz. You find a 250-Hz tuning fork and a 270-Hz tuning fork. When you strike the 250-Hz fork and the fork of unknown frequency, a beat frequency of 5 Hz is produced. When you strike the unknown with the 270-Hz fork, the beat frequency is 15 Hz. What is the unknown frequency? Could you have deduced the frequency using just the 250-Hz fork?

Solution:

The frequency of the unknown fork is 255 Hz. No, if only the 250 Hz fork is used, listening to the beat frequency could only limit the possible frequencies to 245 Hz or 255 Hz.

Exercise:**Problem:**

Referring to the preceding question, if you had only the 250-Hz fork, could you come up with a solution to the problem of finding the unknown frequency?

Exercise:**Problem:**

A “showy” custom-built car has two brass horns that are supposed to produce the same frequency but actually emit 263.8 and 264.5 Hz. What beat frequency is produced?

Solution:

The beat frequency is 0.7 Hz.

Problems

Exercise:**Problem:**

What beat frequencies are present: (a) If the musical notes A and C are played together (frequencies of 220 and 264 Hz)? (b) If D and F are played together (frequencies of 297 and 352 Hz)? (c) If all four are played together?

Exercise:**Problem:**

What beat frequencies result if a piano hammer hits three strings that emit frequencies of 127.8, 128.1, and 128.3 Hz?

Solution:

$$\begin{aligned}f_B &= |f_1 - f_2| \\|128.3 \text{ Hz} - 128.1 \text{ Hz}| &= 0.2 \text{ Hz}; \\|128.3 \text{ Hz} - 127.8 \text{ Hz}| &= 0.5 \text{ Hz}; \\|128.1 \text{ Hz} - 127.8 \text{ Hz}| &= 0.3 \text{ Hz}\end{aligned}$$

Exercise:**Problem:**

A piano tuner hears a beat every 2.00 s when listening to a 264.0-Hz tuning fork and a single piano string. What are the two possible frequencies of the string?

Exercise:**Problem:**

Two identical strings, of identical lengths of 2.00 m and linear mass density of $\mu = 0.0065 \text{ kg/m}$, are fixed on both ends. String A is under a tension of 120.00 N. String B is under a tension of 130.00 N. They are each plucked and produce sound at the $n = 10$ mode. What is the beat frequency?

Solution:

$$v_A = 135.87 \text{ m/s}, \quad v_B = 141.42 \text{ m/s},$$

$$\lambda_A = \lambda_B = 0.40 \text{ m}$$

$$\Delta f = 15.00 \text{ Hz}$$

Exercise:**Problem:**

A piano tuner uses a 512-Hz tuning fork to tune a piano. He strikes the fork and hits a key on the piano and hears a beat frequency of 5 Hz. He tightens the string of the piano, and repeats the procedure. Once again he hears a beat frequency of 5 Hz. What happened?

Exercise:**Problem:**

A string with a linear mass density of $\mu = 0.0062 \text{ kg/m}$ is stretched between two posts 1.30 m apart. The tension in the string is 150.00 N. The string oscillates and produces a sound wave. A 1024-Hz tuning fork is struck and the beat frequency between the two sources is 52.83 Hz. What are the possible frequency and wavelength of the wave on the string?

Solution:

$$v = 155.54 \text{ m/s},$$

$$f_{\text{string}} = 971.17 \text{ Hz}, \quad n = 16.23$$

$$f_{\text{string}} = 1076.83 \text{ Hz}, \quad n = 18.00$$

The frequency is 1076.83 Hz and the wavelength is 0.14 m.

Exercise:

Problem:

A car has two horns, one emitting a frequency of 199 Hz and the other emitting a frequency of 203 Hz. What beat frequency do they produce?

Exercise:**Problem:**

The middle C hammer of a piano hits two strings, producing beats of 1.50 Hz. One of the strings is tuned to 260.00 Hz. What frequencies could the other string have?

Solution:

$$f_2 = f_1 \pm f_B = 260.00 \text{ Hz} \pm 1.50 \text{ Hz},$$

so that $f_2 = 261.50 \text{ Hz}$ or $f_2 = 258.50 \text{ Hz}$

Exercise:**Problem:**

Two tuning forks having frequencies of 460 and 464 Hz are struck simultaneously. What average frequency will you hear, and what will the beat frequency be?

Exercise:**Problem:**

Twin jet engines on an airplane are producing an average sound frequency of 4100 Hz with a beat frequency of 0.500 Hz. What are their individual frequencies?

Solution:

$$f_{\text{ace}} = \frac{f_1 + f_2}{2}; f_{\text{B}} = f_1 - f_2 (\text{assume } f_1 > f_2)$$

$$f_{\text{ace}} = \frac{(f_{\text{B}} + f_2) + f_2}{2} \Rightarrow$$

$$f_2 = 4099.750 \text{ Hz}$$

$$f_1 = 4100.250 \text{ Hz}$$

Exercise:

Problem:

Three adjacent keys on a piano (F, F-sharp, and G) are struck simultaneously, producing frequencies of 349, 370, and 392 Hz. What beat frequencies are produced by this discordant combination?

Glossary

beat frequency

frequency of beats produced by sound waves that differ in frequency

beats

constructive and destructive interference of two or more frequencies of sound

The Doppler Effect

By the end of this section, you will be able to:

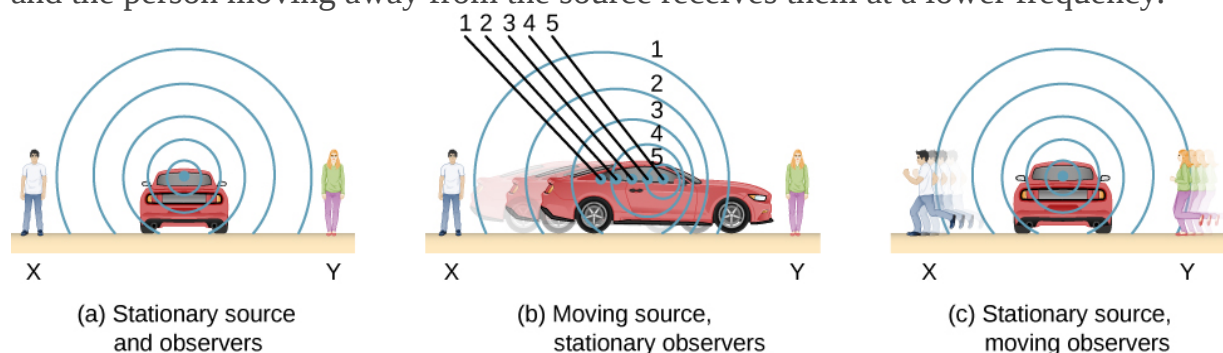
- Explain the change in observed frequency as a moving source of sound approaches or departs from a stationary observer
- Explain the change in observed frequency as an observer moves toward or away from a stationary source of sound

The characteristic sound of a motorcycle buzzing by is an example of the **Doppler effect**. Specifically, if you are standing on a street corner and observe an ambulance with a siren sounding passing at a constant speed, you notice two characteristic changes in the sound of the siren. First, the sound increases in loudness as the ambulance approaches and decreases in loudness as it moves away, which is expected. But in addition, the high-pitched siren shifts dramatically to a lower-pitched sound. As the ambulance passes, the frequency of the sound heard by a stationary observer changes from a constant high frequency to a constant lower frequency, even though the siren is producing a constant source frequency. The closer the ambulance brushes by, the more abrupt the shift. Also, the faster the ambulance moves, the greater the shift. We also hear this characteristic shift in frequency for passing cars, airplanes, and trains.

The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer. Although less familiar, this effect is easily noticed for a stationary source and moving observer. For example, if you ride a train past a stationary warning horn, you will hear the horn's frequency shift from high to low as you pass by. The actual change in frequency due to relative motion of source and observer is called a **Doppler shift**. The Doppler effect and Doppler shift are named for the Austrian physicist and mathematician Christian Johann Doppler (1803–1853), who did experiments with both moving sources and moving observers. Doppler, for example, had musicians play on a moving open train car and also play standing next to the train tracks as a train passed by. Their music was observed both on and off the train, and changes in frequency were measured.

What causes the Doppler shift? [\[link\]](#) illustrates sound waves emitted by stationary and moving sources in a stationary air mass. Each disturbance spreads out spherically from the point at which the sound is emitted. If the source is stationary, then all of the spheres representing the air compressions in the sound wave are centered on the same point, and the stationary observers on either side hear the same wavelength and frequency as emitted by the source (case a). If the source is moving, the situation is different. Each compression of the air moves out in a sphere from the point at which it was emitted, but the point of emission moves. This moving emission point causes

the air compressions to be closer together on one side and farther apart on the other. Thus, the wavelength is shorter in the direction the source is moving (on the right in case b), and longer in the opposite direction (on the left in case b). Finally, if the observers move, as in case (c), the frequency at which they receive the compressions changes. The observer moving toward the source receives them at a higher frequency, and the person moving away from the source receives them at a lower frequency.



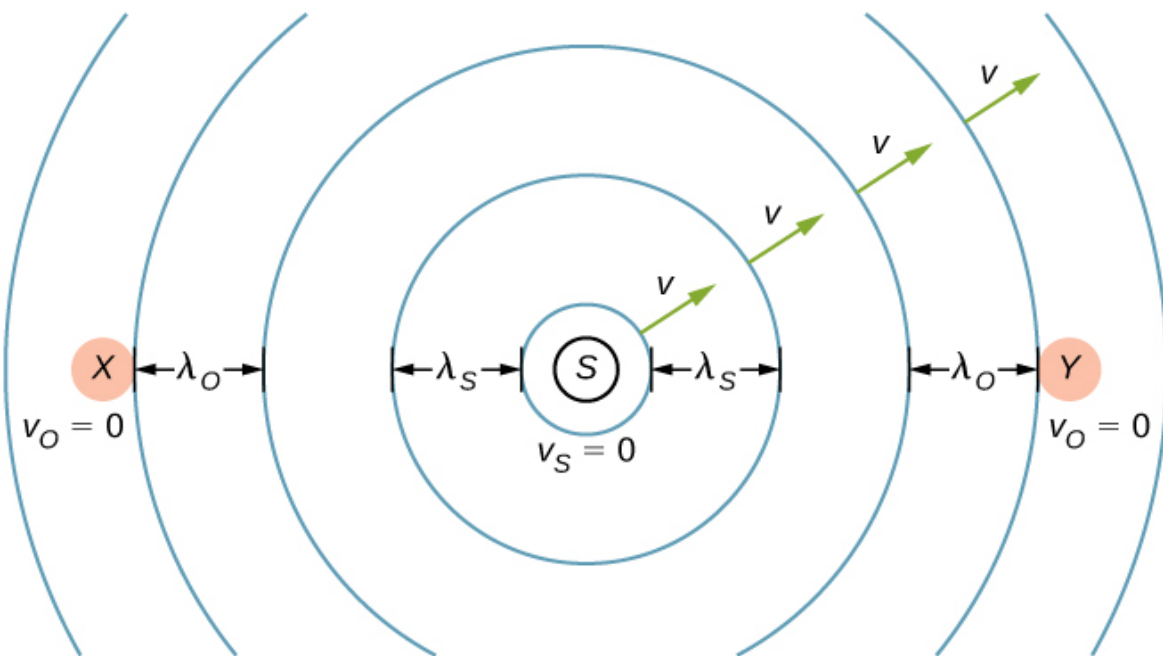
Sounds emitted by a source spread out in spherical waves. (a) When the source, observers, and air are stationary, the wavelength and frequency are the same in all directions and to all observers. (b) Sounds emitted by a source moving to the right spread out from the points at which they were emitted. The wavelength is reduced, and consequently, the frequency is increased in the direction of motion, so that the observer on the right hears a higher-pitched sound. The opposite is true for the observer on the left, where the wavelength is increased and the frequency is reduced. (c) The same effect is produced when the observers move relative to the source. Motion toward the source increases frequency as the observer on the right passes through more wave crests than she would if stationary. Motion away from the source decreases frequency as the observer on the left passes through fewer wave crests than he would if stationary.

We know that wavelength and frequency are related by $v = f\lambda$, where v is the fixed speed of sound. The sound moves in a medium and has the same speed v in that medium whether the source is moving or not. Thus, f multiplied by λ is a constant. Because the observer on the right in case (b) receives a shorter wavelength, the frequency she receives must be higher. Similarly, the observer on the left receives a longer wavelength, and hence he hears a lower frequency. The same thing happens in case (c). A higher frequency is received by the observer moving toward the source, and a lower frequency is received by an observer moving away from the source. In general, then, relative motion of source and observer toward one another increases the received frequency. Relative motion apart decreases frequency. The greater the relative speed, the greater the effect.

The Doppler effect occurs not only for sound, but for any wave when there is relative motion between the observer and the source. Doppler shifts occur in the frequency of sound, light, and water waves, for example. Doppler shifts can be used to determine velocity, such as when ultrasound is reflected from blood in a medical diagnostic. The relative velocities of stars and galaxies is determined by the shift in the frequencies of light received from them and has implied much about the origins of the universe. Modern physics has been profoundly affected by observations of Doppler shifts.

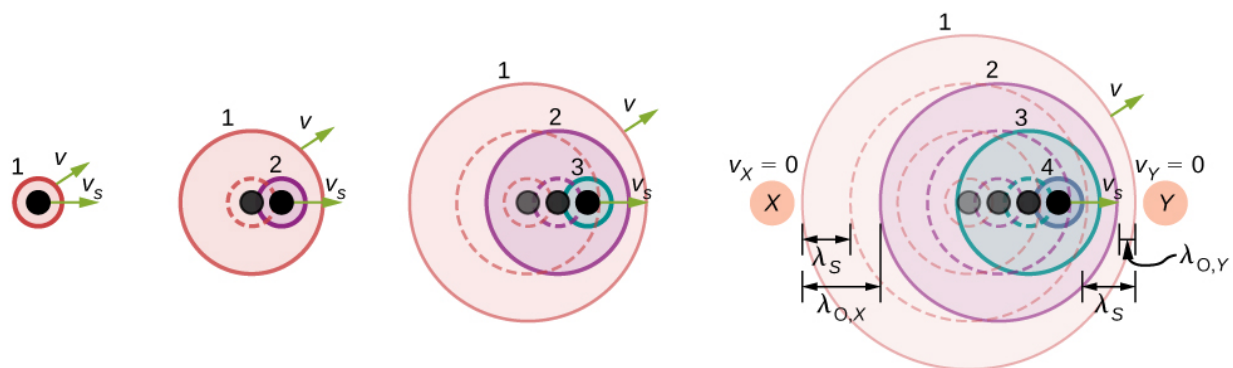
Derivation of the Observed Frequency due to the Doppler Shift

Consider two stationary observers X and Y in [\[link\]](#), located on either side of a stationary source. Each observer hears the same frequency, and that frequency is the frequency produced by the stationary source.



A stationary source sends out sound waves at a constant frequency f_s , with a constant wavelength λ_s , at the speed of sound v . Two stationary observers X and Y , on either side of the source, observe a frequency $f_o = f_s$, with a wavelength $\lambda_o = \lambda_s$.

Now consider a stationary observer X with a source moving away from the observer with a constant speed $v_s < v$ ([link](#)). At time $t = 0$, the source sends out a sound wave, indicated in black. This wave moves out at the speed of sound v . The position of the sound wave at each time interval of period T_s is shown as dotted lines. After one period, the source has moved $\Delta x = v_s T_s$ and emits a second sound wave, which moves out at the speed of sound. The source continues to move and produce sound waves, as indicated by the circles numbered 3 and 4. Notice that as the waves move out, they remained centered at their respective point of origin.



A source moving at a constant speed v_s away from an observer X . The moving source sends out sound waves at a constant frequency f_s , with a constant wavelength λ_s , at the speed of sound v . Snapshots of the source at an interval of T_s are shown as the source moves away from the stationary observer X . The solid lines represent the position of the sound waves after four periods from the initial time. The dotted lines are used to show the positions of the waves at each time period. The observer hears a wavelength of $\lambda_o = \lambda_s + \Delta x = \lambda_s + v_s T_s$.

Using the fact that the wavelength is equal to the speed times the period, and the period is the inverse of the frequency, we can derive the observed frequency:

Equation:

$$\begin{aligned}\lambda_o &= \lambda_s + \Delta x \\ vT_o &= vT_s + v_s T_s \\ \frac{v}{f_o} &= \frac{v}{f_s} = \frac{v_s}{f_s} = \frac{v+v_s}{f_s} \\ f_o &= f_s \left(\frac{v}{v+v_s} \right).\end{aligned}$$

As the source moves away from the observer, the observed frequency is lower than the source frequency.

Now consider a source moving at a constant velocity v_s , moving toward a stationary observer Y, also shown in [\[link\]](#). The wavelength is observed by Y as

$\lambda_o = \lambda_s - \Delta x = \lambda_s - v_s T_s$. Once again, using the fact that the wavelength is equal to the speed times the period, and the period is the inverse of the frequency, we can derive the observed frequency:

Equation:

$$\begin{aligned}\lambda_o &= \lambda_s - \Delta x \\ vT_o &= vT_s - v_s T_s \\ \frac{v}{f_o} &= \frac{v}{f_s} - \frac{v_s}{f_s} = \frac{v-v_s}{f_s} \\ f_o &= f_s \left(\frac{v}{v-v_s} \right).\end{aligned}$$

When a source is moving and the observer is stationary, the observed frequency is

Note:

Equation:

$$f_o = f_s \left(\frac{v}{v \mp v_s} \right),$$

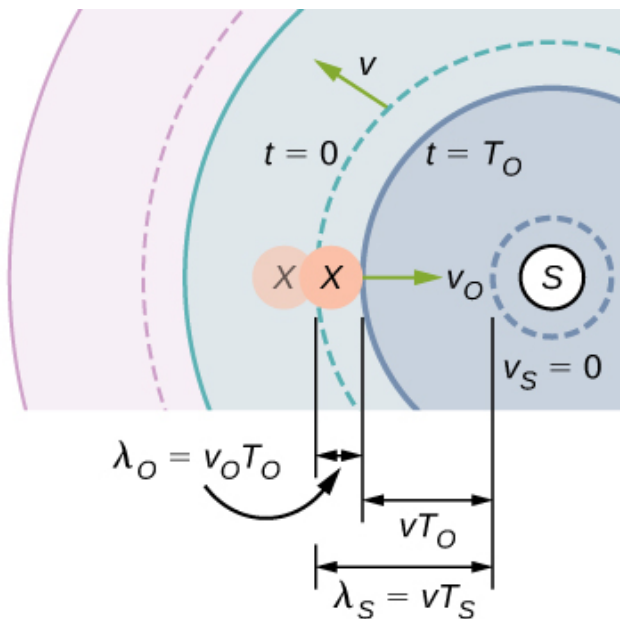
where f_o is the frequency observed by the stationary observer, f_s is the frequency produced by the moving source, v is the speed of sound, v_s is the constant speed of the source, and the top sign is for the source approaching the observer and the bottom sign is for the source departing from the observer.

What happens if the observer is moving and the source is stationary? If the observer moves toward the stationary source, the observed frequency is higher than the source frequency. If the observer is moving away from the stationary source, the observed frequency is lower than the source frequency. Consider observer X in [\[link\]](#) as the observer moves toward a stationary source with a speed v_o . The source emits a tone with a constant frequency f_s and constant period T_s . The observer hears the first wave emitted by the source. If the observer were stationary, the time for one

wavelength of sound to pass should be equal to the period of the source T_s . Since the observer is moving toward the source, the time for one wavelength to pass is less than T_s and is equal to the observed period $T_o = T_s - \Delta t$. At time $t = 0$, the observer starts at the beginning of a wavelength and moves toward the second wavelength as the wavelength moves out from the source. The wavelength is equal to the distance the observer traveled plus the distance the sound wave traveled until it is met by the observer:

Equation:

$$\begin{aligned}\lambda_s &= vT_o + v_oT_o \\ vT_s &= (v + v_o)T_o \\ v\left(\frac{1}{f_s}\right) &= (v + v_o)\left(\frac{1}{f_o}\right) \\ f_o &= f_s\left(\frac{v+v_o}{v}\right).\end{aligned}$$



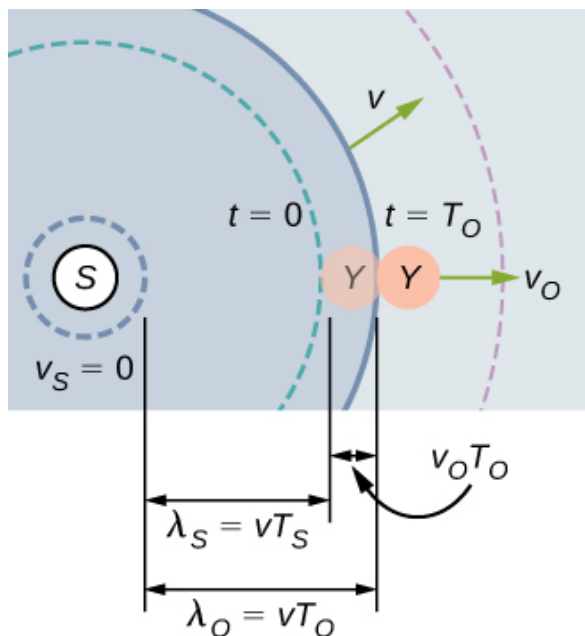
A stationary source emits a sound wave with a constant frequency f_s , with a constant wavelength λ_s moving at the speed of sound v . Observer X moves toward the source with a constant speed v_o , and the figure shows the initial and final position of

observer X. Observer X observes a frequency higher than the source frequency. The dotted lines show the position of the waves at $t = 0$. The solid lines show the position of the waves at $t = T_o$.

If the observer is moving away from the source ([link](#)), the observed frequency can be found:

Equation:

$$\begin{aligned}\lambda_s &= vT_o - v_oT_o \\ vT_s &= (v - v_o)T_o \\ v\left(\frac{1}{f_s}\right) &= (v - v_o)\left(\frac{1}{f_o}\right) \\ f_o &= f_s\left(\frac{v - v_o}{v}\right).\end{aligned}$$



A stationary source emits a sound wave with a constant frequency f_s , with a constant wavelength λ_s moving at the speed of sound v .

moving at the speed of sound v .
 Observer Y moves away from the source with a constant speed v_o , and the figure shows initial and final position of the observer Y . Observer Y observes a frequency lower than the source frequency. The dotted lines show the position of the waves at $t = 0$. The solid lines show the position of the waves at $t = T_o$.

The equations for an observer moving toward or away from a stationary source can be combined into one equation:

Note:

Equation:

$$f_o = f_s \left(\frac{v \pm v_o}{v} \right),$$

where f_o is the observed frequency, f_s is the source frequency, v is the speed of sound, v_o is the speed of the observer, the top sign is for the observer approaching the source and the bottom sign is for the observer departing from the source.

[\[link\]](#) and [\[link\]](#) can be summarized in one equation (the top sign is for approaching) and is further illustrated in [\[link\]](#):

Note:

Equation:

$$f_o = f_s \left(\frac{v \pm v_o}{v \mp v_s} \right),$$

Doppler shift $f_o = f_s \left(\frac{v \pm v_o}{v \mp v_s} \right)$	Stationary observer	Observer moving towards source	Observer moving away from source
Stationary source	$f_o = f_s$	$f_o = f_s \left(\frac{v + v_o}{v} \right)$	$f_o = f_s \left(\frac{v - v_o}{v} \right)$
Source moving towards observer	$f_o = f_s \left(\frac{v}{v - v_s} \right)$	$f_o = f_s \left(\frac{v + v_o}{v - v_s} \right)$	$f_o = f_s \left(\frac{v - v_o}{v - v_s} \right)$
Source moving away from observer	$f_o = f_s \left(\frac{v}{v + v_s} \right)$	$f_o = f_s \left(\frac{v + v_o}{v + v_s} \right)$	$f_o = f_s \left(\frac{v - v_o}{v + v_s} \right)$

where f_o is the observed frequency, f_s is the source frequency, v is the speed of sound, v_o is the speed of the observer, v_s is the speed of the source, the top sign is for approaching and the bottom sign is for departing.

Note:

The Doppler effect involves motion and a [video](#) will help visualize the effects of a moving observer or source. This video shows a moving source and a stationary observer, and a moving observer and a stationary source. It also discusses the Doppler effect and its application to light.

Example:

Calculating a Doppler Shift

Suppose a train that has a 150-Hz horn is moving at 35.0 m/s in still air on a day when the speed of sound is 340 m/s.

(a) What frequencies are observed by a stationary person at the side of the tracks as the train approaches and after it passes?

(b) What frequency is observed by the train's engineer traveling on the train?

Strategy

To find the observed frequency in (a), we must use $f_{\text{obs}} = f_s \left(\frac{v}{v \mp v_s} \right)$ because the source is moving. The minus sign is used for the approaching train, and the plus sign

for the receding train. In (b), there are two Doppler shifts—one for a moving source and the other for a moving observer.

Solution

a. Enter known values into $f_o = f_s \left(\frac{v}{v - v_s} \right)$:

Equation:

$$f_o = f_s \left(\frac{v}{v - v_s} \right) = (150 \text{ Hz}) \left(\frac{340 \text{ m/s}}{340 \text{ m/s} - 35.0 \text{ m/s}} \right).$$

Calculate the frequency observed by a stationary person as the train approaches:

Equation:

$$f_o = (150 \text{ Hz}) (1.11) = 167 \text{ Hz}.$$

Use the same equation with the plus sign to find the frequency heard by a stationary person as the train recedes:

Equation:

$$f_o = f_s \left(\frac{v}{v + v_s} \right) = (150 \text{ Hz}) \left(\frac{340 \text{ m/s}}{340 \text{ m/s} + 35.0 \text{ m/s}} \right).$$

Calculate the second frequency:

Equation:

$$f_o = (150 \text{ Hz}) (0.907) = 136 \text{ Hz}.$$

b. Identify knowns:

- It seems reasonable that the engineer would receive the same frequency as emitted by the horn, because the relative velocity between them is zero.
- Relative to the medium (air), the speeds are $v_s = v_o = 35.0 \text{ m/s}$.
- The first Doppler shift is for the moving observer; the second is for the moving source.

Use the following equation:

Equation:

$$f_o = \left[f_s \left(\frac{v \pm v_o}{v} \right) \right] \left(\frac{v}{v \mp v_s} \right).$$

The quantity in the square brackets is the Doppler-shifted frequency due to a moving observer. The factor on the right is the effect of the moving source. Because the train engineer is moving in the direction toward the horn, we must use the plus sign for v_{obs} ; however, because the horn is also moving in the direction away from the engineer, we also use the plus sign for v_s . But the train is carrying both the engineer and the horn at the same velocity, so $v_s = v_o$. As a result, everything but f_s cancels, yielding

Equation:

$$f_o = f_s.$$

Significance

For the case where the source and the observer are not moving together, the numbers calculated are valid when the source (in this case, the train) is far enough away that the motion is nearly along the line joining source and observer. In both cases, the shift is significant and easily noticed. Note that the shift is 17.0 Hz for motion toward and 14.0 Hz for motion away. The shifts are not symmetric.

For the engineer riding in the train, we may expect that there is no change in frequency because the source and observer move together. This matches your experience. For example, there is no Doppler shift in the frequency of conversations between driver and passenger on a motorcycle. People talking when a wind moves the air between them also observe no Doppler shift in their conversation. The crucial point is that source and observer are not moving relative to each other.

Note:

Exercise:

Problem:

Check Your Understanding Describe a situation in your life when you might rely on the Doppler shift to help you either while driving a car or walking near traffic.

Solution:

If I am driving and I hear Doppler shift in an ambulance siren, I would be able to tell when it was getting closer and also if it has passed by. This would help me to know whether I needed to pull over and let the ambulance through.

The Doppler effect and the Doppler shift have many important applications in science and engineering. For example, the Doppler shift in ultrasound can be used to measure blood velocity, and police use the Doppler shift in radar (a microwave) to measure car velocities. In meteorology, the Doppler shift is used to track the motion of storm clouds; such “Doppler Radar” can give the velocity and direction of rain or snow in weather fronts. In astronomy, we can examine the light emitted from distant galaxies and determine their speed relative to ours. As galaxies move away from us, their light is shifted to a lower frequency, and so to a longer wavelength—the so-called red shift. Such information from galaxies far, far away has allowed us to estimate the age of the universe (from the Big Bang) as about 14 billion years.

Summary

- The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer.
- The actual change in frequency is called the Doppler shift.

Conceptual Questions

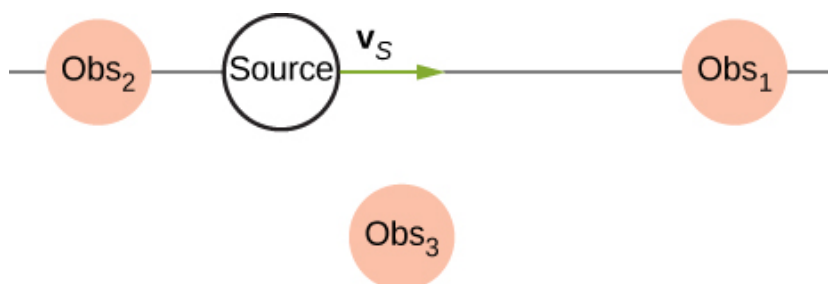
Exercise:

Problem: Is the Doppler shift real or just a sensory illusion?

Exercise:

Problem:

Three stationary observers observe the Doppler shift from a source moving at a constant velocity. The observers are stationed as shown below. Which observer will observe the highest frequency? Which observer will observe the lowest frequency? What can be said about the frequency observed by observer 3?



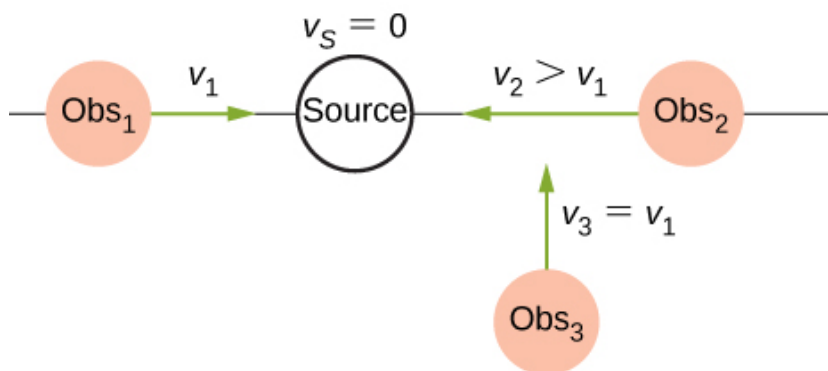
Solution:

Observer 1 will observe the highest frequency. Observer 2 will observe the lowest frequency. Observer 3 will hear a higher frequency than the source frequency, but lower than the frequency observed by observer 1, as the source approaches and a lower frequency than the source frequency, but higher than the frequency observed by observer 1, as the source moves away from observer 3.

Exercise:

Problem:

Shown below is a stationary source and moving observers. Describe the frequencies observed by the observers for this configuration.



Exercise:

Problem:

Prior to 1980, conventional radar was used by weather forecasters. In the 1960s, weather forecasters began to experiment with Doppler radar. What do you think is the advantage of using Doppler radar?

Solution:

Doppler radar can not only detect the distance to a storm, but also the speed and direction at which the storm is traveling.

Problems

Exercise:

Problem:

(a) What frequency is received by a person watching an oncoming ambulance moving at 110 km/h and emitting a steady 800-Hz sound from its siren? The speed of sound on this day is 345 m/s. (b) What frequency does she receive after the ambulance has passed?

Solution:

a. 878 Hz; b. 735 Hz

Exercise:**Problem:**

(a) At an air show a jet flies directly toward the stands at a speed of 1200 km/h, emitting a frequency of 3500 Hz, on a day when the speed of sound is 342 m/s. What frequency is received by the observers? (b) What frequency do they receive as the plane flies directly away from them?

Exercise:**Problem:**

What frequency is received by a mouse just before being dispatched by a hawk flying at it at 25.0 m/s and emitting a screech of frequency 3500 Hz? Take the speed of sound to be 331 m/s.

Solution:

$3.79 \times 10^3 \text{ Hz}$

Exercise:**Problem:**

A spectator at a parade receives an 888-Hz tone from an oncoming trumpeter who is playing an 880-Hz note. At what speed is the musician approaching if the speed of sound is 338 m/s?

Exercise:

Problem:

A commuter train blows its 200-Hz horn as it approaches a crossing. The speed of sound is 335 m/s. (a) An observer waiting at the crossing receives a frequency of 208 Hz. What is the speed of the train? (b) What frequency does the observer receive as the train moves away?

Solution:

a. 12.9 m/s; b. 193 Hz

Exercise:**Problem:**

Can you perceive the shift in frequency produced when you pull a tuning fork toward you at 10.0 m/s on a day when the speed of sound is 344 m/s? To answer this question, calculate the factor by which the frequency shifts and see if it is greater than 0.300%.

Exercise:**Problem:**

Two eagles fly directly toward one another, the first at 15.0 m/s and the second at 20.0 m/s. Both screech, the first one emitting a frequency of 3200 Hz and the second one emitting a frequency of 3800 Hz. What frequencies do they receive if the speed of sound is 330 m/s?

Solution:

The first eagle hears 4.23×10^3 Hz. The second eagle hears 3.56×10^3 Hz.

Exercise:**Problem:**

Student A runs down the hallway of the school at a speed of $v_o = 5.00$ m/s, carrying a ringing 1024.00-Hz tuning fork toward a concrete wall. The speed of sound is $v = 343.00$ m/s. Student B stands at rest at the wall. (a) What is the frequency heard by student B? (b) What is the beat frequency heard by student A?

Exercise:

Problem:

An ambulance with a siren ($f = 1.00\text{kHz}$) blaring is approaching an accident scene. The ambulance is moving at 70.00 mph. A nurse is approaching the scene from the opposite direction, running at $v_o = 7.00\text{ m/s}$. What frequency does the nurse observe? Assume the speed of sound is $v = 343.00\text{ m/s}$.

Solution:

$$v_s = 31.29\text{ m/s}$$

$$f_o = 1.12\text{ kHz}$$

Exercise:**Problem:**

The frequency of the siren of an ambulance is 900 Hz and is approaching you. You are standing on a corner and observe a frequency of 960 Hz. What is the speed of the ambulance (in mph) if the speed of sound is $v = 340.00\text{ m/s}$?

Exercise:**Problem:**

What is the minimum speed at which a source must travel toward you for you to be able to hear that its frequency is Doppler shifted? That is, what speed produces a shift of 0.300% on a day when the speed of sound is 331 m/s?

Solution:

$$\text{An audible shift occurs when } \frac{f_{\text{obs}}}{f_s} \geq 1.003; f_{\text{obs}} = f_s \frac{v}{v-v_s} \Rightarrow \frac{f_{\text{obs}}}{f_s} = \frac{v}{v-v_s} \Rightarrow v_s = 0.990\text{ m/s}$$

Glossary**Doppler effect**

alteration in the observed frequency of a sound due to motion of either the source or the observer

Doppler shift

actual change in frequency due to relative motion of source and observer

Shock Waves

By the end of this section, you will be able to:

- Explain the mechanism behind sonic booms
- Describe the difference between sonic booms and shock waves
- Describe a bow wake

When discussing the Doppler effect of a moving source and a stationary observer, the only cases we considered were cases where the source was moving at speeds that were less than the speed of sound. Recall that the observed frequency for a moving source approaching a stationary observer is $f_o = f_s \left(\frac{v}{v - v_s} \right)$. As the source approaches the speed of sound, the observed frequency increases. According to the equation, if the source moves at the speed of sound, the denominator is equal to zero, implying the observed frequency is infinite. If the source moves at speeds greater than the speed of sound, the observed frequency is negative.

What could this mean? What happens when a source approaches the speed of sound? It was once argued by some scientists that such a large pressure wave would result from the constructive interference of the sound waves, that it would be impossible for a plane to exceed the speed of sound because the pressures would be great enough to destroy the airplane. But now planes routinely fly faster than the speed of sound. On July 28, 1976, Captain Eldon W. Joersz and Major George T. Morgan flew a Lockheed SR-71 Blackbird #61-7958 at 3529.60 km/h (2193.20 mi/h), which is Mach 2.85. The Mach number is the speed of the source divided by the speed of sound:

Note:

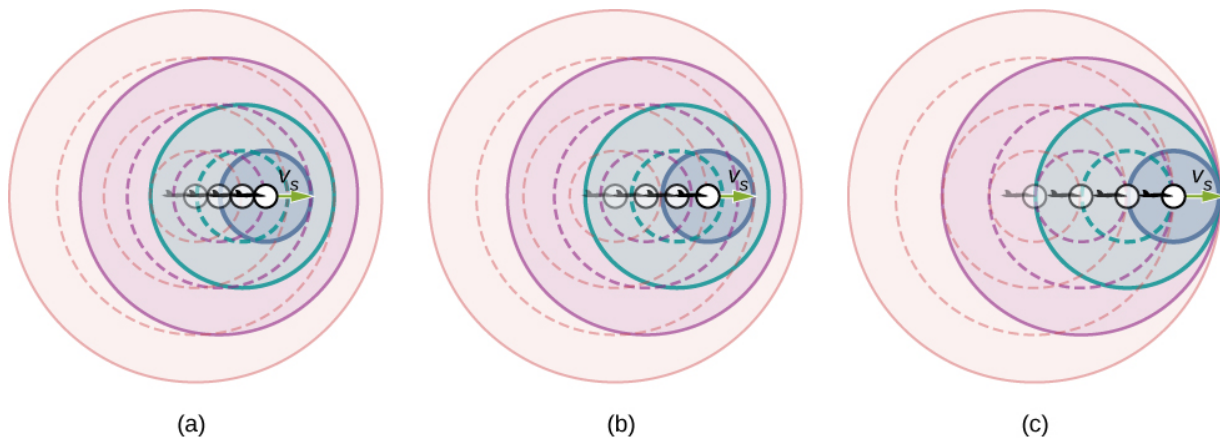
Equation:

$$M = \frac{v_s}{v}.$$

You will see that interesting phenomena occur when a source approaches and exceeds the speed of sound.

Doppler Effect and High Velocity

What happens to the sound produced by a moving source, such as a jet airplane, that approaches or even exceeds the speed of sound? The answer to this question applies not only to sound but to all other waves as well. Suppose a jet plane is coming nearly straight at you, emitting a sound of frequency f_s . The greater the plane's speed v_s , the greater the Doppler shift and the greater the value observed for f_o ([link](#)).

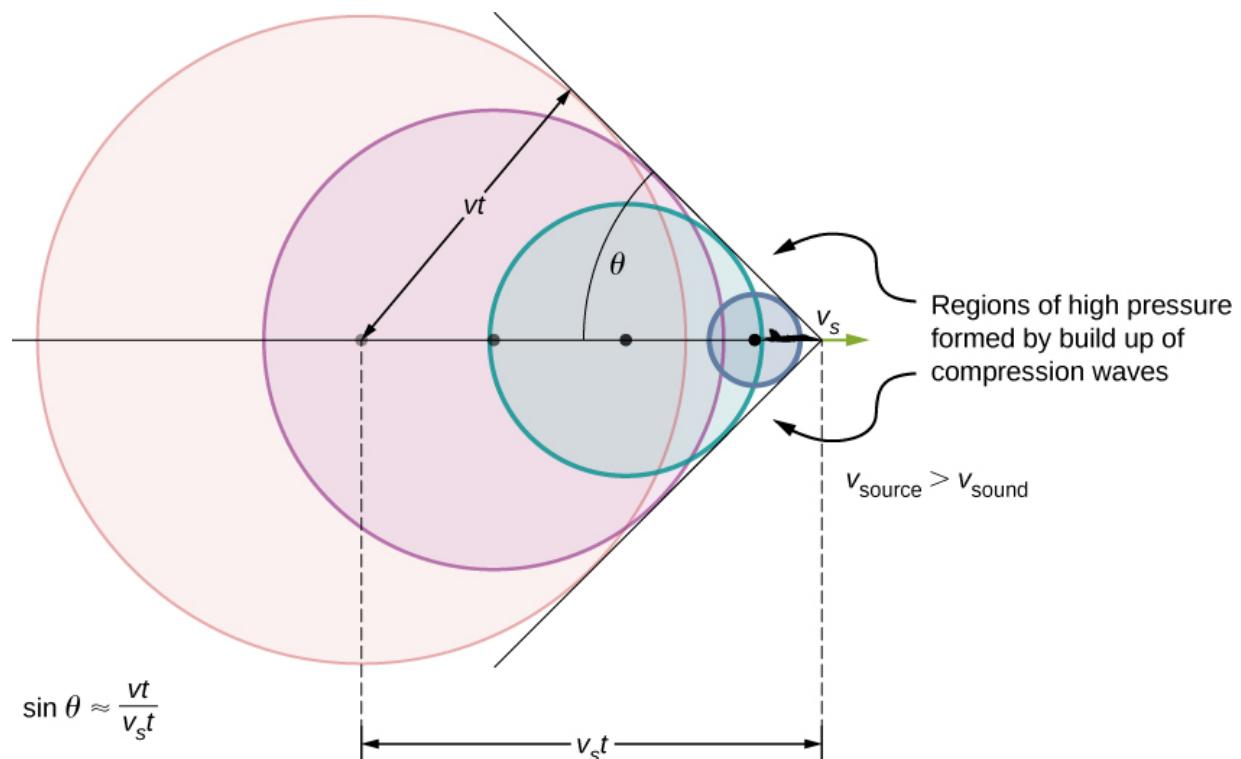


Because of the Doppler shift, as a moving source approaches a stationary observer, the observed frequency is higher than the source frequency. The faster the source is moving, the higher the observed frequency. In this figure, the source in (b) is moving faster than the source in (a). Shown are four time steps, the first three shown as dotted lines. (c) If a source moves at the speed of sound, each successive wave interfere with the previous one and the observer observes them all at the same instant.

Now, as v_s approaches the speed of sound, f_o approaches infinity, because the denominator in $f_o = f_s \left(\frac{v}{v \mp v_s} \right)$ approaches zero. At the speed of sound, this result means that in front of the source, each successive wave interferes with the previous one because the source moves forward at the speed of sound. The observer gets them all at the same instant, so the frequency is infinite [part (c) of the figure].

Shock Waves and Sonic Booms

If the source exceeds the speed of sound, no sound is received by the observer until the source has passed, so that the sounds from the approaching source are mixed with those from it when receding. This mixing appears messy, but something interesting happens—a shock wave is created ([link](#)).



Sound waves from a source that moves faster than the speed of sound spread spherically from the point where they are emitted, but the

source moves ahead of each wave. Constructive interference along the lines shown (actually a cone in three dimensions) creates a shock wave called a sonic boom. The faster the speed of the source, the smaller the angle θ .

Constructive interference along the lines shown (a cone in three dimensions) from similar sound waves arriving there simultaneously. This superposition forms a disturbance called a **shock wave**, a constructive interference of sound created by an object moving faster than sound. Inside the cone, the interference is mostly destructive, so the sound intensity there is much less than on the shock wave. The angle of the shock wave can be found from the geometry. In time t the source has moved $v_s t$ and the sound wave has moved a distance vt and the angle can be found using $\sin \theta = \frac{vt}{v_s t} = \frac{v}{v_s}$. Note that the Mach number is defined as $\frac{v_s}{v}$ so the sine of the angle equals the inverse of the Mach number,

Note:

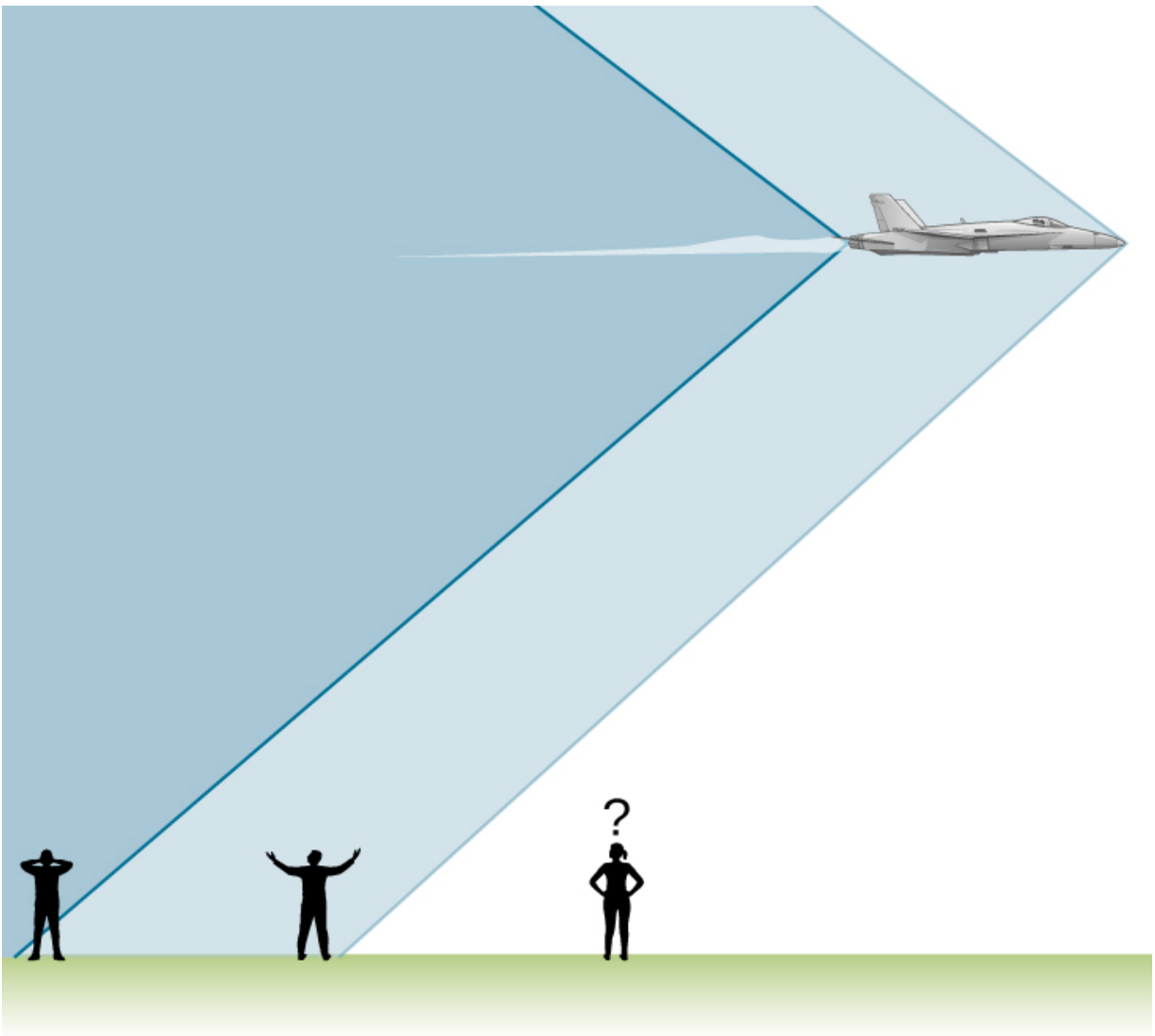
Equation:

$$\sin \theta = \frac{v}{v_s} = \frac{1}{M}.$$

You may have heard of the common term ‘**sonic boom**.’ A common misconception is that the sonic boom occurs as the plane breaks the sound barrier; that is, accelerates to a speed higher than the speed of sound. Actually, the sonic boom occurs as the shock wave sweeps along the ground.

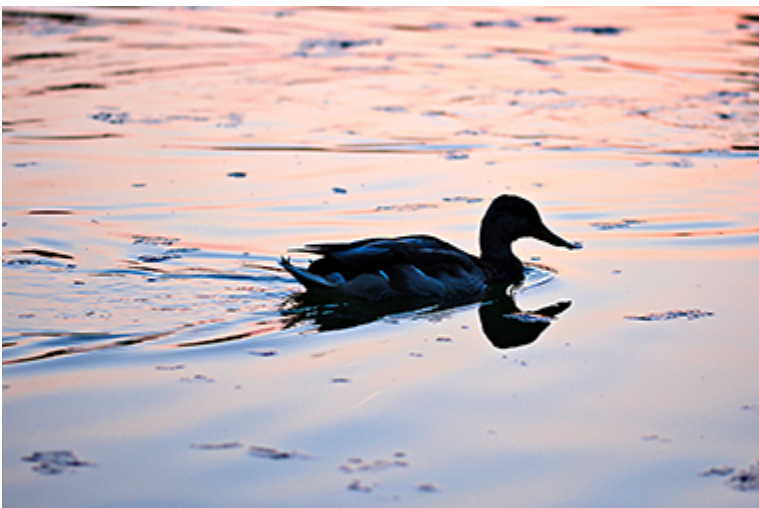
An aircraft creates two shock waves, one from its nose and one from its tail ([link](#)). During television coverage of space shuttle landings, two distinct booms could often be heard. These were separated by exactly the time it

would take the shuttle to pass by a point. Observers on the ground often do not see the aircraft creating the sonic boom, because it has passed by before the shock wave reaches them, as seen in the figure. If the aircraft flies close by at low altitude, pressures in the sonic boom can be destructive and break windows as well as rattle nerves. Because of how destructive sonic booms can be, supersonic flights are banned over populated areas.



Two sonic booms experienced by observers, created by the nose and tail of an aircraft as the shock wave sweeps along the ground, are observed on the ground after the plane has passed by.

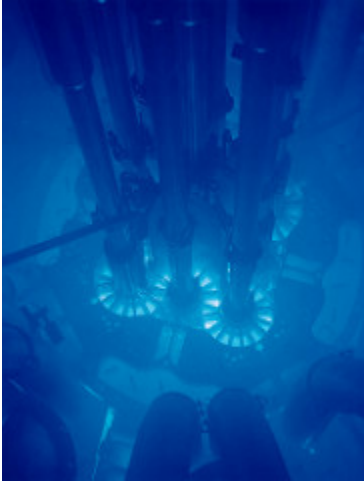
Shock waves are one example of a broader phenomenon called bow wakes. A **bow wake**, such as the one in [\[link\]](#), is created when the wave source moves faster than the wave propagation speed. Water waves spread out in circles from the point where created, and the bow wake is the familiar V-shaped wake, trailing the source. A more exotic bow wake is created when a subatomic particle travels through a medium faster than the speed of light travels in that medium. (In a vacuum, the maximum speed of light is $c = 3.00 \times 10^8$ m/s; in the medium of water, the speed of light is closer to $0.75c$.) If the particle creates light in its passage, that light spreads on a cone with an angle indicative of the speed of the particle, as illustrated in [\[link\]](#). Such a bow wake is called Cerenkov radiation and is commonly observed in particle physics.



Bow wake created by a duck.

Constructive interference produces the rather structured wake, whereas relatively little wave action occurs inside the wake, where interference is mostly destructive.

(credit: Horia Varlan)



The blue glow in this research reactor pool is Cerenkov radiation caused by subatomic particles traveling faster than the speed of light in water.
(credit: Idaho National Laboratory)

Summary

- The Mach number is the velocity of a source divided by the speed of sound, $M = \frac{v_s}{v}$.
- When a sound source moves faster than the speed of sound, a shock wave is produced as the sound waves interfere.

- A sonic boom is the intense sound that occurs as the shock wave moves along the ground.
- The angle the shock wave produces can be found as $\sin \theta = \frac{v}{v_s} = \frac{1}{M}$.
- A bow wake is produced when an object moves faster than the speed of a mechanical wave in the medium, such as a boat moving through the water.

Key Equations

Pressure of a sound wave	$\Delta P = \Delta P_{\max} \sin(kx \mp \omega t + \phi)$
Displacement of the oscillating molecules of a sound wave	$s(x, t) = s_{\max} \cos(kx \mp \omega t + \phi)$
Velocity of a wave	$v = f\lambda$
Speed of sound in a fluid	$v = \sqrt{\frac{\beta}{\rho}}$
Speed of sound in a solid	$v = \sqrt{\frac{Y}{\rho}}$
Speed of sound in an ideal gas	$v = \sqrt{\frac{\gamma RT}{M}}$
Speed of sound in air as a function of temperature	$v = 331 \frac{\text{m}}{\text{s}} \sqrt{\frac{T_K}{273 \text{ K}}} = 331 \frac{\text{m}}{\text{s}} \sqrt{1 + \frac{T_C}{273^\circ \text{C}}}$

Decrease in intensity as a spherical wave expands	$I_2 = I_1 \left(\frac{r_1}{r_2} \right)^2$
Intensity averaged over a period	$I = \frac{\langle P \rangle}{A}$
Intensity of sound	$I = \frac{(\Delta p_{\max})^2}{2\rho v}$
Sound intensity level	$\beta \text{ (dB)} = 10 \log_{10} \left(\frac{I}{I_0} \right)$
Resonant wavelengths of a tube closed at one end	$\lambda_n = \frac{4}{n}L, \quad n = 1, 3, 5, \dots$
Resonant frequencies of a tube closed at one end	$f_n = n \frac{v}{4L}, \quad n = 1, 3, 5, \dots$
Resonant wavelengths of a tube open at both ends	$\lambda_n = \frac{2}{n}L, \quad n = 1, 2, 3, \dots$
Resonant frequencies of a tube open at both ends	$f_n = n \frac{v}{2L}, \quad n = 1, 2, 3, \dots$
Beat frequency produced by two waves that differ in frequency	$f_{\text{beat}} = f_2 - f_1 $
Observed frequency for a stationary observer and a moving source	$f_o = f_s \left(\frac{v}{v \mp v_s} \right)$

Observed frequency for a moving observer and a stationary source	$f_o = f_s \left(\frac{v \pm v_o}{v} \right)$
Doppler shift for the observed frequency	$f_o = f_s \left(\frac{v \pm v_o}{v \mp v_s} \right)$
Mach number	$M = \frac{v_s}{v}$
Sine of angle formed by shock wave	$\sin \theta = \frac{v}{v_s} = \frac{1}{M}$

Conceptual Questions

Exercise:

Problem:

What is the difference between a sonic boom and a shock wave?

Exercise:

Problem:

Due to efficiency considerations related to its bow wake, the supersonic transport aircraft must maintain a cruising speed that is a constant ratio to the speed of sound (a constant Mach number). If the aircraft flies from warm air into colder air, should it increase or decrease its speed? Explain your answer.

Solution:

The speed of sound decreases as the temperature decreases. The Mach number is equal to $M = \frac{v_s}{v}$, so the plane should slow down.

Exercise:

Problem:

When you hear a sonic boom, you often cannot see the plane that made it. Why is that?

Problems**Exercise:****Problem:**

An airplane is flying at Mach 1.50 at an altitude of 7500.00 meters, where the speed of sound is $v = 343.00$ m/s. How far away from a stationary observer will the plane be when the observer hears the sonic boom?

Exercise:**Problem:**

A jet flying at an altitude of 8.50 km has a speed of Mach 2.00, where the speed of sound is $v = 340.00$ m/s. How long after the jet is directly overhead, will a stationary observer hear a sonic boom?

Solution:

$$\theta = 30.02^\circ$$

$$v_s = 680.00 \text{ m/s}$$

$$\tan \theta = \frac{y}{v_s t}, \quad t = 21.65 \text{ s}$$

Exercise:**Problem:**

The shock wave off the front of a fighter jet has an angle of $\theta = 70.00^\circ$. The jet is flying at 1200 km/h. What is the speed of sound?

Exercise:**Problem:**

A plane is flying at Mach 1.2, and an observer on the ground hears the sonic boom 15.00 seconds after the plane is directly overhead. What is the altitude of the plane? Assume the speed of sound is $v_w = 343.00 \text{ m/s}$.

Solution:

$$\sin \theta = \frac{1}{M}, \quad \theta = 56.47^\circ$$

$$y = 9.31 \text{ km}$$

Exercise:**Problem:**

A bullet is fired and moves at a speed of 1342 mph. Assume the speed of sound is $v = 340.00 \text{ m/s}$. What is the angle of the shock wave produced?

Exercise:**Problem:**

A speaker is placed at the opening of a long horizontal tube. The speaker oscillates at a frequency of f , creating a sound wave that moves down the tube. The wave moves through the tube at a speed of $v = 340.00 \text{ m/s}$. The sound wave is modeled with the wave function $s(x, t) = s_{\max} \cos(kx - \omega t + \phi)$. At time $t = 0.00 \text{ s}$, an air molecule at $x = 2.3 \text{ m}$ is at the maximum displacement of 6.34 nm . At the same time, another molecule at $x = 2.7 \text{ m}$ has a displacement of 2.30 nm . What is the wave function of the sound wave, that is, find the wave number, angular frequency, and the initial phase shift?

Solution:

$$s_1 = 6.34 \text{ nm}$$

$$s_2 = 2.30 \text{ nm}$$

$$kx_1 + \phi = 0 \text{ rad}$$

$$kx_2 + \phi = 1.20 \text{ rad}$$

$$k(x_2 - x_1) = 1.20 \text{ rad}$$

$$k = 3.00 \text{ m}^{-1}$$

$$\omega = 1019.62 \text{ s}^{-1}$$

$$s_1 = s_{\max} \cos(kx_1 - \phi)$$

$$\phi = 5.66 \text{ rad}$$

$$s(x, t) = 6.30 \text{ nm} \cos(3.00 \text{ m}^{-1}x - 1019.62 \text{ s}^{-1}t + 5.66)$$

Exercise:

Problem:

An airplane moves at Mach 1.2 and produces a shock wave. (a) What is the speed of the plane in meters per second? (b) What is the angle that the shock wave moves?

Additional Problems

Exercise:

Problem:

A 0.80-m-long tube is opened at both ends. The air temperature is 26°C . The air in the tube is oscillated using a speaker attached to a signal generator. What are the wavelengths and frequencies of first two modes of sound waves that resonate in the tube?

Solution:

$$v_s = 346.40 \text{ m/s};$$

$$\lambda_n = \frac{2}{n} L \quad f_n = \frac{v_s}{\lambda_n}$$

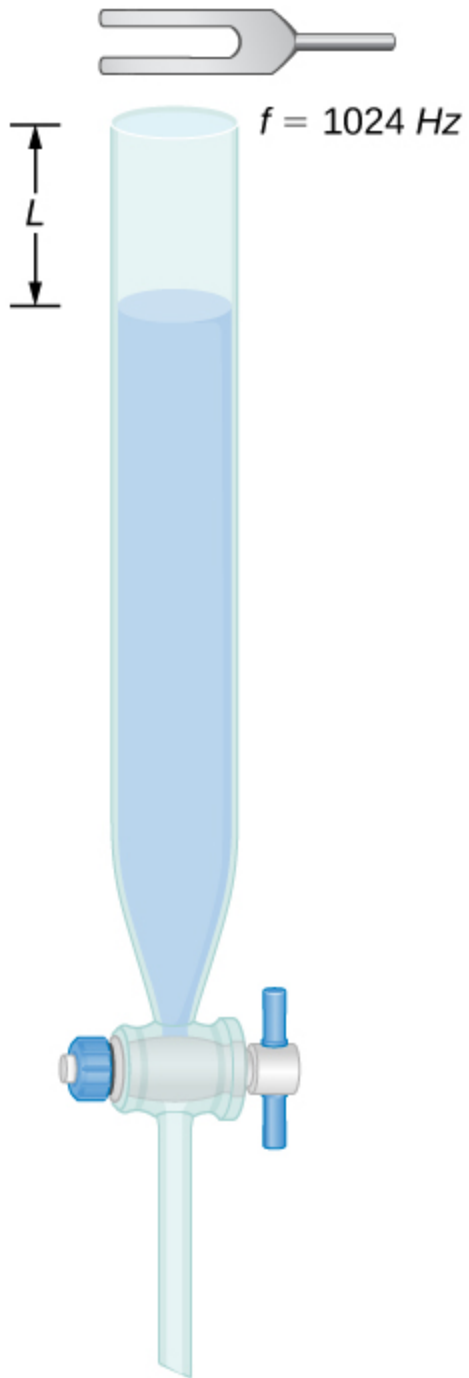
$$\lambda_1 = 1.60 \text{ m} \quad f_1 = 216.50 \text{ Hz}$$

$$\lambda_2 = 0.80 \text{ m} \quad f_1 = 433.00 \text{ Hz}$$

Exercise:

Problem:

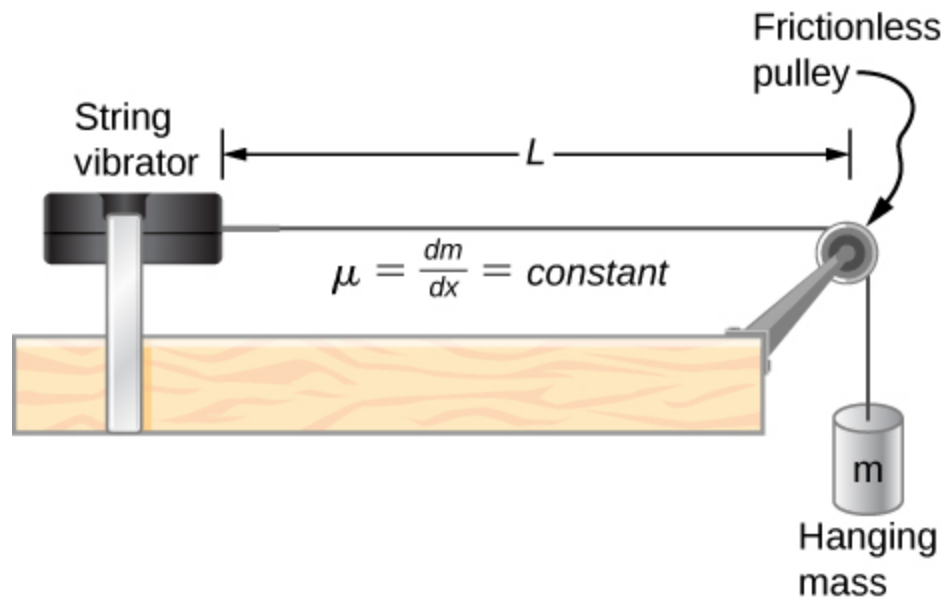
A tube filled with water has a valve at the bottom to allow the water to flow out of the tube. As the water is emptied from the tube, the length L of the air column changes. A 1024-Hz tuning fork is placed at the opening of the tube. Water is removed from the tube until the $n = 5$ mode of a sound wave resonates. What is the length of the air column if the temperature of the air in the room is 18°C ?



Exercise:

Problem:

Consider the following figure. The length of the string between the string vibrator and the pulley is $L = 1.00$ m. The linear density of the string is $\mu = 0.006$ kg/m. The string vibrator can oscillate at any frequency. The hanging mass is 2.00 kg. (a) What are the wavelength and frequency of $n = 6$ mode? (b) The string oscillates the air around the string. What is the wavelength of the sound if the speed of the sound is $v_s = 343.00$ m/s?



Solution:

$$\lambda_6 = 0.40 \text{ m}$$

a. $v = 57.15 \frac{\text{m}}{\text{s}}$; b. $\lambda_s = 2.40 \text{ m}$

$$f_6 = 142.89 \text{ Hz}$$

Exercise:

Problem:

Early Doppler shift experiments were conducted using a band playing music on a train. A trumpet player on a moving railroad flatcar plays a 320-Hz note. The sound waves heard by a stationary observer on a train platform hears a frequency of 350 Hz. What is the flatcar's speed in mph? The temperature of the air is $T_C = 22^\circ\text{C}$.

Exercise:**Problem:**

Two cars move toward one another, both sounding their horns ($f_s = 800\text{ Hz}$). Car A is moving at 65 mph and Car B is at 75 mph. What is the beat frequency heard by each driver? The air temperature is $T_C = 22.00^\circ\text{C}$.

Solution:

$$\begin{aligned}v &= 344.08 \frac{\text{m}}{\text{s}} \\v_A &= 29.05 \frac{\text{m}}{\text{s}}, \quad v_B = 33.52 \text{ m/s} \\f_A &= 961.18 \text{ Hz}, \\f_B &= 958.89 \text{ Hz} \\f_{A,\text{beat}} &= 161.18 \text{ Hz}, \quad f_{B,\text{beat}} = 158.89 \text{ Hz}\end{aligned}$$

Exercise:**Problem:**

Student A runs after Student B. Student A carries a tuning fork ringing at 1024 Hz, and student B carries a tuning fork ringing at 1000 Hz. Student A is running at a speed of $v_A = 5.00\text{ m/s}$ and Student B is running at $v_B = 6.00\text{ m/s}$. What is the beat frequency heard by each student? The speed of sound is $v = 343.00\text{ m/s}$.

Exercise:

Problem:

Suppose that the sound level from a source is 75 dB and then drops to 52 dB, with a frequency of 600 Hz. Determine the (a) initial and (b) final sound intensities and the (c) initial and (d) final sound wave amplitudes. The air temperature is $T_C = 24.00^\circ\text{C}$ and the air density is $\rho = 1.184\text{ kg/m}^3$.

Solution:

$$v = 345.24 \frac{\text{m}}{\text{s}}; \text{ a. } I = 31.62 \frac{\mu\text{W}}{\text{m}^2}; \text{ b. } I = 0.16 \frac{\mu\text{W}}{\text{m}^2}; \text{ c. } s_{\text{max}} = 104.39 \mu\text{m}; \text{ d. } s_{\text{max}} = 7.43 \mu\text{m}$$

Exercise:**Problem:**

The Doppler shift for a Doppler radar is found by $f = f_R \left(\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}} \right)$, where f_R is the frequency of the radar, f is the frequency observed by the radar, c is the speed of light, and v is the speed of the target. What is the beat frequency observed at the radar, assuming the speed of the target is much slower than the speed of light?

Exercise:**Problem:**

A stationary observer hears a frequency of 1000.00 Hz as a source approaches and a frequency of 850.00 Hz as a source departs. The source moves at a constant velocity of 75 mph. What is the temperature of the air?

Solution:

$$\frac{f_A}{f_D} = \frac{v + v_s}{v - v_s}, \quad (v - v_s) \frac{f_A}{f_D} = v + v_s, \quad v = 347.39 \frac{\text{m}}{\text{s}}$$

$$T_C = 27.70^\circ$$

Exercise:

Problem:

A flute plays a note with a frequency of 600 Hz. The flute can be modeled as a pipe open at both ends, where the flute player changes the length with his finger positions. What is the length of the tube if this is the fundamental frequency?

Challenge Problems**Exercise:****Problem:**

Two sound speakers are separated by a distance d , each sounding a frequency f . An observer stands at one speaker and walks in a straight line a distance x , perpendicular to the line between the two speakers, until he comes to the first maximum intensity of sound. The speed of sound is v . How far is he from the speaker?

Solution:

$$\sqrt{x^2 + d^2} - x = \lambda, \quad x^2 + d^2 = (\lambda + x)^2$$

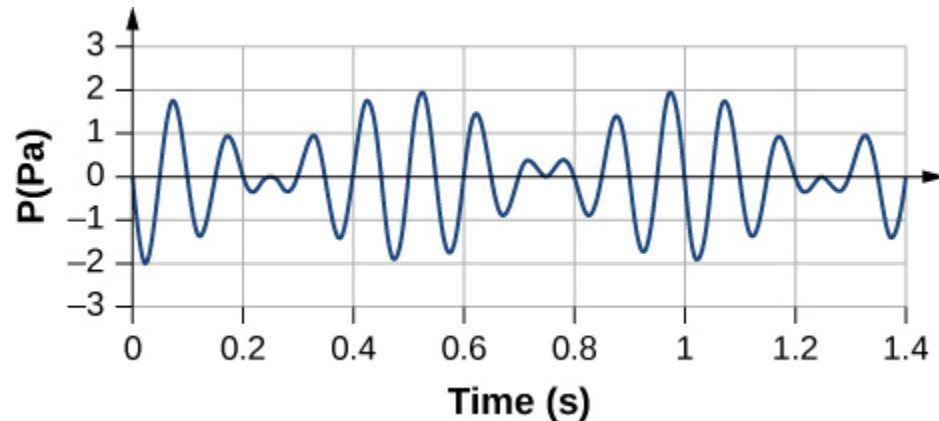
$$x^2 + d^2 = \lambda^2 + 2x\lambda + x^2, \quad d^2 = \lambda^2 + 2x\lambda$$

$$x = \frac{d^2 - \left(\frac{v}{f}\right)^2}{2\frac{v}{f}}$$

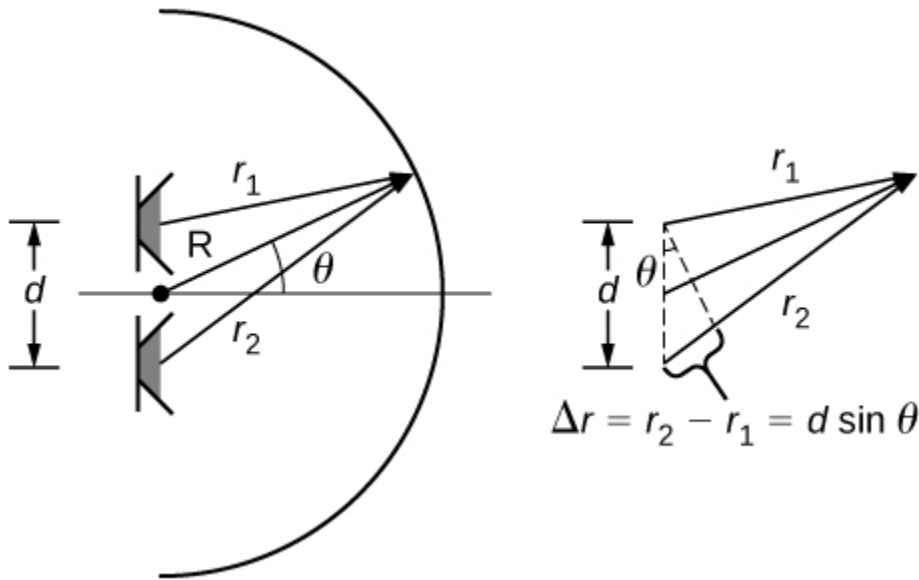
Exercise:

Problem:

Consider the beats shown below. This is a graph of the gauge pressure versus time for the position $x = 0.00$ m. The wave moves with a speed of $v = 343.00$ m/s. (a) How many beats are there per second? (b) How many times does the wave oscillate per second? (c) Write a wave function for the gauge pressure as a function of time.

**Exercise:****Problem:**

Two speakers producing the same frequency of sound are a distance of d apart. Consider an arc along a circle of radius R , centered at the midpoint of the speakers, as shown below. (a) At what angles will there be maxima? (b) At what angle will there be minima?



Solution:

a. For maxima

$$\Delta r = d \sin \theta$$

$$d \sin \theta = n\lambda \quad n = 0, \pm 1, \pm 2, \dots, \quad \theta = \sin^{-1} \left(n \frac{\lambda}{d} \right) \quad n = 0, \pm 1, \pm 2, \dots$$

$$\Delta r = d \sin \theta$$

b. For minima, $d \sin \theta = \left(n + \frac{1}{2} \right) \lambda \quad n = 0, \pm 1, \pm 2, \dots$

$$\theta = \sin^{-1} \left(\left(n + \frac{1}{2} \right) \frac{\lambda}{d} \right) \quad n = 0, \pm 1, \pm 2, \dots$$

Exercise:

Problem:

A string has a length of 1.5 m, a linear mass density $\mu = 0.008 \text{ kg/m}$, and a tension of 120 N. If the air temperature is $T = 22^\circ \text{C}$, what should the length of a pipe open at both ends for it to have the same frequency for the $n = 3$ mode?

Exercise:

Problem:

A string ($\mu = 0.006 \frac{\text{kg}}{\text{m}}$, $L = 1.50 \text{ m}$) is fixed at both ends and is under a tension of 155 N. It oscillates in the $n = 10$ mode and produces sound. A tuning fork is ringing nearby, producing a beat frequency of 23.76 Hz. (a) What is the frequency of the sound from the string? (b) What is the frequency of the tuning fork if the tuning fork frequency is lower? (c) What should be the tension of the string for the beat frequency to be zero?

Solution:

a. $v_{\text{string}} = 160.73 \frac{\text{m}}{\text{s}}$, $f_{\text{string}} = 535.77 \text{ Hz}$; b. $f_{\text{fork}} = 512 \text{ Hz}$; c.

$$f_{\text{fork}} = \frac{n\sqrt{\frac{F_T}{\mu}}}{2L}, \quad F_T = 141.56 \text{ N}$$

Exercise:**Problem:**

A string has a linear mass density μ , a length L , and a tension of F_T , and oscillates in a mode n at a frequency f . Find the ratio of $\frac{\Delta f}{f}$ for a small change in tension.

Exercise:**Problem:**

A string has a linear mass density $\mu = 0.007 \text{ kg/m}$, a length $L = 0.70 \text{ m}$, a tension of $F_T = 110 \text{ N}$, and oscillates in a mode $n = 3$. (a) What is the frequency of the oscillations? (b) Use the result in the preceding problem to find the change in the frequency when the tension is increased by 1.00%.

Solution:

a. $f = 268.62 \text{ Hz}$; b. $\Delta f \approx \frac{1}{2} \frac{\Delta F_T}{F_T} f = 1.34 \text{ Hz}$

Exercise:**Problem:**

A speaker powered by a signal generator is used to study resonance in a tube. The signal generator can be adjusted from a frequency of 1000 Hz to 1800 Hz. First, a 0.75-m-long tube, open at both ends, is studied. The temperature in the room is $T_F = 85.00^\circ\text{F}$. (a) Which normal modes of the pipe can be studied? What are the frequencies and wavelengths? Next a cap is placed on one end of the 0.75-meter-long pipe. (b) Which normal modes of the pipe can be studied? What are the frequencies and wavelengths?

Exercise:**Problem:**

A string on the violin has a length of 23.00 cm and a mass of 0.900 grams. The tension in the string is 850.00 N. The temperature in the room is $T_C = 24.00^\circ\text{C}$. The string is plucked and oscillates in the $n = 9$ mode. (a) What is the speed of the wave on the string? (b) What is the wavelength of the sounding wave produced? (c) What is the frequency of the oscillating string? (d) What is the frequency of the sound produced? (e) What is the wavelength of the sound produced?

Solution:

- a. $v = 466.07 \frac{\text{m}}{\text{s}}$; b. $\lambda_9 = 51.11 \text{ mm}$; c. $f_9 = 9.12 \text{ kHz}$;
d. $f_{\text{sound}} = 9.12 \text{ kHz}$; e. $\lambda_{\text{air}} = 37.86 \text{ mm}$

Glossary

bow wake

v-shaped disturbance created when the wave source moves faster than the wave propagation speed

shock wave

wave front that is produced when a sound source moves faster than the speed of sound

sonic boom

loud noise that occurs as a shock wave as it sweeps along the ground

Introduction

class="introduction"

These snowshoers on Mount Hood in Oregon are enjoying the heat flow and light caused by high temperature. All three mechanisms of heat transfer are relevant to this picture. The heat flowing out of the fire also turns the solid snow to liquid water and vapor. (credit: modification of work by “Mt. Hood Territory”/Flickr)



Heat and temperature are important concepts for each of us, every day. How we dress in the morning depends on whether the day is hot or cold, and most of what we do requires energy that ultimately comes from the Sun. The study of heat and temperature is part of an area of physics known as thermodynamics. The laws of thermodynamics govern the flow of energy throughout the universe. They are studied in all areas of science and engineering, from chemistry to biology to environmental science.

In this chapter, we explore heat and temperature. It is not always easy to distinguish these terms. Heat is the flow of energy from one object to another. This flow of energy is caused by a difference in temperature. The transfer of heat can change temperature, as can work, another kind of energy transfer that is central to thermodynamics. We return to these basic ideas several times throughout the next four chapters, and you will see that they affect everything from the behavior of atoms and molecules to cooking to our weather on Earth to the life cycles of stars.

Temperature and Thermal Equilibrium

By the end of this section, you will be able to:

- Define temperature and describe it qualitatively
- Explain thermal equilibrium
- Explain the zeroth law of thermodynamics

Heat is familiar to all of us. We can feel heat entering our bodies from the summer Sun or from hot coffee or tea after a winter stroll. We can also feel heat leaving our bodies as we feel the chill of night or the cooling effect of sweat after exercise.

What is heat? How do we define it and how is it related to temperature? What are the effects of heat and how does it flow from place to place? We will find that, in spite of the richness of the phenomena, a small set of underlying physical principles unites these subjects and ties them to other fields. We start by examining temperature and how to define and measure it.

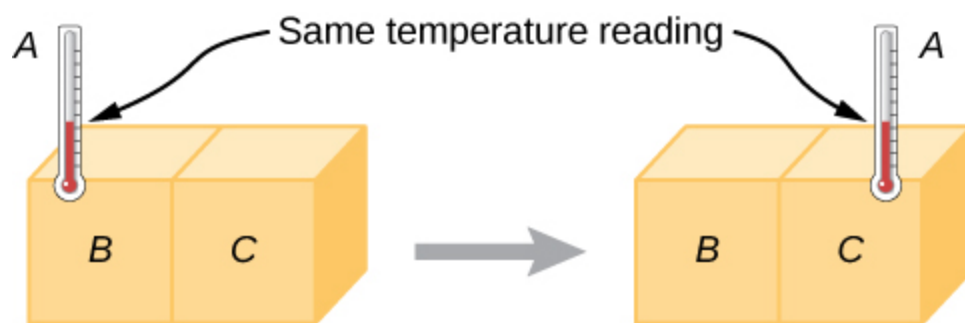
Temperature

The concept of temperature has evolved from the common concepts of hot and cold. The scientific definition of temperature explains more than our senses of hot and cold. As you may have already learned, many physical quantities are defined solely in terms of how they are observed or measured, that is, they are defined *operationally*. **Temperature** is operationally defined as the quantity of what we measure with a thermometer. As we will see in detail in a later chapter on the kinetic theory of gases, temperature is proportional to the average kinetic energy of translation, a fact that provides a more physical definition. Differences in temperature maintain the transfer of heat, or *heat transfer*, throughout the universe. **Heat transfer** is the movement of energy from one place or material to another as a result of a difference in temperature. (You will learn more about heat transfer later in this chapter.)

Thermal Equilibrium

An important concept related to temperature is **thermal equilibrium**. Two objects are in thermal equilibrium if they are in close contact that allows either to gain energy from the other, but nevertheless, no net energy is transferred between them. Even when not in contact, they are in thermal equilibrium if, when they are placed in contact, no net energy is transferred between them. If two objects remain in contact for a long time, they typically come to equilibrium. In other words, two objects in thermal equilibrium do not exchange energy.

Experimentally, if object *A* is in equilibrium with object *B*, and object *B* is in equilibrium with object *C*, then (as you may have already guessed) object *A* is in equilibrium with object *C*. That statement of transitivity is called the **zeroth law of thermodynamics**. (The number “zeroth” was suggested by British physicist Ralph Fowler in the 1930s. The first, second, and third laws of thermodynamics were already named and numbered then. The zeroth law had seldom been stated, but it needs to be discussed before the others, so Fowler gave it a smaller number.) Consider the case where *A* is a thermometer. The zeroth law tells us that if *A* reads a certain temperature when in equilibrium with *B*, and it is then placed in contact with *C*, it will not exchange energy with *C*; therefore, its temperature reading will remain the same ([\[link\]](#)). In other words, *if two objects are in thermal equilibrium, they have the same temperature*.



If thermometer *A* is in thermal equilibrium with object *B*, and *B* is in thermal equilibrium with *C*, then *A* is in thermal equilibrium with *C*. Therefore, the reading on *A* stays the same when *A* is moved over to make contact with *C*.

A thermometer measures its own temperature. It is through the concepts of thermal equilibrium and the zeroth law of thermodynamics that we can say that a thermometer measures the temperature of *something else*, and to make sense of the statement that two objects are at the same temperature.

In the rest of this chapter, we will often refer to “systems” instead of “objects.” As in the chapter on linear momentum and collisions, a system consists of one or more objects—but in thermodynamics, we require a system to be macroscopic, that is, to consist of a huge number (such as 10^{23}) of molecules. Then we can say that a system is in thermal equilibrium with itself if all parts of it are at the same temperature. (We will return to the definition of a thermodynamic system in the chapter on the first law of thermodynamics.)

Summary

- Temperature is operationally defined as the quantity measured by a thermometer. It is proportional to the average kinetic energy of atoms and molecules in a system.
- Thermal equilibrium occurs when two bodies are in contact with each other and can freely exchange energy. Systems are in thermal equilibrium when they have the same temperature.
- The zeroth law of thermodynamics states that when two systems, *A* and *B*, are in thermal equilibrium with each other, and *B* is in thermal equilibrium with a third system *C*, then *A* is also in thermal equilibrium with *C*.

Conceptual Questions

Exercise:

Problem:

What does it mean to say that two systems are in thermal equilibrium?

Solution:

They are at the same temperature, and if they are placed in contact, no net heat flows between them.

Exercise:**Problem:**

Give an example in which A has some kind of non-thermal equilibrium relationship with B , and B has the same relationship with C , but A does not have that relationship with C .

Glossary**heat transfer**

movement of energy from one place or material to another as a result of a difference in temperature

temperature

quantity measured by a thermometer, which reflects the mechanical energy of molecules in a system

thermal equilibrium

condition in which heat no longer flows between two objects that are in contact; the two objects have the same temperature

zeroth law of thermodynamics

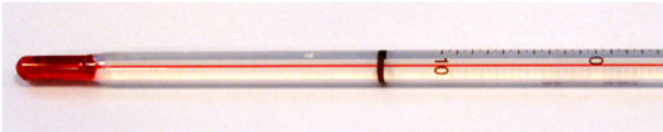
law that states that if two objects are in thermal equilibrium, and a third object is in thermal equilibrium with one of those objects, it is also in thermal equilibrium with the other object

Thermometers and Temperature Scales

By the end of this section, you will be able to:

- Describe several different types of thermometers
- Convert temperatures between the Celsius, Fahrenheit, and Kelvin scales

Any physical property that depends consistently and reproducibly on temperature can be used as the basis of a thermometer. For example, volume increases with temperature for most substances. This property is the basis for the common alcohol thermometer and the original mercury thermometers. Other properties used to measure temperature include electrical resistance, color, and the emission of infrared radiation ([\[link\]](#)).



(a)



(b)



(c)

Because many physical properties depend on temperature, the variety of thermometers is remarkable. (a) In this common type of thermometer, the alcohol, containing a red dye, expands more rapidly than the glass encasing it. When the thermometer's temperature increases, the liquid from the bulb is forced into the narrow tube,

producing a large change in the length of the column for a small change in temperature. (b) Each of the six squares on this plastic (liquid crystal) thermometer contains a film of a different heat-sensitive liquid crystal material. Below 95 °F, all six squares are black. When the plastic thermometer is exposed to a temperature of 95 °F, the first liquid crystal square changes color. When the temperature reaches above 96.8 °F, the second liquid crystal square also changes color, and so forth. (c) A firefighter uses a pyrometer to check the temperature of an aircraft carrier's ventilation system. The pyrometer measures infrared radiation (whose emission varies with temperature) from the vent and quickly produces a temperature readout. Infrared thermometers are also frequently used to measure body temperature by gently placing them in the ear canal. Such thermometers are more accurate than the alcohol thermometers placed under the tongue or in the armpit. (credit b: modification of work by Tess Watson; credit c: modification of work by Lamel J. Hinton, U.S. Navy)

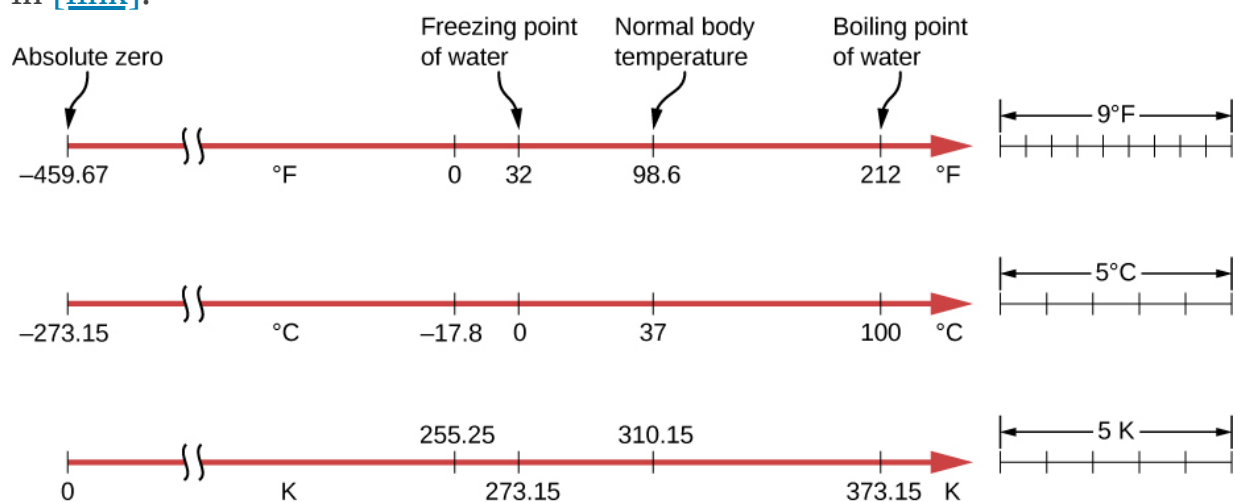
Thermometers measure temperature according to well-defined scales of measurement. The three most common temperature scales are Fahrenheit, Celsius, and Kelvin. Temperature scales are created by identifying two reproducible temperatures. The freezing and boiling temperatures of water at standard atmospheric pressure are commonly used.

On the **Celsius scale**, the freezing point of water is 0 °C and the boiling point is 100 °C. The unit of temperature on this scale is the **degree Celsius** (°C). The **Fahrenheit scale** (still the most frequently used for common purposes in the United States) has the freezing point of water at 32 °F and the boiling point at 212 °F. Its unit is the **degree Fahrenheit** (°F). You can see that 100 Celsius degrees span the same range as 180 Fahrenheit degrees. Thus, a temperature difference of one degree on the Celsius scale is 1.8 times as large as a difference of one degree on the Fahrenheit scale, or $\Delta T_F = \frac{9}{5} \Delta T_C$.

The definition of temperature in terms of molecular motion suggests that there should be a lowest possible temperature, where the average kinetic energy of molecules is zero (or the minimum allowed by quantum mechanics). Experiments confirm the existence of such a temperature, called **absolute zero**. An **absolute temperature scale** is one whose zero point is absolute zero. Such scales are convenient in science because several physical quantities, such as the volume of an ideal gas, are directly related to absolute temperature.

The **Kelvin scale** is the absolute temperature scale that is commonly used in science. The SI temperature unit is the *kelvin*, which is abbreviated K (not accompanied by a degree sign). Thus 0 K is absolute zero. The freezing and boiling points of water are 273.15 K and 373.15 K, respectively. Therefore, temperature differences are the same in units of kelvins and degrees Celsius, or $\Delta T_C = \Delta T_K$.

The relationships between the three common temperature scales are shown in [\[link\]](#). Temperatures on these scales can be converted using the equations in [\[link\]](#).



Relationships between the Fahrenheit, Celsius, and Kelvin temperature scales are shown. The relative sizes of the scales are also shown.

To convert from...	Use this equation...
Celsius to Fahrenheit	$T_F = \frac{9}{5}T_C + 32$
Fahrenheit to Celsius	$T_C = \frac{5}{9}(T_F - 32)$
Celsius to Kelvin	$T_K = T_C + 273.15$
Kelvin to Celsius	$T_C = T_K - 273.15$
Fahrenheit to Kelvin	$T_K = \frac{5}{9}(T_F - 32) + 273.15$
Kelvin to Fahrenheit	$T_F = \frac{9}{5}(T_K - 273.15) + 32$

Temperature Conversions

To convert between Fahrenheit and Kelvin, convert to Celsius as an intermediate step.

Example:

Converting between Temperature Scales: Room Temperature

“Room temperature” is generally defined in physics to be 25 °C. (a) What is room temperature in °F? (b) What is it in K?

Strategy

To answer these questions, all we need to do is choose the correct conversion equations and substitute the known values.

Solution

To convert from °C to °F, use the equation

Equation:

$$T_F = \frac{9}{5}T_C + 32.$$

Substitute the known value into the equation and solve:

Equation:

$$T_F = \frac{9}{5}(25\text{ }^\circ\text{C}) + 32 = 77\text{ }^\circ\text{F}.$$

Similarly, we find that $T_K = T_C + 273.15 = 298\text{ K}$.

The Kelvin scale is part of the SI system of units, so its actual definition is more complicated than the one given above. First, it is not defined in terms of the freezing and boiling points of water, but in terms of the **triple point**. The triple point is the unique combination of temperature and pressure at which ice, liquid water, and water vapor can coexist stably. As will be discussed in the section on phase changes, the coexistence is achieved by lowering the pressure and consequently the boiling point to reach the freezing point. The triple-point temperature is defined as 273.16 K. This definition has the advantage that although the freezing temperature and boiling temperature of water depend on pressure, there is only one triple-point temperature.

Second, even with two points on the scale defined, different thermometers give somewhat different results for other temperatures. Therefore, a standard thermometer is required. Metrologists (experts in the science of measurement) have chosen the *constant-volume gas thermometer* for this purpose. A vessel of constant volume filled with gas is subjected to temperature changes, and the measured temperature is proportional to the change in pressure. Using “TP” to represent the triple point,

Equation:

$$T = \frac{p}{p_{\text{TP}}} T_{\text{TP}}.$$

The results depend somewhat on the choice of gas, but the less dense the gas in the bulb, the better the results for different gases agree. If the results are extrapolated to zero density, the results agree quite well, with zero pressure corresponding to a temperature of absolute zero.

Constant-volume gas thermometers are big and come to equilibrium slowly, so they are used mostly as standards to calibrate other thermometers.

Note:

Visit this [site](#) to learn more about the constant-volume gas thermometer.

Summary

- Three types of thermometers are alcohol, liquid crystal, and infrared radiation (pyrometer).
- The three main temperature scales are Celsius, Fahrenheit, and Kelvin. Temperatures can be converted from one scale to another using temperature conversion equations.
- The three phases of water (ice, liquid water, and water vapor) can coexist at a single pressure and temperature known as the triple point.

Conceptual Questions

Exercise:

Problem:

If a thermometer is allowed to come to equilibrium with the air, and a glass of water is not in equilibrium with the air, what will happen to the thermometer reading when it is placed in the water?

Solution:

The reading will change.

Exercise:

Problem:

Give an example of a physical property that varies with temperature and describe how it is used to measure temperature.

Problems**Exercise:****Problem:**

While traveling outside the United States, you feel sick. A companion gets you a thermometer, which says your temperature is 39. What scale is that on? What is your Fahrenheit temperature? Should you seek medical help?

Solution:

That must be Celsius. Your Fahrenheit temperature is 102 °F. Yes, it is time to get treatment.

Exercise:

Problem: What are the following temperatures on the Kelvin scale?

- (a) 68.0 °F, an indoor temperature sometimes recommended for energy conservation in winter
- (b) 134 °F, one of the highest atmospheric temperatures ever recorded on Earth (Death Valley, California, 1913)
- (c) 9890 °F, the temperature of the surface of the Sun

Exercise:

Problem:

(a) Suppose a cold front blows into your locale and drops the temperature by 40.0 Fahrenheit degrees. How many degrees Celsius does the temperature decrease when it decreases by 40.0 °F? (b) Show that any change in temperature in Fahrenheit degrees is nine-fifths the change in Celsius degrees

Solution:

a. $\Delta T_C = 22.2^\circ\text{C}$; b. We know that $\Delta T_F = T_{F2} - T_{F1}$. We also know that $T_{F2} = \frac{9}{5}T_{C2} + 32$ and $T_{F1} = \frac{9}{5}T_{C1} + 32$. So, substituting, we have $\Delta T_F = \left(\frac{9}{5}T_{C2} + 32\right) - \left(\frac{9}{5}T_{C1} + 32\right)$. Partially solving and rearranging the equation, we have $\Delta T_F = \frac{9}{5}(T_{C2} - T_{C1})$. Therefore, $\Delta T_F = \frac{9}{5}\Delta T_C$.

Exercise:**Problem:**

An Associated Press article on climate change said, “Some of the ice shelf’s disappearance was probably during times when the planet was 36 degrees Fahrenheit (2 degrees Celsius) to 37 degrees Fahrenheit (3 degrees Celsius) warmer than it is today.” What mistake did the reporter make?

Exercise:**Problem:**

(a) At what temperature do the Fahrenheit and Celsius scales have the same numerical value? (b) At what temperature do the Fahrenheit and Kelvin scales have the same numerical value?

Solution:

a. -40° ; b. 575 K

Exercise:

Problem:

A person taking a reading of the temperature in a freezer in Celsius makes two mistakes: first omitting the negative sign and then thinking the temperature is Fahrenheit. That is, the person reads $-x\text{ }^{\circ}\text{C}$ as $x\text{ }^{\circ}\text{F}$. Oddly enough, the result is the correct Fahrenheit temperature. What is the original Celsius reading? Round your answer to three significant figures.

Glossary

absolute temperature scale

scale, such as Kelvin, with a zero point that is absolute zero

absolute zero

temperature at which the average kinetic energy of molecules is zero

Celsius scale

temperature scale in which the freezing point of water is $0\text{ }^{\circ}\text{C}$ and the boiling point of water is $100\text{ }^{\circ}\text{C}$

degree Celsius

($^{\circ}\text{C}$) unit on the Celsius temperature scale

degree Fahrenheit

($^{\circ}\text{F}$) unit on the Fahrenheit temperature scale

Fahrenheit scale

temperature scale in which the freezing point of water is $32\text{ }^{\circ}\text{F}$ and the boiling point of water is $212\text{ }^{\circ}\text{F}$

Kelvin scale (K)

temperature scale in which 0 K is the lowest possible temperature, representing absolute zero

triple point

pressure and temperature at which a substance exists in equilibrium as a solid, liquid, and gas

Thermal Expansion

By the end of this section, you will be able to:

- Answer qualitative questions about the effects of thermal expansion
- Solve problems involving thermal expansion, including those involving thermal stress

The expansion of alcohol in a thermometer is one of many commonly encountered examples of **thermal expansion**, which is the change in size or volume of a given system as its temperature changes. The most visible example is the expansion of hot air. When air is heated, it expands and becomes less dense than the surrounding air, which then exerts an (upward) force on the hot air and makes steam and smoke rise, hot air balloons float, and so forth. The same behavior happens in all liquids and gases, driving natural heat transfer upward in homes, oceans, and weather systems, as we will discuss in an upcoming section. Solids also undergo thermal expansion. Railroad tracks and bridges, for example, have expansion joints to allow them to freely expand and contract with temperature changes, as shown in [\[link\]](#).



(a)



(b)

(a) Thermal expansion joints like these in the (b) Auckland Harbour Bridge in New Zealand allow bridges to change length without buckling. (credit: modification of works by “ŠJů”/Wikimedia Commons)

What is the underlying cause of thermal expansion? As previously mentioned, an increase in temperature means an increase in the kinetic energy of individual atoms. In a solid, unlike in a gas, the molecules are held in place by forces from neighboring molecules; as we saw in [Oscillations](#), the forces can be modeled as in harmonic springs described by the Lennard-Jones potential. [Energy in Simple Harmonic Motion](#) shows that such potentials are asymmetrical in that the potential energy increases more steeply when the molecules get closer to each other than when they get farther away. Thus, at a given kinetic energy, the distance moved is greater when neighbors move away from each other than when they move toward each other. The result is that increased kinetic energy (increased temperature) increases the average distance between molecules—the substance expands.

For most substances under ordinary conditions, it is an excellent approximation that there is no preferred direction (that is, the solid is “isotropic”), and an increase in temperature increases the solid’s size by a

certain fraction in each dimension. Therefore, if the solid is free to expand or contract, its proportions stay the same; only its overall size changes.

Note:

Linear Thermal Expansion

According to experiments, the dependence of thermal expansion on temperature, substance, and original initial length is summarized in the equation

Equation:

$$\frac{dL}{dT} = \alpha L$$

where $\frac{dL}{dT}$ is the instantaneous change in length per temperature, L is the length, and α is the **coefficient of linear expansion**, a material property that varies slightly with temperature. As α is nearly constant and also very small, for practical purposes, we use the linear approximation:

Equation:

$$\Delta L = \alpha L \Delta T$$

where ΔL is the change in length and ΔT is the change in temperature.

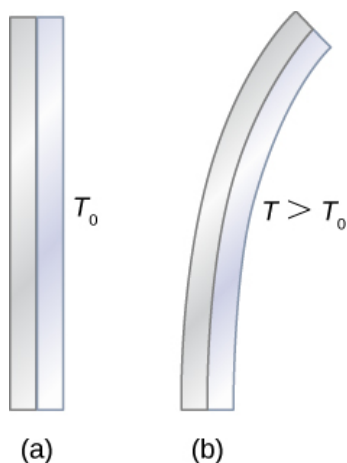
[\[link\]](#) lists representative values of the coefficient of linear expansion. As noted earlier, ΔT is the same whether it is expressed in units of degrees Celsius or kelvins; thus, α may have units of $1/^{\circ}\text{C}$ or $1/\text{K}$ with the same value in either case. Approximating α as a constant is quite accurate for small changes in temperature and sufficient for most practical purposes, even for large changes in temperature. We examine this approximation more closely in the next example.

Material	Coefficient of Linear Expansion α ($1/^{\circ}\text{C}$)	Coefficient of Volume Expansion β ($1/^{\circ}\text{C}$)
<i>Solids</i>		
Aluminum	25×10^{-6}	75×10^{-6}
Brass	19×10^{-6}	56×10^{-6}
Copper	17×10^{-6}	51×10^{-6}
Gold	14×10^{-6}	42×10^{-6}
Iron or steel	12×10^{-6}	35×10^{-6}
Invar (nickel-iron alloy)	0.9×10^{-6}	2.7×10^{-6}

Material	Coefficient of Linear Expansion α ($1/^{\circ}\text{C}$)	Coefficient of Volume Expansion β ($1/^{\circ}\text{C}$)
Lead	29×10^{-6}	87×10^{-6}
Silver	18×10^{-6}	54×10^{-6}
Glass (ordinary)	9×10^{-6}	27×10^{-6}
Glass (Pyrex®)	3×10^{-6}	9×10^{-6}
Quartz	0.4×10^{-6}	1×10^{-6}
Concrete, brick	$\sim 12 \times 10^{-6}$	$\sim 36 \times 10^{-6}$
Marble (average)	2.5×10^{-6}	7.5×10^{-6}
<i>Liquids</i>		
Ether		1650×10^{-6}
Ethyl alcohol		1100×10^{-6}
Gasoline		950×10^{-6}
Glycerin		500×10^{-6}
Mercury		180×10^{-6}
Water		210×10^{-6}
<i>Gases</i>		
Air and most other gases at atmospheric pressure		3400×10^{-6}

Thermal Expansion Coefficients

Thermal expansion is exploited in the bimetallic strip ([link](#)). This device can be used as a thermometer if the curving strip is attached to a pointer on a scale. It can also be used to automatically close or open a switch at a certain temperature, as in older or analog thermostats.



The curvature of a bimetallic strip depends on temperature. (a) The strip is straight at the starting temperature, where its two components have the same length. (b) At a higher temperature, this strip bends to the right, because the metal on the left has expanded more than the metal on the right. At a lower temperature, the strip would bend to the left.

Example:

Calculating Linear Thermal Expansion

The main span of San Francisco's Golden Gate Bridge is 1275 m long at its coldest. The bridge is exposed to temperatures ranging from -15°C to 40°C . What is its change in length between these temperatures? Assume that the bridge is made entirely of steel.

Strategy

Use the equation for linear thermal expansion $\Delta L = \alpha L \Delta T$ to calculate the change in length, ΔL . Use the coefficient of linear expansion α for steel from [\[link\]](#), and note that the change in temperature ΔT is 55°C .

Solution

Substitute all of the known values into the equation to solve for ΔL :

Equation:

$$\Delta L = \alpha L \Delta T = \left(\frac{12 \times 10^{-6}}{^{\circ}\text{C}} \right) (1275 \text{ m}) (55^{\circ}\text{C}) = 0.84 \text{ m}.$$

Significance

Although not large compared with the length of the bridge, this change in length is observable. It is generally spread over many expansion joints so that the expansion at each joint is small.

Thermal Expansion in Two and Three Dimensions

Unconstrained objects expand in all dimensions, as illustrated in [\[link\]](#). That is, their areas and volumes, as well as their lengths, increase with temperature. Because the proportions stay the same, holes and container volumes also get larger with temperature. If you cut a hole in a metal plate, the remaining material will expand exactly as it would if the piece you removed were still in place. The piece would get bigger, so the hole must get bigger too.

Note:

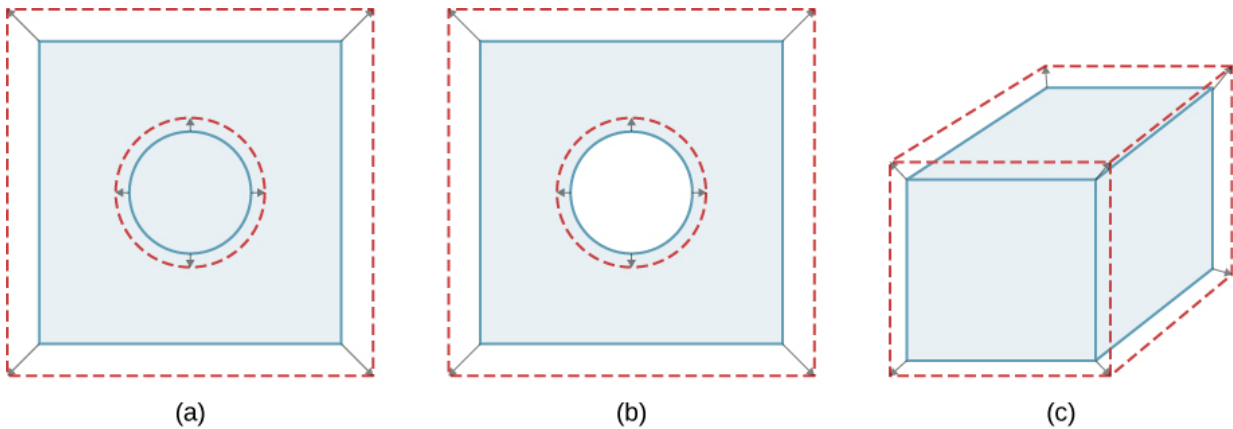
Thermal Expansion in Two Dimensions

For small temperature changes, the change in area ΔA is given by

Equation:

$$\Delta A = 2\alpha A \Delta T$$

where ΔA is the change in area A , ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature. (The derivation of this equation is analogous to that of the more important equation for three dimensions, below.)



In general, objects expand in all directions as temperature increases. In these drawings, the original boundaries of the objects are shown with solid lines, and the expanded boundaries with dashed lines.

(a) Area increases because both length and width increase. The area of a circular plug also increases.

(b) If the plug is removed, the hole it leaves becomes larger with increasing temperature, just as if the expanding plug were still in place. (c) Volume also increases, because all three dimensions increase.

Note:**Thermal Expansion in Three Dimensions**

The relationship between volume and temperature $\frac{dV}{dT}$ is given by $\frac{dV}{dT} = \beta V$, where β is the **coefficient of volume expansion**. As you can show in [\[link\]](#), $\beta = 3\alpha$. This equation is usually written as

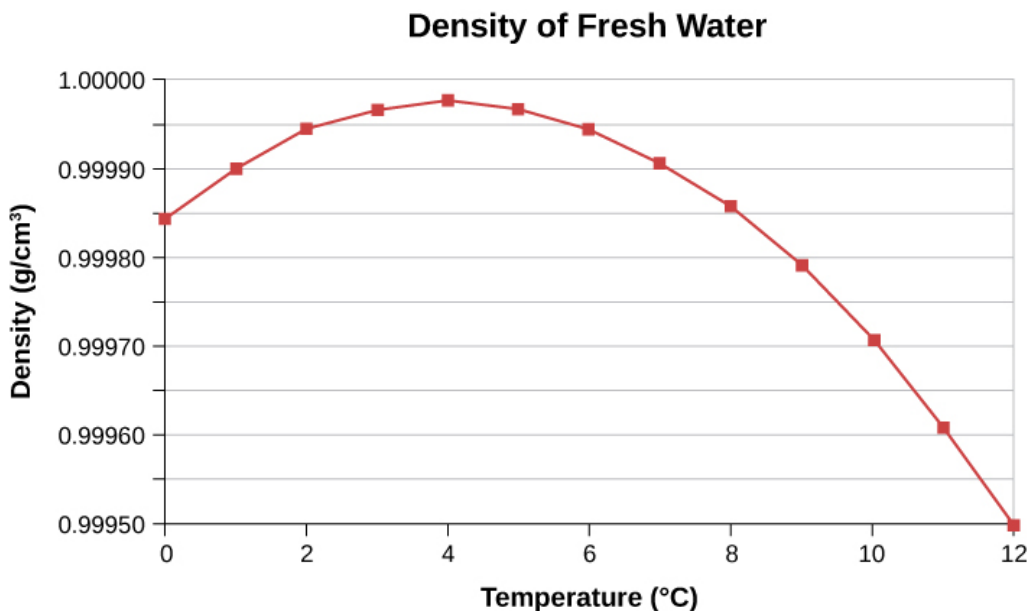
Equation:

$$\Delta V = \beta V \Delta T.$$

Note that the values of β in [\[link\]](#) are equal to 3α except for rounding.

Volume expansion is defined for liquids, but linear and area expansion are not, as a liquid's changes in linear dimensions and area depend on the shape of its container. Thus, [\[link\]](#) shows liquids' values of β but not α .

In general, objects expand with increasing temperature. Water is the most important exception to this rule. Water does expand with increasing temperature (its density *decreases*) at temperatures greater than 4 °C (40 °F). However, it is densest at +4 °C and expands with *decreasing* temperature between +4 °C and 0 °C (40 °F to 32 °F), as shown in [\[link\]](#). A striking effect of this phenomenon is the freezing of water in a pond. When water near the surface cools down to 4 °C, it is denser than the remaining water and thus sinks to the bottom. This “turnover” leaves a layer of warmer water near the surface, which is then cooled. However, if the temperature in the surface layer drops below 4 °C, that water is less dense than the water below, and thus stays near the top. As a result, the pond surface can freeze over. The layer of ice insulates the liquid water below it from low air temperatures. Fish and other aquatic life can survive in 4 °C water beneath ice, due to this unusual characteristic of water.



This curve shows the density of water as a function of temperature. Note that the thermal expansion at low temperatures is very small. The maximum density at 4 °C is only 0.0075 % greater than the density at 2 °C, and 0.012 % greater than that at 0 °C. The decrease of density below 4 °C occurs because the liquid water

approaches the solid crystal form of ice, which contains more empty space than the liquid.

Example:**Calculating Thermal Expansion**

Suppose your 60.0-L (15.9 -gal-gal) steel gasoline tank is full of gas that is cool because it has just been pumped from an underground reservoir. Now, both the tank and the gasoline have a temperature of 15.0 °C. How much gasoline has spilled by the time they warm to 35.0 °C?

Strategy

The tank and gasoline increase in volume, but the gasoline increases more, so the amount spilled is the difference in their volume changes. We can use the equation for volume expansion to calculate the change in volume of the gasoline and of the tank. (The gasoline tank can be treated as solid steel.)

Solution

1. Use the equation for volume expansion to calculate the increase in volume of the steel tank:

Equation:

$$\Delta V_s = \beta_s V_s \Delta T.$$

2. The increase in volume of the gasoline is given by this equation:

Equation:

$$\Delta V_{\text{gas}} = \beta_{\text{gas}} V_{\text{gas}} \Delta T.$$

3. Find the difference in volume to determine the amount spilled as

Equation:

$$V_{\text{spill}} = \Delta V_{\text{gas}} - \Delta V_s.$$

Alternatively, we can combine these three equations into a single equation. (Note that the original volumes are equal.)

Equation:

$$\begin{aligned} V_{\text{spill}} &= (\beta_{\text{gas}} - \beta_s) V \Delta T \\ &= [(950 - 35) \times 10^{-6} / ^\circ\text{C}] (60.0 \text{ L}) (20.0 ^\circ\text{C}) \\ &= 1.10 \text{ L.} \end{aligned}$$

Significance

This amount is significant, particularly for a 60.0-L tank. The effect is so striking because the gasoline and steel expand quickly. The rate of change in thermal properties is discussed later in this chapter. If you try to cap the tank tightly to prevent overflow, you will find that it leaks anyway, either around the cap or by bursting the tank. Tightly constricting the expanding gas is equivalent to compressing it, and both liquids and solids resist compression with extremely large forces. To avoid rupturing rigid containers, these containers have air gaps, which allow them to expand and contract without stressing them.

Note:**Exercise:****Problem:**

Check Your Understanding Does a given reading on a gasoline gauge indicate more gasoline in cold weather or in hot weather, or does the temperature not matter?

Solution:

The actual amount (mass) of gasoline left in the tank when the gauge hits “empty” is less in the summer than in the winter. The gasoline has the same volume as it does in the winter when the “add fuel” light goes on, but because the gasoline has expanded, there is less mass.

Thermal Stress

If you change the temperature of an object while preventing it from expanding or contracting, the object is subjected to stress that is compressive if the object would expand in the absence of constraint and tensile if it would contract. This stress resulting from temperature changes is known as **thermal stress**. It can be quite large and can cause damage.

To avoid this stress, engineers may design components so they can expand and contract freely. For instance, in highways, gaps are deliberately left between blocks to prevent thermal stress from developing. When no gaps can be left, engineers must consider thermal stress in their designs. Thus, the reinforcing rods in concrete are made of steel because steel’s coefficient of linear expansion is nearly equal to that of concrete.

To calculate the thermal stress in a rod whose ends are both fixed rigidly, we can think of the stress as developing in two steps. First, let the ends be free to expand (or contract) and find the expansion (or contraction). Second, find the stress necessary to compress (or extend) the rod to its original length by the methods you studied in [Static Equilibrium and Elasticity](#) on static equilibrium and elasticity. In other words, the ΔL of the thermal expansion equals the ΔL of the elastic distortion (except that the signs are opposite).

Example:**Calculating Thermal Stress**

Concrete blocks are laid out next to each other on a highway without any space between them, so they cannot expand. The construction crew did the work on a winter day when the temperature was 5°C . Find the stress in the blocks on a hot summer day when the temperature is 38°C . The compressive Young’s modulus of concrete is $Y = 20 \times 10^9 \text{ N/m}^2$.

Strategy

According to the chapter on static equilibrium and elasticity, the stress F/A is given by

Equation:

$$\frac{F}{A} = Y \frac{\Delta L}{L_0},$$

where Y is the Young’s modulus of the material—concrete, in this case. In thermal expansion, $\Delta L = \alpha L_0 \Delta T$. We combine these two equations by noting that the two ΔL /s are equal, as stated above.

Because we are not given L_0 or A , we can obtain a numerical answer only if they both cancel out.

Solution

We substitute the thermal-expansion equation into the elasticity equation to get

Equation:

$$\frac{F}{A} = Y \frac{\alpha L_0 \Delta T}{L_0} = Y \alpha \Delta T,$$

and as we hoped, L_0 has canceled and A appears only in F/A , the notation for the quantity we are calculating.

Now we need only insert the numbers:

Equation:

$$\frac{F}{A} = (20 \times 10^9 \text{ N/m}^2) (12 \times 10^{-6} / ^\circ\text{C}) (38 ^\circ\text{C} - 5 ^\circ\text{C}) = 7.9 \times 10^6 \text{ N/m}^2.$$

Significance

The ultimate compressive strength of concrete is $20 \times 10^6 \text{ N/m}^2$, so the blocks are unlikely to break. However, the ultimate shear strength of concrete is only $2 \times 10^6 \text{ N/m}^2$, so some might chip off.

Note:

Exercise:

Problem:

Check Your Understanding Two objects A and B have the same dimensions and are constrained identically. A is made of a material with a higher thermal expansion coefficient than B . If the objects are heated identically, will A feel a greater stress than B ?

Solution:

Not necessarily, as the thermal stress is also proportional to Young's modulus.

Summary

- Thermal expansion is the increase of the size (length, area, or volume) of a body due to a change in temperature, usually a rise. Thermal contraction is the decrease in size due to a change in temperature, usually a fall in temperature.
- Thermal stress is created when thermal expansion or contraction is constrained.

Conceptual Questions

Exercise:

Problem:

Pouring cold water into hot glass or ceramic cookware can easily break it. What causes the breaking? Explain why Pyrex®, a glass with a small coefficient of linear expansion, is less susceptible.

Solution:

The cold water cools part of the inner surface, making it contract, while the rest remains expanded. The strain is too great for the strength of the material. Pyrex contracts less, so it experiences less strain.

Exercise:**Problem:**

One method of getting a tight fit, say of a metal peg in a hole in a metal block, is to manufacture the peg slightly larger than the hole. The peg is then inserted when at a different temperature than the block. Should the block be hotter or colder than the peg during insertion? Explain your answer.

Exercise:**Problem:**

Does it really help to run hot water over a tight metal lid on a glass jar before trying to open it? Explain your answer.

Solution:

In principle, the lid expands more than the jar because metals have higher coefficients of expansion than glass. That should make unscrewing the lid easier. (In practice, getting the lid and jar wet may make gripping them more difficult.)

Exercise:**Problem:**

When a cold alcohol thermometer is placed in a hot liquid, the column of alcohol goes *down* slightly before going up. Explain why.

Exercise:**Problem:**

Calculate the length of a 1-meter rod of a material with thermal expansion coefficient α when the temperature is raised from 300 K to 600 K. Taking your answer as the new initial length, find the length after the rod is cooled back down to 300 K. Is your answer 1 meter? Should it be? How can you account for the result you got?

Solution:

After being heated, the length is $(1 + 300\alpha)$ (1 m). After being cooled, the length is $(1 - 300\alpha)(1 + 300\alpha)$ (1 m). That answer is not 1 m, but it should be. The explanation is that even if α is exactly constant, the relation $\Delta L = \alpha L \Delta T$ is strictly true only in the limit of small ΔT . Since α values are small, the discrepancy is unimportant in practice.

Exercise:

Problem:

Noting the large stresses that can be caused by thermal expansion, an amateur weapon inventor decides to use it to make a new kind of gun. He plans to jam a bullet against an aluminum rod inside a closed invar tube. When he heats the tube, the rod will expand more than the tube and a very strong force will build up. Then, by a method yet to be determined, he will open the tube in a split second and let the force of the rod launch the bullet at very high speed. What is he overlooking?

Problems**Exercise:****Problem:**

The height of the Washington Monument is measured to be 170.00 m on a day when the temperature is $35.0\text{ }^{\circ}\text{C}$. What will its height be on a day when the temperature falls to $-10.0\text{ }^{\circ}\text{C}$? Although the monument is made of limestone, assume that its coefficient of thermal expansion is the same as that of marble. Give your answer to five significant figures.

Solution:

Using [\[link\]](#) to find the coefficient of thermal expansion of marble:

$$L = L_0 + \Delta L = L_0 (1 + \alpha \Delta T) = 170\text{ m} [1 + (2.5 \times 10^{-6}/^{\circ}\text{C}) (-45.0\text{ }^{\circ}\text{C})] = 169.98\text{ m.}$$

(Answer rounded to five significant figures to show the slight difference in height.)

Exercise:**Problem:**

How much taller does the Eiffel Tower become at the end of a day when the temperature has increased by $15\text{ }^{\circ}\text{C}$? Its original height is 321 m and you can assume it is made of steel.

Exercise:**Problem:**

What is the change in length of a 3.00-cm-long column of mercury if its temperature changes from $37.0\text{ }^{\circ}\text{C}$ to $40.0\text{ }^{\circ}\text{C}$, assuming the mercury is constrained to a cylinder but unconstrained in length? Your answer will show why thermometers contain bulbs at the bottom instead of simple columns of liquid.

Solution:

We use β instead of α since this is a volume expansion with constant surface area. Therefore:

$$\Delta L = \alpha L \Delta T = (6.0 \times 10^{-5}/^{\circ}\text{C}) (0.0300\text{ m}) (3.00\text{ }^{\circ}\text{C}) = 5.4 \times 10^{-6}\text{ m.}$$
Exercise:**Problem:**

How large an expansion gap should be left between steel railroad rails if they may reach a maximum temperature $35.0\text{ }^{\circ}\text{C}$ greater than when they were laid? Their original length is 10.0 m.

Exercise:

Problem:

You are looking to buy a small piece of land in Hong Kong. The price is “only” \$60,000 per square meter. The land title says the dimensions are 20 m \times 30 m. By how much would the total price change if you measured the parcel with a steel tape measure on a day when the temperature was 20 °C above the temperature that the tape measure was designed for? The dimensions of the land do not change.

Solution:

On the warmer day, our tape measure will expand linearly. Therefore, each measured dimension will be smaller than the actual dimension of the land. Calling these measured dimensions l' and w' , we will find a new area, A . Let's calculate these measured dimensions:

$$l' = l_0 - \Delta l = (20 \text{ m}) - (20^\circ\text{C})(20 \text{ m}) \left(\frac{1.2 \times 10^{-5}}{^\circ\text{C}} \right) = 19.9952 \text{ m};$$

$$A' = l' \times w' = (29.9928 \text{ m})(19.9952 \text{ m}) = 599.71 \text{ m}^2;$$

$$\text{Cost change} = (A - A') \left(\frac{\$60,000}{\text{m}^2} \right) = ((600 - 599.71) \text{ m}^2) \left(\frac{\$60,000}{\text{m}^2} \right) = \$17,000.$$

Because the area gets smaller, the price of the land *decreases* by about \$17,000.

Exercise:**Problem:**

Global warming will produce rising sea levels partly due to melting ice caps and partly due to the expansion of water as average ocean temperatures rise. To get some idea of the size of this effect, calculate the change in length of a column of water 1.00 km high for a temperature increase of 1.00 °C. Assume the column is not free to expand sideways. As a model of the ocean, that is a reasonable approximation, as only parts of the ocean very close to the surface can expand sideways onto land, and only to a limited degree. As another approximation, neglect the fact that ocean warming is not uniform with depth.

Exercise:**Problem:**

(a) Suppose a meter stick made of steel and one made of aluminum are the same length at 0 °C. What is their difference in length at 22.0 °C? (b) Repeat the calculation for two 30.0-m-long surveyor's tapes.

Solution:

a. Use [\[link\]](#) to find the coefficients of thermal expansion of steel and aluminum. Then

$$\Delta L_{\text{Al}} - \Delta L_{\text{steel}} = (\alpha_{\text{Al}} - \alpha_{\text{steel}})L_0\Delta T = \left(\frac{2.5 \times 10^{-5}}{^\circ\text{C}} - \frac{1.2 \times 10^{-5}}{^\circ\text{C}} \right)(1.00 \text{ m})(22^\circ\text{C}) = 2.9 \times 10^{-4} \text{ m}$$

b. By the same method with $L_0 = 30.0 \text{ m}$, we have $\Delta L = 8.6 \times 10^{-3} \text{ m}$.

Exercise:**Problem:**

(a) If a 500-mL glass beaker is filled to the brim with ethyl alcohol at a temperature of 5.00 °C, how much will overflow when the alcohol's temperature reaches the room temperature of 22.0 °C? (b) How much less water would overflow under the same conditions?

Exercise:

Problem:

Most cars have a coolant reservoir to catch radiator fluid that may overflow when the engine is hot. A radiator is made of copper and is filled to its 16.0-L capacity when at 10.0 °C. What volume of radiator fluid will overflow when the radiator and fluid reach a temperature of 95.0 °C, given that the fluid's volume coefficient of expansion is $\beta = 400 \times 10^{-6}/^{\circ}\text{C}$? (Your answer will be a conservative estimate, as most car radiators have operating temperatures greater than 95.0 °C).

Solution:

$$\Delta V = 0.475 \text{ L}$$

Exercise:**Problem:**

A physicist makes a cup of instant coffee and notices that, as the coffee cools, its level drops 3.00 mm in the glass cup. Show that this decrease cannot be due to thermal contraction by calculating the decrease in level if the 350 cm³ of coffee is in a 7.00-cm-diameter cup and decreases in temperature from 95.0 °C to 45.0 °C. (Most of the drop in level is actually due to escaping bubbles of air.)

Exercise:**Problem:**

The density of water at 0 °C is very nearly 1000 kg/m³ (it is actually 999.84 kg/m³), whereas the density of ice at 0 °C is 917 kg/m³. Calculate the pressure necessary to keep ice from expanding when it freezes, neglecting the effect such a large pressure would have on the freezing temperature. (This problem gives you only an indication of how large the forces associated with freezing water might be.)

Solution:

If we start with the freezing of water, then it would expand to

$$(1 \text{ m}^3) \left(\frac{1000 \text{ kg/m}^3}{917 \text{ kg/m}^3} \right) = 1.09 \text{ m}^3 = 1.98 \times 10^8 \text{ N/m}^2 \text{ of ice.}$$
Exercise:**Problem:**

Show that $\beta = 3\alpha$, by calculating the infinitesimal change in volume dV of a cube with sides of length L when the temperature changes by dT .

Glossary

coefficient of linear expansion

(α) material property that gives the change in length, per unit length, per 1-°C change in temperature; a constant used in the calculation of linear expansion; the coefficient of linear expansion depends to some degree on the temperature of the material

coefficient of volume expansion

(β) similar to α but gives the change in volume, per unit volume, per 1-°C change in temperature

thermal expansion

change in size or volume of an object with change in temperature

thermal stress

stress caused by thermal expansion or contraction

Heat Transfer, Specific Heat, and Calorimetry

By the end of this section, you will be able to:

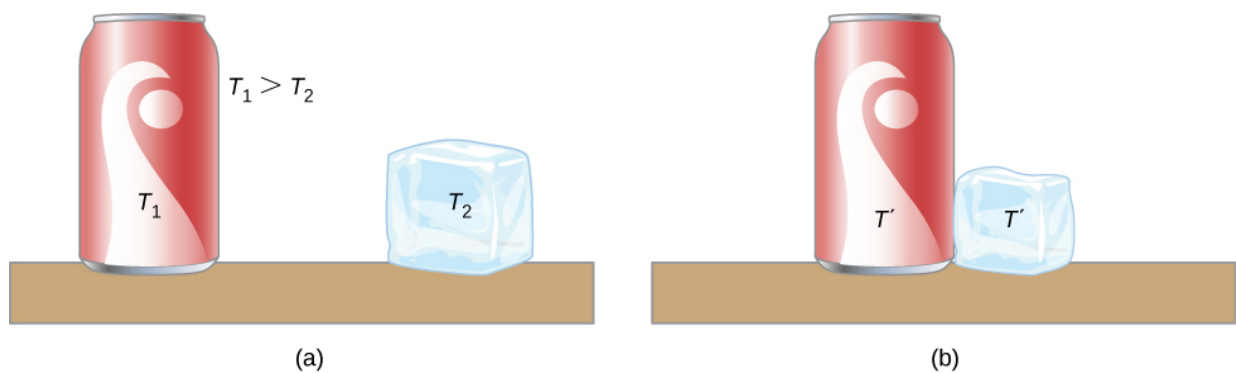
- Explain phenomena involving heat as a form of energy transfer
- Solve problems involving heat transfer

We have seen in previous chapters that energy is one of the fundamental concepts of physics. **Heat** is a type of energy transfer that is caused by a temperature difference, and it can change the temperature of an object. As we learned earlier in this chapter, heat transfer is the movement of energy from one place or material to another as a result of a difference in temperature. Heat transfer is fundamental to such everyday activities as home heating and cooking, as well as many industrial processes. It also forms a basis for the topics in the remainder of this chapter.

We also introduce the concept of internal energy, which can be increased or decreased by heat transfer. We discuss another way to change the internal energy of a system, namely doing work on it. Thus, we are beginning the study of the relationship of heat and work, which is the basis of engines and refrigerators and the central topic (and origin of the name) of thermodynamics.

Internal Energy and Heat

A thermal system has *internal energy* (also called thermal energy), which is the sum of the mechanical energies of its molecules. A system's internal energy is proportional to its temperature. As we saw earlier in this chapter, if two objects at different temperatures are brought into contact with each other, energy is transferred from the hotter to the colder object until the bodies reach thermal equilibrium (that is, they are at the same temperature). No work is done by either object because no force acts through a distance (as we discussed in [Work and Kinetic Energy](#)). These observations reveal that heat is energy transferred spontaneously due to a temperature difference. [\[link\]](#) shows an example of heat transfer.



(a) Here, the soft drink has a higher temperature than the ice, so they are not in thermal equilibrium. (b) When the soft drink and ice are allowed to interact, heat is transferred from the drink to the ice due to the difference in temperatures until they reach the same temperature, T' , achieving equilibrium. In fact, since the soft drink and ice are both in contact with the surrounding air and the bench, the ultimate equilibrium temperature will be the same as that of the surroundings.

The meaning of “heat” in physics is different from its ordinary meaning. For example, in conversation, we may say “the heat was unbearable,” but in physics, we would say that the temperature was high. Heat is a form of energy flow, whereas temperature is not. Incidentally, humans are sensitive to *heat flow* rather than to temperature.

Since heat is a form of energy, its SI unit is the joule (J). Another common unit of energy often used for heat is the **calorie** (cal), defined as the energy needed to change the temperature of 1.00 g of water by 1.00 °C —specifically, between 14.5 °C and 15.5 °C, since there is a slight temperature dependence. Also commonly used is the **kilocalorie** (kcal), which is the energy needed to change the temperature of 1.00 kg of water by 1.00 °C. Since mass is most often specified in kilograms, the kilocalorie is convenient. Confusingly, food calories (sometimes called “big calories,” abbreviated Cal) are actually kilocalories, a fact not easily determined from package labeling.

Mechanical Equivalent of Heat

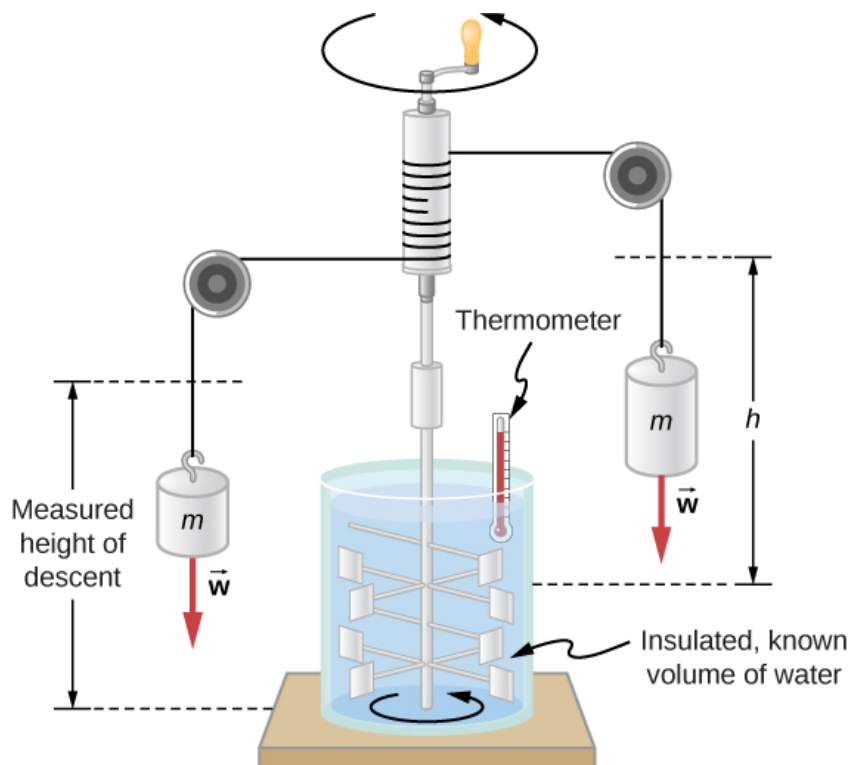
It is also possible to change the temperature of a substance by doing work, which transfers energy into or out of a system. This realization helped establish that heat is a form of energy. James Prescott Joule (1818–1889) performed many experiments to establish the **mechanical equivalent of heat**—*the work needed to produce the same effects as heat transfer*. In the units used for these two quantities, the value for this equivalence is

Equation:

$$1.000 \text{ kcal} = 4186 \text{ J.}$$

We consider this equation to represent the conversion between two units of energy. (Other numbers that you may see refer to calories defined for temperature ranges other than 14.5 °C to 15.5 °C.)

[\[link\]](#) shows one of Joule’s most famous experimental setups for demonstrating that work and heat can produce the same effects and measuring the mechanical equivalent of heat. It helped establish the principle of conservation of energy. Gravitational potential energy (U) was converted into kinetic energy (K), and then randomized by viscosity and turbulence into increased average kinetic energy of atoms and molecules in the system, producing a temperature increase. Joule’s contributions to thermodynamics were so significant that the SI unit of energy was named after him.



Joule's experiment established the equivalence of heat and work.

As the masses descended, they caused the paddles to do work, $W = mgh$, on the water. The result was a temperature increase, ΔT , measured by the thermometer. Joule found that ΔT was proportional to W and thus determined the mechanical equivalent of heat.

Increasing internal energy by heat transfer gives the same result as increasing it by doing work. Therefore, although a system has a well-defined internal energy, we cannot say that it has a certain "heat content" or "work content." A well-defined quantity that depends only on the current state of the system, rather than on the history of that system, is known as a *state variable*. Temperature and internal energy are state variables. To sum up this paragraph, *heat and work are not state variables*.

Incidentally, increasing the internal energy of a system does not necessarily increase its temperature. As we'll see in the next section, the temperature does not change when a substance changes from one phase to another. An example is the melting of ice, which can be accomplished by adding heat or by doing frictional work, as when an ice cube is rubbed against a rough surface.

Temperature Change and Heat Capacity

We have noted that heat transfer often causes temperature change. Experiments show that with no phase change and no work done on or by the system, the transferred heat is typically directly proportional to the change in temperature and to the mass of the system, to a good approximation. (Below we show how to handle situations where the approximation is not valid.) The constant of proportionality depends on the substance and its phase, which may be gas, liquid, or solid. We omit discussion of the fourth

phase, plasma, because although it is the most common phase in the universe, it is rare and short-lived on Earth.

We can understand the experimental facts by noting that the transferred heat is the change in the internal energy, which is the total energy of the molecules. Under typical conditions, the total kinetic energy of the molecules K_{total} is a constant fraction of the internal energy (for reasons and with exceptions that we'll see in the next chapter). The average kinetic energy of a molecule K_{ave} is proportional to the absolute temperature. Therefore, the change in internal energy of a system is typically proportional to the change in temperature and to the number of molecules, N . Mathematically, $\Delta U \propto \Delta K_{\text{total}} = NK_{\text{ave}} \propto N\Delta T$. The dependence on the substance results in large part from the different masses of atoms and molecules. We are considering its heat capacity in terms of its mass, but as we will see in the next chapter, in some cases, heat capacities *per molecule* are similar for different substances. The dependence on substance and phase also results from differences in the potential energy associated with interactions between atoms and molecules.

Note:

Heat Transfer and Temperature Change

A practical approximation for the relationship between heat transfer and temperature change is:

Equation:

$$Q = mc\Delta T,$$

where Q is the symbol for heat transfer (“quantity of heat”), m is the mass of the substance, and ΔT is the change in temperature. The symbol c stands for the **specific heat** (also called “*specific heat capacity*”) and depends on the material and phase. The specific heat is numerically equal to the amount of heat necessary to change the temperature of 1.00 kg of mass by 1.00 °C. The SI unit for specific heat is J/(kg × K) or J/(kg × °C). (Recall that the temperature change ΔT is the same in units of kelvin and degrees Celsius.)

Values of specific heat must generally be measured, because there is no simple way to calculate them precisely. [\[link\]](#) lists representative values of specific heat for various substances. We see from this table that the specific heat of water is five times that of glass and 10 times that of iron, which means that it takes five times as much heat to raise the temperature of water a given amount as for glass, and 10 times as much as for iron. In fact, water has one of the largest specific heats of any material, which is important for sustaining life on Earth.

The specific heats of gases depend on what is maintained constant during the heating—typically either the volume or the pressure. In the table, the first specific heat value for each gas is measured at constant volume, and the second (in parentheses) is measured at constant pressure. We will return to this topic in the chapter on the kinetic theory of gases.

Substances	Specific Heat (c)
------------	-------------------

Substances	Specific Heat (c)	
<i>Solids</i>	J/(kg · °C)	kcal/(kg · °C) ^[2]
Aluminum	900	0.215
Asbestos	800	0.19
Concrete, granite (average)	840	0.20
Copper	387	0.0924
Glass	840	0.20
Gold	129	0.0308
Human body (average at 37 °C)	3500	0.83
Ice (average, −50 °C to 0 °C)	2090	0.50
Iron, steel	452	0.108
Lead	128	0.0305
Silver	235	0.0562
Wood	1700	0.40
<i>Liquids</i>		
Benzene	1740	0.415
Ethanol	2450	0.586
Glycerin	2410	0.576
Mercury	139	0.0333
Water (15.0 °C)	4186	1.000
<i>Gases</i> ^[3]		
Air (dry)	721 (1015)	0.172 (0.242)
Ammonia	1670 (2190)	0.399 (0.523)
Carbon dioxide	638 (833)	0.152 (0.199)
Nitrogen	739 (1040)	0.177 (0.248)

Substances	Specific Heat (c)	
Oxygen	651 (913)	0.156 (0.218)
Steam (100 °C)	1520 (2020)	0.363 (0.482)

Specific Heats of Various Substances^{[1][1]}The values for solids and liquids are at constant volume and 25 °C, except as noted. ^[2]These values are identical in units of cal/g · °C. ^[3]Specific heats at constant volume and at 20.0 °C except as noted, and at 1.00 atm pressure. Values in parentheses are specific heats at a constant pressure of 1.00 atm.

In general, specific heat also depends on temperature. Thus, a precise definition of c for a substance must be given in terms of an infinitesimal change in temperature. To do this, we note that $c = \frac{1}{m} \frac{\Delta Q}{\Delta T}$ and replace Δ with d :

Equation:

$$c = \frac{1}{m} \frac{dQ}{dT}.$$

Except for gases, the temperature and volume dependence of the specific heat of most substances is weak at normal temperatures. Therefore, we will generally take specific heats to be constant at the values given in the table.

Example:

Calculating the Required Heat

A 0.500-kg aluminum pan on a stove and 0.250 L of water in it are heated from 20.0 °C to 80.0 °C.

(a) How much heat is required? What percentage of the heat is used to raise the temperature of (b) the pan and (c) the water?

Strategy

We can assume that the pan and the water are always at the same temperature. When you put the pan on the stove, the temperature of the water and that of the pan are increased by the same amount. We use the equation for the heat transfer for the given temperature change and mass of water and aluminum. The specific heat values for water and aluminum are given in [\[link\]](#).

Solution

1. Calculate the temperature difference:

Equation:

$$\Delta T = T_f - T_i = 60.0 \text{ °C}.$$

2. Calculate the mass of water. Because the density of water is 1000 kg/m³, 1 L of water has a mass of 1 kg, and the mass of 0.250 L of water is $m_w = 0.250$ kg.
3. Calculate the heat transferred to the water. Use the specific heat of water in [\[link\]](#):

Equation:

$$Q_w = m_w c_w \Delta T = (0.250 \text{ kg}) (4186 \text{ J/kg °C}) (60.0 \text{ °C}) = 62.8 \text{ kJ}.$$

4. Calculate the heat transferred to the aluminum. Use the specific heat for aluminum in [\[link\]](#):

Equation:

$$Q_{\text{Al}} = m_{\text{Al}} c_{\text{Al}} \Delta T = (0.500 \text{ kg}) (900 \text{ J/kg } ^\circ\text{C}) (60.0 ^\circ\text{C}) = 27.0 \text{ kJ}.$$

5. Find the total transferred heat:

Equation:

$$Q_{\text{Total}} = Q_{\text{W}} + Q_{\text{Al}} = 89.8 \text{ kJ}.$$

Significance

In this example, the heat transferred to the water is more than the aluminum pan. Although the mass of the pan is twice that of the water, the specific heat of water is over four times that of aluminum. Therefore, it takes a bit more than twice as much heat to achieve the given temperature change for the water as for the aluminum pan.

[\[link\]](#) illustrates a temperature rise caused by doing work. (The result is the same as if the same amount of energy had been added with a blowtorch instead of mechanically.)

Example:**Calculating the Temperature Increase from the Work Done on a Substance**

Truck brakes used to control speed on a downhill run do work, converting gravitational potential energy into increased internal energy (higher temperature) of the brake material ([\[link\]](#)). This conversion prevents the gravitational potential energy from being converted into kinetic energy of the truck. Since the mass of the truck is much greater than that of the brake material absorbing the energy, the temperature increase may occur too fast for sufficient heat to transfer from the brakes to the environment; in other words, the brakes may overheat.



The smoking brakes on a braking truck are visible evidence of the mechanical equivalent of heat.

Calculate the temperature increase of 10 kg of brake material with an average specific heat of $800 \text{ J/kg} \cdot ^\circ\text{C}$ if the material retains 10% of the energy from a 10,000-kg truck descending 75.0 m (in vertical displacement) at a constant speed.

Strategy

We calculate the gravitational potential energy (Mgh) that the entire truck loses in its descent, equate it to the increase in the brakes' internal energy, and then find the temperature increase produced in the brake material alone.

Solution

First we calculate the change in gravitational potential energy as the truck goes downhill:

Equation:

$$Mgh = (10,000 \text{ kg}) (9.80 \text{ m/s}^2) (75.0 \text{ m}) = 7.35 \times 10^6 \text{ J}.$$

Because the kinetic energy of the truck does not change, conservation of energy tells us the lost potential energy is dissipated, and we assume that 10% of it is transferred to internal energy of the brakes, so take $Q = Mgh/10$. Then we calculate the temperature change from the heat transferred, using

Equation:

$$\Delta T = \frac{Q}{mc},$$

where m is the mass of the brake material. Insert the given values to find

Equation:

$$\Delta T = \frac{7.35 \times 10^5 \text{ J}}{(10 \text{ kg})(800 \text{ J/kg} \cdot ^\circ\text{C})} = 92 ^\circ\text{C}.$$

Significance

If the truck had been traveling for some time, then just before the descent, the brake temperature would probably be higher than the ambient temperature. The temperature increase in the descent would likely raise the temperature of the brake material very high, so this technique is not practical. Instead, the truck would use the technique of engine braking. A different idea underlies the recent technology of hybrid and electric cars, where mechanical energy (kinetic and gravitational potential energy) is converted by the brakes into electrical energy in the battery, a process called regenerative braking.

In a common kind of problem, objects at different temperatures are placed in contact with each other but isolated from everything else, and they are allowed to come into equilibrium. A container that prevents heat transfer in or out is called a **calorimeter**, and the use of a calorimeter to make measurements (typically of heat or specific heat capacity) is called **calorimetry**.

We will use the term “calorimetry problem” to refer to any problem in which the objects concerned are thermally isolated from their surroundings. An important idea in solving calorimetry problems is that during a heat transfer between objects isolated from their surroundings, the heat gained by the colder object must equal the heat lost by the hotter object, due to conservation of energy:

Note:

Equation:

$$Q_{\text{cold}} + Q_{\text{hot}} = 0.$$

We express this idea by writing that the sum of the heats equals zero because the heat gained is usually considered positive; the heat lost, negative.

Example:**Calculating the Final Temperature in Calorimetry**

Suppose you pour 0.250 kg of 20.0-°C water (about a cup) into a 0.500-kg aluminum pan off the stove with a temperature of 150 °C. Assume no heat transfer takes place to anything else: The pan is placed on an insulated pad, and heat transfer to the air is neglected in the short time needed to reach equilibrium. Thus, this is a calorimetry problem, even though no isolating container is specified. Also assume that a negligible amount of water boils off. What is the temperature when the water and pan reach thermal equilibrium?

Strategy

Originally, the pan and water are not in thermal equilibrium: The pan is at a higher temperature than the water. Heat transfer restores thermal equilibrium once the water and pan are in contact; it stops once thermal equilibrium between the pan and the water is achieved. The heat lost by the pan is equal to the heat gained by the water—that is the basic principle of calorimetry.

Solution

1. Use the equation for heat transfer $Q = mc\Delta T$ to express the heat lost by the aluminum pan in terms of the mass of the pan, the specific heat of aluminum, the initial temperature of the pan, and the final temperature:

Equation:

$$Q_{\text{hot}} = m_{\text{Al}}c_{\text{Al}}(T_{\text{f}} - 150\text{ }^{\circ}\text{C}).$$

2. Express the heat gained by the water in terms of the mass of the water, the specific heat of water, the initial temperature of the water, and the final temperature:

Equation:

$$Q_{\text{cold}} = m_{\text{w}}c_{\text{w}}(T_{\text{f}} - 20.0\text{ }^{\circ}\text{C}).$$

3. Note that $Q_{\text{hot}} < 0$ and $Q_{\text{cold}} > 0$ and that as stated above, they must sum to zero:

Equation:

$$\begin{aligned} Q_{\text{cold}} + Q_{\text{hot}} &= 0 \\ Q_{\text{cold}} &= -Q_{\text{hot}} \\ m_{\text{w}}c_{\text{w}}(T_{\text{f}} - 20.0\text{ }^{\circ}\text{C}) &= -m_{\text{Al}}c_{\text{Al}}(T_{\text{f}} - 150\text{ }^{\circ}\text{C}). \end{aligned}$$

4. Bring all terms involving T_{f} on the left hand side and all other terms on the right hand side. Solving for T_{f} ,

Equation:

$$T_f = \frac{m_{A1}c_{A1}(150^\circ\text{C}) + m_w c_w (20.0^\circ\text{C})}{m_{A1}c_{A1} + m_w c_w},$$

and insert the numerical values:

Equation:

$$T_f = \frac{(0.500\text{ kg})(900\text{ J/kg}^\circ\text{C})(150^\circ\text{C}) + (0.250\text{ kg})(4186\text{ J/kg}^\circ\text{C})(20.0^\circ\text{C})}{(0.500\text{ kg})(900\text{ J/kg}^\circ\text{C}) + (0.250\text{ kg})(4186\text{ J/kg}^\circ\text{C})} = 59.1^\circ\text{C}.$$

Significance

Why is the final temperature so much closer to 20.0°C than to 150°C ? The reason is that water has a greater specific heat than most common substances and thus undergoes a smaller temperature change for a given heat transfer. A large body of water, such as a lake, requires a large amount of heat to increase its temperature appreciably. This explains why the temperature of a lake stays relatively constant during the day even when the temperature change of the air is large. However, the water temperature does change over longer times (e.g., summer to winter).

Note:

Exercise:

Problem:

Check Your Understanding If 25 kJ is necessary to raise the temperature of a rock from 25°C to 30°C , how much heat is necessary to heat the rock from 45°C to 50°C ?

Solution:

To a good approximation, the heat transfer depends only on the temperature difference. Since the temperature differences are the same in both cases, the same 25 kJ is necessary in the second case. (As we will see in the next section, the answer would have been different if the object had been made of some substance that changes phase anywhere between 30°C and 50°C .)

Example:

Temperature-Dependent Heat Capacity

At low temperatures, the specific heats of solids are typically proportional to T^3 . The first understanding of this behavior was due to the Dutch physicist Peter Debye, who in 1912, treated atomic oscillations with the quantum theory that Max Planck had recently used for radiation. For instance, a good approximation for the specific heat of salt, NaCl, is $c = 3.33 \times 10^4 \frac{\text{J}}{\text{kg}\cdot\text{K}} \left(\frac{T}{321\text{ K}}\right)^3$. The constant 321 K is called the *Debye temperature* of NaCl, Θ_D , and the formula works well when $T < 0.04\Theta_D$. Using this formula, how much heat is required to raise the temperature of 24.0 g of NaCl from 5 K to 15 K?

Solution

Because the heat capacity depends on the temperature, we need to use the equation

Equation:

$$c = \frac{1}{m} \frac{dQ}{dT}.$$

We solve this equation for Q by integrating both sides: $Q = m \int_{T_1}^{T_2} c dT$.

Then we substitute the given values in and evaluate the integral:

Equation:

$$Q = (0.024 \text{ kg}) \int_{T_1}^{T_2} 3.33 \times 10^{-6} \frac{\text{J}}{\text{kg} \cdot \text{K}} \left(\frac{T}{321 \text{ K}} \right)^3 dT = \left(6.04 \times 10^{-4} \frac{\text{J}}{\text{K}^4} \right) T^4 \bigg|_{5 \text{ K}}^{15 \text{ K}} = 0.302 \text{ J}.$$

Significance

If we had used the equation $Q = mc\Delta T$ and the room-temperature specific heat of salt, $880 \text{ J/kg} \cdot \text{K}$, we would have gotten a very different value.

Summary

- Heat and work are the two distinct methods of energy transfer.
- Heat transfer to an object when its temperature changes is often approximated well by $Q = mc\Delta T$, where m is the object's mass and c is the specific heat of the substance.

Conceptual Questions

Exercise:

Problem: How is heat transfer related to temperature?

Solution:

Temperature differences cause heat transfer.

Exercise:

Problem: Describe a situation in which heat transfer occurs.

Exercise:

Problem: When heat transfers into a system, is the energy stored as heat? Explain briefly.

Solution:

No, it is stored as thermal energy. A thermodynamic system does not have a well-defined quantity of heat.

Exercise:

Problem:

The brakes in a car increase in temperature by ΔT when bringing the car to rest from a speed v . How much greater would ΔT be if the car initially had twice the speed? You may assume the car stops fast enough that no heat transfers out of the brakes.

Problems**Exercise:****Problem:**

On a hot day, the temperature of an 80,000-L swimming pool increases by 1.50°C . What is the net heat transfer during this heating? Ignore any complications, such as loss of water by evaporation.

Solution:

$$m = 5.02 \times 10^8 \text{ J}$$

Exercise:**Problem:**

To sterilize a 50.0-g glass baby bottle, we must raise its temperature from 22.0°C to 95.0°C . How much heat transfer is required?

Exercise:**Problem:**

The same heat transfer into identical masses of different substances produces different temperature changes. Calculate the final temperature when 1.00 kcal of heat transfers into 1.00 kg of the following, originally at 20.0°C : (a) water; (b) concrete; (c) steel; and (d) mercury.

Solution:

$$Q = mc\Delta T \Rightarrow \Delta T = \frac{Q}{mc}; \text{ a. } 21.0^\circ\text{C}; \text{ b. } 25.0^\circ\text{C}; \text{ c. } 29.3^\circ\text{C}; \text{ d. } 50.0^\circ\text{C}$$

Exercise:**Problem:**

Rubbing your hands together warms them by converting work into thermal energy. If a woman rubs her hands back and forth for a total of 20 rubs, at a distance of 7.50 cm per rub, and with an average frictional force of 40.0 N, what is the temperature increase? The mass of tissues warmed is only 0.100 kg, mostly in the palms and fingers.

Exercise:**Problem:**

A 0.250-kg block of a pure material is heated from 20.0°C to 65.0°C by the addition of 4.35 kJ of energy. Calculate its specific heat and identify the substance of which it is most likely composed.

Solution:

$$Q = mc\Delta T \Rightarrow c = \frac{Q}{m\Delta T} = \frac{1.04 \text{ kcal}}{(0.250 \text{ kg})(45.0^\circ\text{C})} = 0.0924 \text{ kcal/kg} \cdot ^\circ\text{C}. \text{ It is copper.}$$

Exercise:**Problem:**

Suppose identical amounts of heat transfer into different masses of copper and water, causing identical changes in temperature. What is the ratio of the mass of copper to water?

Exercise:**Problem:**

(a) The number of kilocalories in food is determined by calorimetry techniques in which the food is burned and the amount of heat transfer is measured. How many kilocalories per gram are there in a 5.00-g peanut if the energy from burning it is transferred to 0.500 kg of water held in a 0.100-kg aluminum cup, causing a 54.9°C temperature increase? Assume the process takes place in an ideal calorimeter, in other words a perfectly insulated container. (b) Compare your answer to the following labeling information found on a package of dry roasted peanuts: a serving of 33 g contains 200 calories. Comment on whether the values are consistent.

Solution:

$$\begin{aligned} \text{a. } Q &= m_w c_w \Delta T + m_{A1} c_{A1} \Delta T = (m_w c_w + m_{A1} c_{A1}) \Delta T; \\ Q &= \left[(0.500 \text{ kg}) (1.00 \text{ kcal/kg} \cdot ^\circ\text{C}) + (0.100 \text{ kg}) (0.215 \text{ kcal/kg} \cdot ^\circ\text{C}) \right] (54.9^\circ\text{C}) = 28.63 \text{ kcal}; \\ \frac{Q}{m_p} &= \frac{28.63 \text{ kcal}}{5.00 \text{ g}} = 5.73 \text{ kcal/g}; \text{ b. } \frac{Q}{m_p} = \frac{200 \text{ kcal}}{33 \text{ g}} = 6 \text{ kcal/g, which is consistent with our results} \\ &\text{to part (a), to one significant figure.} \end{aligned}$$

Exercise:**Problem:**

Following vigorous exercise, the body temperature of an 80.0 kg person is 40.0°C . At what rate in watts must the person transfer thermal energy to reduce the body temperature to 37.0°C in 30.0 min, assuming the body continues to produce energy at the rate of 150 W? (1 watt = 1 joule/second or 1 W = 1 J/s)

Exercise:**Problem:**

In a study of healthy young men^[footnote], doing 20 push-ups in 1 minute burned an amount of energy per kg that for a 70.0-kg man corresponds to 8.06 calories (kcal). How much would a 70.0-kg man's temperature rise if he did not lose any heat during that time?

JW Vezina, "An examination of the differences between two methods of estimating energy expenditure in resistance training activities," *Journal of Strength and Conditioning Research*, April 28, 2014, <http://www.ncbi.nlm.nih.gov/pubmed/24402448>

Solution:

$$0.139^\circ\text{C}$$

Exercise:**Problem:**

A 1.28-kg sample of water at $10.0\text{ }^{\circ}\text{C}$ is in a calorimeter. You drop a piece of steel with a mass of 0.385 kg at $215\text{ }^{\circ}\text{C}$ into it. After the sizzling subsides, what is the final equilibrium temperature? (Make the reasonable assumptions that any steam produced condenses into liquid water during the process of equilibration and that the evaporation and condensation don't affect the outcome, as we'll see in the next section.)

Exercise:**Problem:**

Repeat the preceding problem, assuming the water is in a glass beaker with a mass of 0.200 kg, which in turn is in a calorimeter. The beaker is initially at the same temperature as the water. Before doing the problem, should the answer be higher or lower than the preceding answer? Comparing the mass and specific heat of the beaker to those of the water, do you think the beaker will make much difference?

Solution:

It should be lower. The beaker will not make much difference: $16.3\text{ }^{\circ}\text{C}$

Glossary

calorie (cal)

energy needed to change the temperature of 1.00 g of water by $1.00\text{ }^{\circ}\text{C}$

calorimeter

container that prevents heat transfer in or out

calorimetry

study of heat transfer inside a container impervious to heat

heat

energy transferred solely due to a temperature difference

kilocalorie (kcal)

energy needed to change the temperature of 1.00 kg of water between $14.5\text{ }^{\circ}\text{C}$ and $15.5\text{ }^{\circ}\text{C}$

mechanical equivalent of heat

work needed to produce the same effects as heat transfer

specific heat

amount of heat necessary to change the temperature of 1.00 kg of a substance by $1.00\text{ }^{\circ}\text{C}$; also called "specific heat capacity"

Phase Changes

By the end of this section, you will be able to:

- Describe phase transitions and equilibrium between phases
- Solve problems involving latent heat
- Solve calorimetry problems involving phase changes

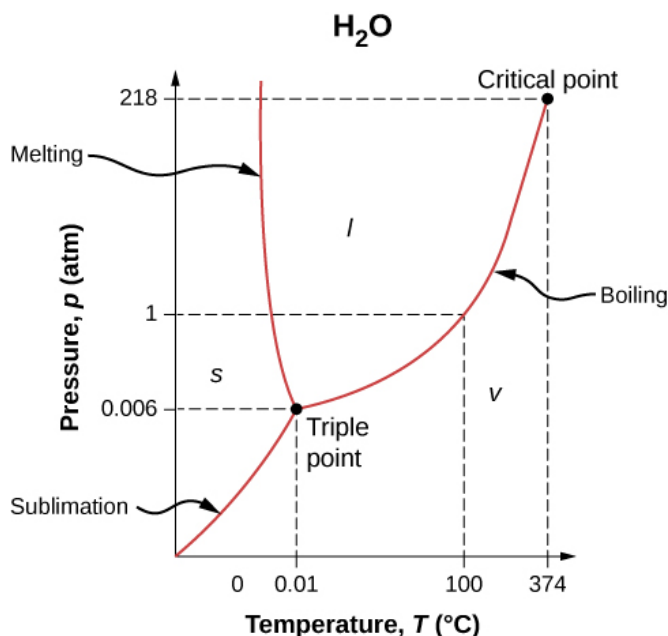
Phase transitions play an important theoretical and practical role in the study of heat flow. In melting (or “fusion”), a solid turns into a liquid; the opposite process is freezing. In evaporation, a liquid turns into a gas; the opposite process is condensation.

A substance melts or freezes at a temperature called its melting point, and boils (evaporates rapidly) or condenses at its boiling point. These temperatures depend on pressure. High pressure favors the denser form, so typically, high pressure raises the melting point and boiling point, and low pressure lowers them. For example, the boiling point of water is 100 °C at 1.00 atm. At higher pressure, the boiling point is higher, and at lower pressure, it is lower. The main exception is the melting and freezing of water, discussed in the next section.

Phase Diagrams

The phase of a given substance depends on the pressure and temperature. Thus, plots of pressure versus temperature showing the phase in each region provide considerable insight into thermal properties of substances. Such a pT graph is called a **phase diagram**.

[\[link\]](#) shows the phase diagram for water. Using the graph, if you know the pressure and temperature, you can determine the phase of water. The solid curves—boundaries between phases—indicate phase transitions, that is, temperatures and pressures at which the phases coexist. For example, the boiling point of water is 100 °C at 1.00 atm. As the pressure increases, the boiling temperature rises gradually to 374 °C at a pressure of 218 atm. A pressure cooker (or even a covered pot) cooks food faster than an open pot, because the water can exist as a liquid at temperatures greater than 100 °C without all boiling away. (As we’ll see in the next section, liquid water conducts heat better than steam or hot air.) The boiling point curve ends at a certain point called the **critical point**—that is, a **critical temperature**, above which the liquid and gas phases cannot be distinguished; the substance is called a *supercritical fluid*. At sufficiently high pressure above the critical point, the gas has the density of a liquid but does not condense. Carbon dioxide, for example, is supercritical at all temperatures above 31.0 °C. **Critical pressure** is the pressure of the critical point.



The phase diagram (pT graph) for water shows solid (s), liquid (l), and vapor (v) phases. At temperatures and pressure above those of the critical point, there is no distinction between liquid and vapor. Note that the axes are nonlinear and the graph is not to scale. This graph is simplified—it omits several exotic phases of ice at higher pressures. The phase diagram of water is unusual because the melting-point curve has a negative slope, showing that you can melt ice by *increasing* the pressure.

Similarly, the curve between the solid and liquid regions in [\[link\]](#) gives the melting temperature at various pressures. For example, the melting point is 0 °C at 1.00 atm, as expected. Water has the unusual property that ice is less dense than liquid water at the melting point, so at a fixed temperature, you can change the phase from solid (ice) to liquid (water) by increasing the pressure. That is, the melting temperature of ice falls with increased pressure, as the phase diagram shows. For example, when a car is driven over snow, the increased pressure from the tires melts the snowflakes; afterwards, the water refreezes and forms an ice layer.

As you learned in the earlier section on thermometers and temperature scales, the triple point is the combination of temperature and pressure at which ice, liquid water, and water vapor can coexist stably—that is, all three phases exist in equilibrium. For water, the triple point occurs at 273.16 K (0.01 °C) and 611.2 Pa; that is a more accurate calibration temperature than the melting point of water at 1.00 atm, or 273.15 K (0.0 °C).

Note:

View this [video](#) to see a substance at its triple point.

At pressures below that of the triple point, there is no liquid phase; the substance can exist as either gas or solid. For water, there is no liquid phase at pressures below 0.00600 atm. The phase change from solid to gas is called

sublimation. You may have noticed that snow can disappear into thin air without a trace of liquid water, or that ice cubes can disappear in a freezer. Both are examples of sublimation. The reverse also happens: Frost can form on very cold windows without going through the liquid stage. [\[link\]](#) shows the result, as well as showing a familiar example of sublimation. Carbon dioxide has no liquid phase at atmospheric pressure. Solid CO_2 is known as dry ice because instead of melting, it sublimates. Its sublimation temperature at atmospheric pressure is -78°C . Certain air fresheners use the sublimation of a solid to spread a perfume around a room. Some solids, such as osmium tetroxide, are so toxic that they must be kept in sealed containers to prevent human exposure to their sublimation-produced vapors.



(a)



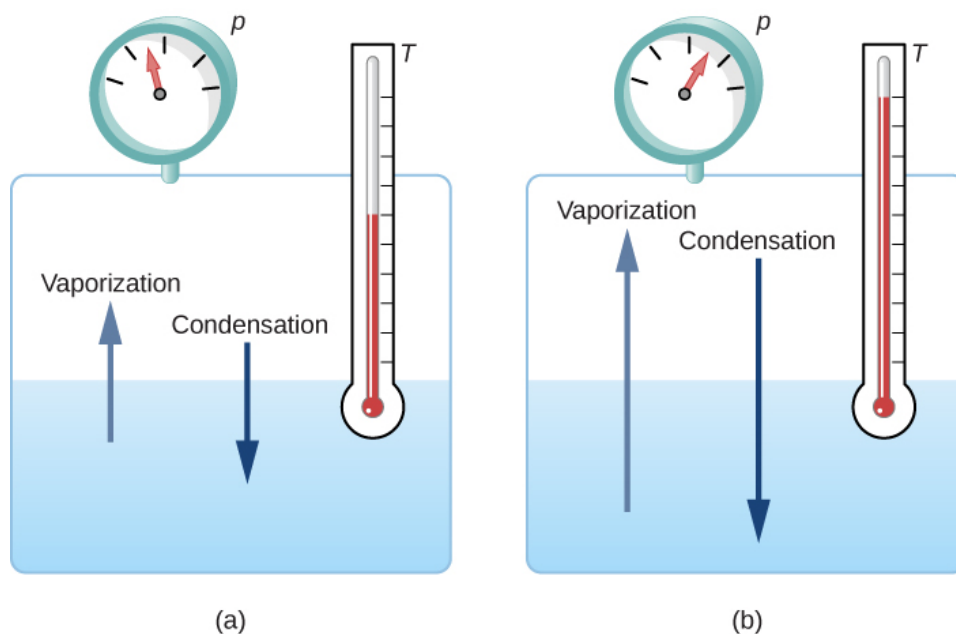
(b)

Direct transitions between solid and vapor are common, sometimes useful, and even beautiful. (a) Dry ice sublimates directly to carbon dioxide gas. The visible “smoke” consists of water droplets that condensed in the air cooled by the dry ice. (b) Frost forms patterns on a very cold window, an example of a solid formed directly from a vapor. (credit a: modification of work by Windell Oskay; credit b: modification of work by Liz West)

Equilibrium

At the melting temperature, the solid and liquid phases are in equilibrium. If heat is added, some of the solid will melt, and if heat is removed, some of the liquid will freeze. The situation is somewhat more complex for liquid-gas equilibrium. Generally, liquid and gas are in equilibrium at any temperature. We call the gas phase a **vapor** when it exists at a temperature below the boiling temperature, as it does for water at 20.0°C . Liquid in a closed container at a fixed temperature evaporates until the pressure of the gas reaches a certain value, called the **vapor pressure**, which depends on the gas and the temperature. At this equilibrium, if heat is added, some of the liquid will evaporate, and if heat is removed, some of the gas will condense; molecules either join the liquid or form suspended droplets. If there is not enough liquid for the gas to reach the vapor pressure in the container, all the liquid eventually evaporates.

If the vapor pressure of the liquid is greater than the *total* ambient pressure, including that of any air (or other gas), the liquid evaporates rapidly; in other words, it boils. Thus, the boiling point of a liquid at a given pressure is the temperature at which its vapor pressure equals the ambient pressure. Liquid and gas phases are in equilibrium at the boiling temperature ([\[link\]](#)). If a substance is in a closed container at the boiling point, then the liquid is boiling and the gas is condensing at the same rate without net change in their amounts.



Equilibrium between liquid and gas at two different boiling points inside a closed container. (a) The rates of boiling and condensation are equal at this combination of temperature and pressure, so the liquid and gas phases are in equilibrium. (b) At a higher temperature, the boiling rate is faster, that is, the rate at which molecules leave the liquid and enter the gas is faster. This increases the number of molecules in the gas, which increases the gas pressure, which in turn increases the rate at which gas molecules condense and enter the liquid. The pressure stops increasing when it reaches the point where the boiling rate and the condensation rate are equal. The gas and liquid are in equilibrium again at this higher temperature and pressure.

For water, $100\text{ }^{\circ}\text{C}$ is the boiling point at 1.00 atm , so water and steam should exist in equilibrium under these conditions. Why does an open pot of water at $100\text{ }^{\circ}\text{C}$ boil completely away? The gas surrounding an open pot is not pure water: it is mixed with air. If pure water and steam are in a closed container at $100\text{ }^{\circ}\text{C}$ and 1.00 atm , they will coexist—but with air over the pot, there are fewer water molecules to condense, and water boils away. Another way to see this is that at the boiling point, the vapor pressure equals the ambient pressure. However, part of the ambient pressure is due to air, so the pressure of the steam is less than the vapor pressure at that temperature, and evaporation continues. Incidentally, the equilibrium vapor pressure of solids is not zero, a fact that accounts for sublimation.

Note:

Exercise:

Problem:

Check Your Understanding Explain why a cup of water (or soda) with ice cubes stays at $0\text{ }^{\circ}\text{C}$, even on a hot summer day.

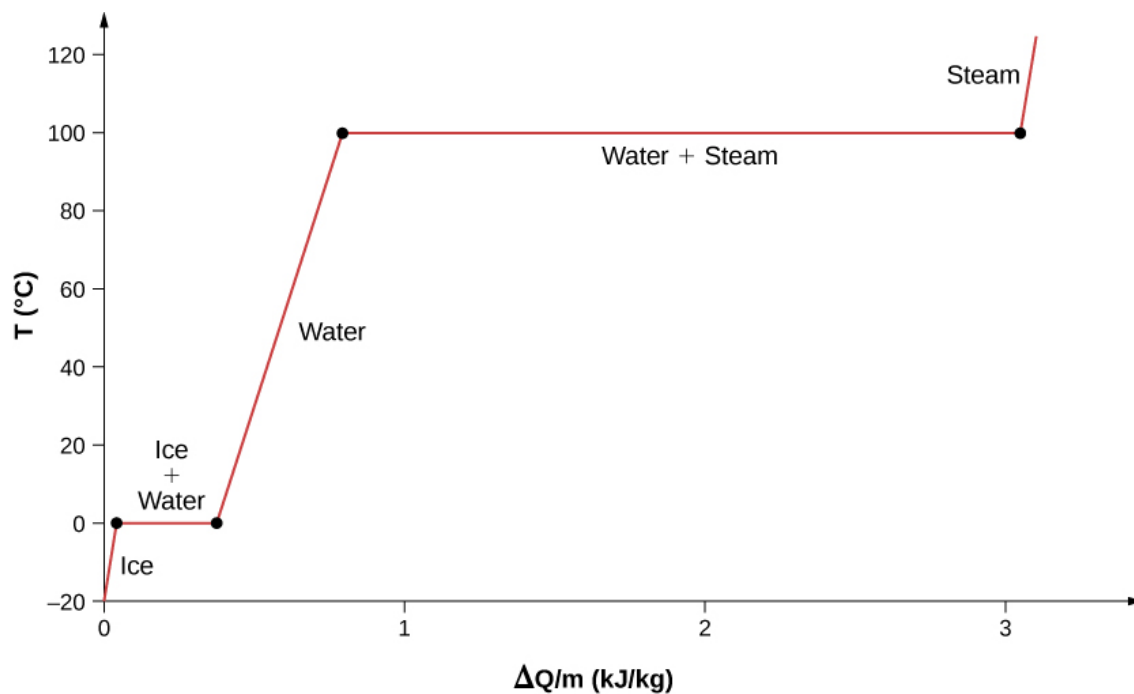
Solution:

The ice and liquid water are in thermal equilibrium, so that the temperature stays at the freezing temperature as long as ice remains in the liquid. (Once all of the ice melts, the water temperature will start to rise.)

Phase Change and Latent Heat

So far, we have discussed heat transfers that cause temperature change. However, in a phase transition, heat transfer does not cause any temperature change.

For an example of phase changes, consider the addition of heat to a sample of ice at $-20\text{ }^{\circ}\text{C}$ ([link](#)) and atmospheric pressure. The temperature of the ice rises linearly, absorbing heat at a constant rate of $2090\text{ J/kg} \cdot ^{\circ}\text{C}$ until it reaches $0\text{ }^{\circ}\text{C}$. Once at this temperature, the ice begins to melt and continues until it has all melted, absorbing 333 kJ/kg of heat. The temperature remains constant at $0\text{ }^{\circ}\text{C}$ during this phase change. Once all the ice has melted, the temperature of the liquid water rises, absorbing heat at a new constant rate of $4186\text{ J/kg} \cdot ^{\circ}\text{C}$. At $100\text{ }^{\circ}\text{C}$, the water begins to boil. The temperature again remains constant during this phase change while the water absorbs 2256 kJ/kg of heat and turns into steam. When all the liquid has become steam, the temperature rises again, absorbing heat at a rate of $2020\text{ J/kg} \cdot ^{\circ}\text{C}$. If we started with steam and cooled it to make it condense into liquid water and freeze into ice, the process would exactly reverse, with the temperature again constant during each phase transition.



Temperature versus heat. The system is constructed so that no vapor evaporates while ice warms to become liquid water, and so that, when vaporization occurs, the vapor remains in the system. The long stretches of constant temperatures at $0\text{ }^{\circ}\text{C}$ and $100\text{ }^{\circ}\text{C}$ reflect the large amounts of heat needed to cause melting and vaporization, respectively.

Where does the heat added during melting or boiling go, considering that the temperature does not change until the transition is complete? Energy is required to melt a solid, because the attractive forces between the molecules in the solid must be broken apart, so that in the liquid, the molecules can move around at comparable kinetic energies; thus, there is no rise in temperature. Energy is needed to vaporize a liquid for similar reasons. Conversely, work is done by attractive forces when molecules are brought together during freezing and condensation. That energy must be transferred out of the system, usually in the form of heat, to allow the molecules to stay together ([link](#)). Thus, condensation occurs in association with cold objects—the glass in [link](#), for example.



Condensation forms on this glass of iced tea because the temperature of the nearby air is reduced. The air cannot hold as much water as it did at room temperature, so water condenses. Energy is released when the water condenses, speeding the melting of the ice in the glass. (credit: Jenny Downing)

The energy released when a liquid freezes is used by orange growers when the temperature approaches 0°C . Growers spray water on the trees so that the water freezes and heat is released to the growing oranges. This prevents the temperature inside the orange from dropping below freezing, which would damage the fruit ([link](#)).



The ice on these trees released large amounts of energy when it froze, helping to prevent the temperature of the trees from dropping below 0°C . Water is intentionally sprayed on orchards to help prevent hard frosts. (credit: Hermann Hammer)

The energy involved in a phase change depends on the number of bonds or force pairs and their strength. The number of bonds is proportional to the number of molecules and thus to the mass of the sample. The energy per unit mass required to change a substance from the solid phase to the liquid phase, or released when the substance changes from liquid to solid, is known as the **heat of fusion**. The energy per unit mass required to change a substance from the liquid phase to the vapor phase is known as the **heat of vaporization**. The strength of the forces depends on the type of molecules. The heat Q absorbed or released in a phase change in a sample of mass m is given by

Note:

Equation:

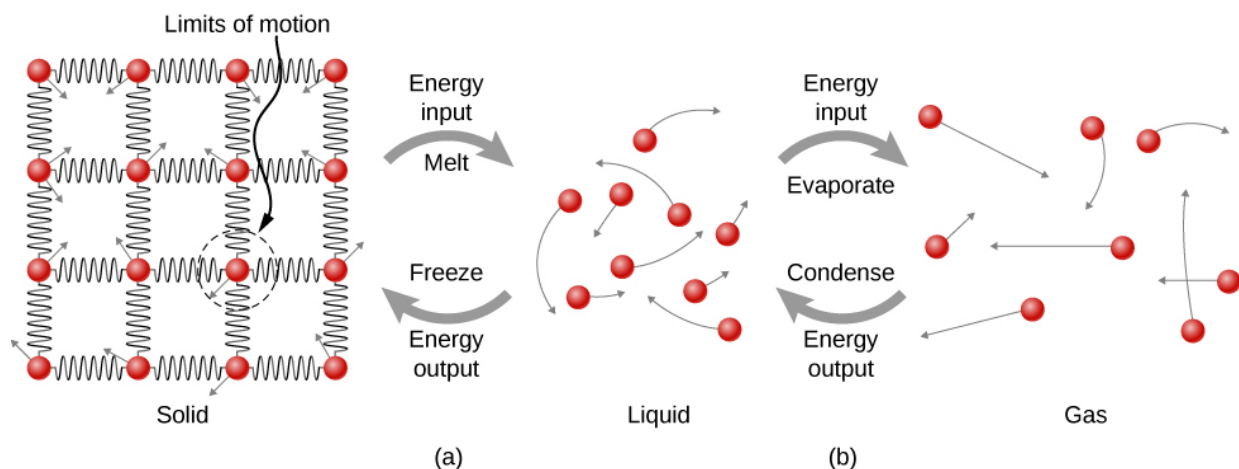
$$Q = mL_f(\text{melting/freezing})$$

Note:

Equation:

$$Q = mL_v(\text{vaporization/condensation})$$

where the latent heat of fusion L_f and latent heat of vaporization L_v are material constants that are determined experimentally. (Latent heats are also called **latent heat coefficients** and heats of transformation.) These constants are “latent,” or hidden, because in phase changes, energy enters or leaves a system without causing a temperature change in the system, so in effect, the energy is hidden.



(a) Energy is required to partially overcome the attractive forces (modeled as springs) between molecules in a solid to form a liquid. That same energy must be removed from the liquid for freezing to take place. (b) Molecules become separated by large distances when going from liquid to vapor, requiring significant energy to completely overcome molecular attraction. The same energy must be removed from the vapor for condensation to take place.

[\[link\]](#) lists representative values of L_f and L_v in kJ/kg, together with melting and boiling points. Note that in general, $L_v > L_f$. The table shows that the amounts of energy involved in phase changes can easily be comparable to or greater than those involved in temperature changes, as [\[link\]](#) and the accompanying discussion also showed.

Substance	Melting Point (°C)	L_f		Boiling Point (°C)	L_v	
		kJ/kg	kcal/kg		kJ/kg	kcal/kg
Helium ^[2]	−272.2 (0.95 K)	5.23	1.25	−268.9 (4.2 K)	20.9	4.99
Hydrogen	−259.3 (13.9 K)	58.6	14.0	−252.9 (20.2 K)	452	108
Nitrogen	−210.0 (63.2 K)	25.5	6.09	−195.8 (77.4 K)	201	48.0
Oxygen	−218.8 (54.4 K)	13.8	3.30	−183.0 (90.2 K)	213	50.9
Ethanol	−114	104	24.9	78.3	854	204
Ammonia	−75	332	79.3	−33.4	1370	327
Mercury	−38.9	11.8	2.82	357	272	65.0
Water	0.00	334	79.8	100.0	2256 ^[3]	539 ^[4]

		L_f			L_v	
Sulfur	119	38.1	9.10	444.6	326	77.9
Lead	327	24.5	5.85	1750	871	208
Antimony	631	165	39.4	1440	561	134
Aluminum	660	380	90	2450	11400	2720
Silver	961	88.3	21.1	2193	2336	558
Gold	1063	64.5	15.4	2660	1578	377
Copper	1083	134	32.0	2595	5069	1211
Uranium	1133	84	20	3900	1900	454
Tungsten	3410	184	44	5900	4810	1150

Heats of Fusion and Vaporization^{[1][1]}Values quoted at the normal melting and boiling temperatures at standard atmospheric pressure (1 atm). ^[2]Helium has no solid phase at atmospheric pressure. The melting point given is at a pressure of 2.5 MPa. ^[3]At 37.0 °C (body temperature), the heat of vaporization L_v for water is 2430 kJ/kg or 580 kcal/kg. ^[4]At 37.0 °C (body temperature), the heat of vaporization, L_v for water is 2430 kJ/kg or 580 kcal/kg.

Phase changes can have a strong stabilizing effect on temperatures that are not near the melting and boiling points, since evaporation and condensation occur even at temperatures below the boiling point. For example, air temperatures in humid climates rarely go above approximately 38.0 °C because most heat transfer goes into evaporating water into the air. Similarly, temperatures in humid weather rarely fall below the dew point—the temperature where condensation occurs given the concentration of water vapor in the air—because so much heat is released when water vapor condenses.

More energy is required to evaporate water below the boiling point than at the boiling point, because the kinetic energy of water molecules at temperatures below 100 °C is less than that at 100 °C, so less energy is available from random thermal motions. For example, at body temperature, evaporation of sweat from the skin requires a heat input of 2428 kJ/kg, which is about 10% higher than the latent heat of vaporization at 100 °C. This heat comes from the skin, and this evaporative cooling effect of sweating helps reduce the body temperature in hot weather. However, high humidity inhibits evaporation, so that body temperature might rise, while unevaporated sweat might be left on your brow.

Example:

Calculating Final Temperature from Phase Change

Three ice cubes are used to chill a soda at 20 °C with mass $m_{\text{soda}} = 0.25$ kg. The ice is at 0 °C and each ice cube has a mass of 6.0 g. Assume that the soda is kept in a foam container so that heat loss can be ignored and that the soda has the same specific heat as water. Find the final temperature when all ice has melted.

Strategy

The ice cubes are at the melting temperature of 0 °C. Heat is transferred from the soda to the ice for melting. Melting yields water at 0 °C, so more heat is transferred from the soda to this water until the water plus soda system reaches thermal equilibrium.

The heat transferred to the ice is

Equation:

$$Q_{\text{ice}} = m_{\text{ice}}L_f + m_{\text{ice}}c_W (T_f - 0^\circ \text{C}).$$

The heat given off by the soda is

Equation:

$$Q_{\text{soda}} = m_{\text{soda}}c_W (T_f - 20^\circ \text{C}).$$

Since no heat is lost, $Q_{\text{ice}} = -Q_{\text{soda}}$, as in [\[link\]](#), so that

Equation:

$$m_{\text{ice}}L_f + m_{\text{ice}}c_W (T_f - 0^\circ \text{C}) = -m_{\text{soda}}c_W (T_f - 20^\circ \text{C}).$$

Solve for the unknown quantity T_f :

Equation:

$$T_f = \frac{m_{\text{soda}}c_W (20^\circ \text{C}) - m_{\text{ice}}L_f}{(m_{\text{soda}} + m_{\text{ice}})c_W}.$$

Solution

First we identify the known quantities. The mass of ice is $m_{\text{ice}} = 3 \times 6.0 \text{ g} = 0.018 \text{ kg}$ and the mass of soda is $m_{\text{soda}} = 0.25 \text{ kg}$. Then we calculate the final temperature:

Equation:

$$T_f = \frac{20,930 \text{ J} - 6012 \text{ J}}{1122 \text{ J}/^\circ \text{C}} = 13^\circ \text{C}.$$

Significance

This example illustrates the large energies involved during a phase change. The mass of ice is about 7% of the mass of the soda but leads to a noticeable change in the temperature of the soda. Although we assumed that the ice was at the freezing temperature, this is unrealistic for ice straight out of a freezer: The typical temperature is -6°C . However, this correction makes no significant change from the result we found. Can you explain why?

Like solid-liquid and liquid-vapor transitions, direct solid-vapor transitions or sublimations involve heat. The energy transferred is given by the equation $Q = mL_s$, where L_s is the **heat of sublimation**, analogous to L_f and L_v . The heat of sublimation at a given temperature is equal to the heat of fusion plus the heat of vaporization at that temperature.

We can now calculate any number of effects related to temperature and phase change. In each case, it is necessary to identify which temperature and phase changes are taking place. Keep in mind that heat transfer and work can cause both temperature and phase changes.

Note:

The Effects of Heat Transfer

1. Examine the situation to determine that there is a change in the temperature or phase. Is there heat transfer into or out of the system? When it is not obvious whether a phase change occurs or not, you may wish to first solve the problem as if there were no phase changes, and examine the temperature change obtained. If it is sufficient to take you past a boiling or melting point, you should then go back and do the problem in steps—temperature change, phase change, subsequent temperature change, and so on.
2. Identify and list all objects that change temperature or phase.
3. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.

4. Make a list of what is given or what can be inferred from the problem as stated (identify the knowns). If there is a temperature change, the transferred heat depends on the specific heat of the substance ([Heat Transfer, Specific Heat, and Calorimetry](#)), and if there is a phase change, the transferred heat depends on the latent heat of the substance ([link](#)).
5. Solve the appropriate equation for the quantity to be determined (the unknown).
6. Substitute the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units. You may need to do this in steps if there is more than one state to the process, such as a temperature change followed by a phase change. However, in a calorimetry problem, each step corresponds to a term in the single equation $Q_{\text{hot}} + Q_{\text{cold}} = 0$.
7. Check the answer to see if it is reasonable. Does it make sense? As an example, be certain that any temperature change does not also cause a phase change that you have not taken into account.

Note:

Exercise:

Problem:

Check Your Understanding Why does snow often remain even when daytime temperatures are higher than the freezing temperature?

Solution:

Snow is formed from ice crystals and thus is the solid phase of water. Because enormous heat is necessary for phase changes, it takes a certain amount of time for this heat to be transferred from the air, even if the air is above 0°C .

Summary

- Most substances have three distinct phases (under ordinary conditions on Earth), and they depend on temperature and pressure.
- Two phases coexist (i.e., they are in thermal equilibrium) at a set of pressures and temperatures.
- Phase changes occur at fixed temperatures for a given substance at a given pressure, and these temperatures are called boiling, freezing (or melting), and sublimation points.

Conceptual Questions

Exercise:

Problem:

A pressure cooker contains water and steam in equilibrium at a pressure greater than atmospheric pressure. How does this greater pressure increase cooking speed?

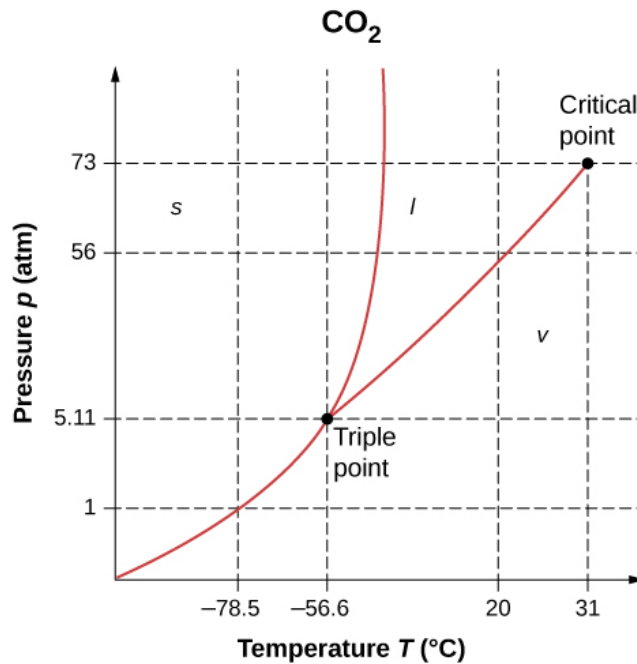
Solution:

It raises the boiling point, so the water, which the food gains heat from, is at a higher temperature.

Exercise:

Problem:

As shown below, which is the phase diagram for carbon dioxide, what is the vapor pressure of solid carbon dioxide (dry ice) at $-78.5\text{ }^{\circ}\text{C}$? (Note that the axes in the figure are nonlinear and the graph is not to scale.)



Exercise:

Problem:

Can carbon dioxide be liquefied at room temperature ($20\text{ }^{\circ}\text{C}$)? If so, how? If not, why not? (See the phase diagram in the preceding problem.)

Solution:

Yes, by raising the pressure above 56 atm.

Exercise:

Problem: What is the distinction between gas and vapor?

Exercise:

Problem: Heat transfer can cause temperature and phase changes. What else can cause these changes?

Solution:

work

Exercise:

Problem:

How does the latent heat of fusion of water help slow the decrease of air temperatures, perhaps preventing temperatures from falling significantly below $0\text{ }^{\circ}\text{C}$, in the vicinity of large bodies of water?

Exercise:

Problem: What is the temperature of ice right after it is formed by freezing water?

Solution:

0 °C (at or near atmospheric pressure)

Exercise:

Problem:

If you place 0 °C ice into 0 °C water in an insulated container, what will the net result be? Will there be less ice and more liquid water, or more ice and less liquid water, or will the amounts stay the same?

Exercise:

Problem:

What effect does condensation on a glass of ice water have on the rate at which the ice melts? Will the condensation speed up the melting process or slow it down?

Solution:

Condensation releases heat, so it speeds up the melting.

Exercise:

Problem:

In Miami, Florida, which has a very humid climate and numerous bodies of water nearby, it is unusual for temperatures to rise above about 38 °C (100 °F). In the desert climate of Phoenix, Arizona, however, temperatures rise above that almost every day in July and August. Explain how the evaporation of water helps limit high temperatures in humid climates.

Exercise:

Problem:

In winter, it is often warmer in San Francisco than in Sacramento, 150 km inland. In summer, it is nearly always hotter in Sacramento. Explain how the bodies of water surrounding San Francisco moderate its extreme temperatures.

Solution:

Because of water's high specific heat, it changes temperature less than land. Also, evaporation reduces temperature rises. The air tends to stay close to equilibrium with the water, so its temperature does not change much where there's a lot of water around, as in San Francisco but not Sacramento.

Exercise:

Problem:

Freeze-dried foods have been dehydrated in a vacuum. During the process, the food freezes and must be heated to facilitate dehydration. Explain both how the vacuum speeds up dehydration and why the food freezes as a result.

Exercise:

Problem:

In a physics classroom demonstration, an instructor inflates a balloon by mouth and then cools it in liquid nitrogen. When cold, the shrunken balloon has a small amount of light blue liquid in it, as well as some snow-like crystals. As it warms up, the liquid boils, and part of the crystals sublime, with some crystals lingering for a while and then producing a liquid. Identify the blue liquid and the two solids in the cold balloon. Justify your identifications using data from [\[link\]](#).

Solution:

The liquid is oxygen, whose boiling point is above that of nitrogen but whose melting point is below the boiling point of liquid nitrogen. The crystals that sublime are carbon dioxide, which has no liquid phase at atmospheric pressure. The crystals that melt are water, whose melting point is above carbon dioxide's sublimation point. The water came from the instructor's breath.

Problems**Exercise:****Problem:**

How much heat transfer (in kilocalories) is required to thaw a 0.450-kg package of frozen vegetables originally at 0°C if their heat of fusion is the same as that of water?

Exercise:**Problem:**

A bag containing 0°C ice is much more effective in absorbing energy than one containing the same amount of 0°C water. (a) How much heat transfer is necessary to raise the temperature of 0.800 kg of water from 0°C to 30.0°C ? (b) How much heat transfer is required to first melt 0.800 kg of 0°C ice and then raise its temperature? (c) Explain how your answer supports the contention that the ice is more effective.

Solution:

a. $1.00 \times 10^5 \text{ J}$; b. $3.68 \times 10^5 \text{ J}$; c. The ice is much more effective in absorbing heat because it first must be melted, which requires a lot of energy, and then it gains the same amount of heat as the bag that started with water. The first $2.67 \times 10^5 \text{ J}$ of heat is used to melt the ice, then it absorbs the $1.00 \times 10^5 \text{ J}$ of heat as water.

Exercise:**Problem:**

(a) How much heat transfer is required to raise the temperature of a 0.750-kg aluminum pot containing 2.50 kg of water from 30.0°C to the boiling point and then boil away 0.750 kg of water? (b) How long does this take if the rate of heat transfer is 500 W?

Exercise:**Problem:**

Condensation on a glass of ice water causes the ice to melt faster than it would otherwise. If 8.00 g of vapor condense on a glass containing both water and 200 g of ice, how many grams of the ice will melt as a result? Assume no other heat transfer occurs. Use L_v for water at 37°C as a better approximation than L_v for water at 100°C .)

Solution:

58.1 g

Exercise:**Problem:**

On a trip, you notice that a 3.50-kg bag of ice lasts an average of one day in your cooler. What is the average power in watts entering the ice if it starts at 0 °C and completely melts to 0 °C water in exactly one day?

Exercise:**Problem:**

On a certain dry sunny day, a swimming pool's temperature would rise by 1.50 °C if not for evaporation. What fraction of the water must evaporate to carry away precisely enough energy to keep the temperature constant?

Solution:

Let M be the mass of pool water and m be the mass of pool water that evaporates.

$$Mc\Delta T = mL_{V(37^\circ\text{C})} \Rightarrow \frac{m}{M} = \frac{c\Delta T}{L_{V(37^\circ\text{C})}} = \frac{(1.00 \text{ kcal/kg}\cdot^\circ\text{C})(1.50^\circ\text{C})}{580 \text{ kcal/kg}} = 2.59 \times 10^{-3};$$

(Note that L_V for water at 37 °C is used here as a better approximation than L_V for 100 °C water.)

Exercise:**Problem:**

(a) How much heat transfer is necessary to raise the temperature of a 0.200-kg piece of ice from -20.0°C to 130.0°C , including the energy needed for phase changes? (b) How much time is required for each stage, assuming a constant 20.0 kJ/s rate of heat transfer? (c) Make a graph of temperature versus time for this process.

Exercise:**Problem:**

In 1986, an enormous iceberg broke away from the Ross Ice Shelf in Antarctica. It was an approximately rectangular prism 160 km long, 40.0 km wide, and 250 m thick. (a) What is the mass of this iceberg, given that the density of ice is 917 kg/m³? (b) How much heat transfer (in joules) is needed to melt it? (c) How many years would it take sunlight alone to melt ice this thick, if the ice absorbs an average of 100 W/m², 12.00 h per day?

Solution:

a. 1.47×10^{15} kg; b. 4.90×10^{20} J; c. 48.5 y

Exercise:**Problem:**

How many grams of coffee must evaporate from 350 g of coffee in a 100-g glass cup to cool the coffee and the cup from 95.0 °C to 45.0 °C? Assume the coffee has the same thermal properties as water and that the average heat of vaporization is 2340 kJ/kg (560 kcal/g). Neglect heat losses through processes other than evaporation, as well as the change in mass of the coffee as it cools. Do the latter two assumptions cause your answer to be higher or lower than the true answer?

Exercise:

Problem:

(a) It is difficult to extinguish a fire on a crude oil tanker, because each liter of crude oil releases $2.80 \times 10^7 \text{ J}$ of energy when burned. To illustrate this difficulty, calculate the number of liters of water that must be expended to absorb the energy released by burning 1.00 L of crude oil, if the water's temperature rises from 20.0°C to 100°C , it boils, and the resulting steam's temperature rises to 300°C at constant pressure. (b) Discuss additional complications caused by the fact that crude oil is less dense than water.

Solution:

a. 9.35 L; b. Crude oil is less dense than water, so it floats on top of the water, thereby exposing it to the oxygen in the air, which it uses to burn. Also, if the water is under the oil, it is less able to absorb the heat generated by the oil.

Exercise:**Problem:**

The energy released from condensation in thunderstorms can be very large. Calculate the energy released into the atmosphere for a small storm of radius 1 km, assuming that 1.0 cm of rain is precipitated uniformly over this area.

Exercise:**Problem:**

To help prevent frost damage, 4.00 kg of water at 0°C is sprayed onto a fruit tree. (a) How much heat transfer occurs as the water freezes? (b) How much would the temperature of the 200-kg tree decrease if this amount of heat transferred from the tree? Take the specific heat to be $3.35 \text{ kJ/kg} \cdot ^\circ\text{C}$, and assume that no phase change occurs in the tree.

Solution:

a. 319 kcal; b. 2.00°C

Exercise:**Problem:**

A 0.250-kg aluminum bowl holding 0.800 kg of soup at 25.0°C is placed in a freezer. What is the final temperature if 388 kJ of energy is transferred from the bowl and soup, assuming the soup's thermal properties are the same as that of water?

Exercise:**Problem:**

A 0.0500-kg ice cube at -30.0°C is placed in 0.400 kg of 35.0°C water in a very well-insulated container. What is the final temperature?

Solution:

First bring the ice up to 0°C and melt it with heat $Q_1 : 4.74 \text{ kcal}$. This lowers the temperature of water by $\Delta T_2 : 23.15^\circ\text{C}$. Now, the heat lost by the hot water equals that gained by the cold water (T_f is the final temperature): 20.6°C

Exercise:

Problem:

If you pour 0.0100 kg of 20.0 °C water onto a 1.20-kg block of ice (which is initially at −15.0 °C), what is the final temperature? You may assume that the water cools so rapidly that effects of the surroundings are negligible.

Exercise:**Problem:**

Indigenous people sometimes cook in watertight baskets by placing hot rocks into water to bring it to a boil. What mass of 500-°C granite must be placed in 4.00 kg of 15.0-°C water to bring its temperature to 100 °C, if 0.0250 kg of water escapes as vapor from the initial sizzle? You may neglect the effects of the surroundings.

Solution:

Let the subscripts r, e, v, and w represent rock, equilibrium, vapor, and water, respectively.

$$m_r c_r (T_1 - T_e) = m_v L_v + m_w c_w (T_e - T_2);$$

$$\begin{aligned} m_r &= \frac{m_v L_v + m_w c_w (T_e - T_2)}{c_r (T_1 - T_e)} \\ &= \frac{(0.0250 \text{ kg})(2256 \times 10^3 \text{ J/kg}) + (3.975 \text{ kg})(4186 \times 10^3 \text{ J/kg} \cdot ^\circ\text{C})(100^\circ\text{C} - 15^\circ\text{C})}{(840 \text{ J/kg} \cdot ^\circ\text{C})(500^\circ\text{C} - 100^\circ\text{C})} \\ &= 4.38 \text{ kg} \end{aligned}$$

Exercise:**Problem:**

What would the final temperature of the pan and water be in [\[link\]](#) if 0.260 kg of water were placed in the pan and 0.0100 kg of the water evaporated immediately, leaving the remainder to come to a common temperature with the pan?

Glossary

critical point

for a given substance, the combination of temperature and pressure above which the liquid and gas phases are indistinguishable

critical pressure

pressure at the critical point

critical temperature

temperature at the critical point

heat of fusion

energy per unit mass required to change a substance from the solid phase to the liquid phase, or released when the substance changes from liquid to solid

heat of sublimation

energy per unit mass required to change a substance from the solid phase to the vapor phase

heat of vaporization

energy per unit mass required to change a substance from the liquid phase to the vapor phase

latent heat coefficient

general term for the heats of fusion, vaporization, and sublimation

phase diagram

graph of pressure vs. temperature of a particular substance, showing at which pressures and temperatures the phases of the substance occur

sublimation

phase change from solid to gas

vapor

gas at a temperature below the boiling temperature

vapor pressure

pressure at which a gas coexists with its solid or liquid phase

Mechanisms of Heat Transfer

By the end of this section, you will be able to:

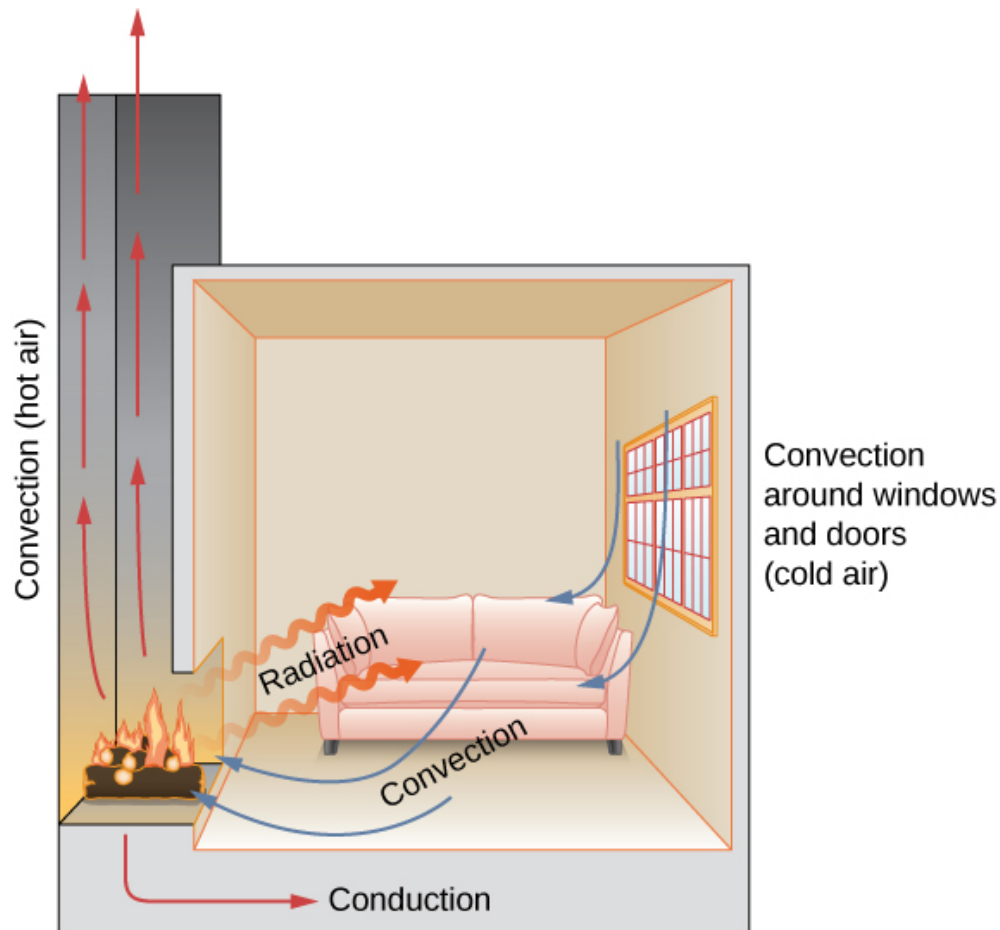
- Explain some phenomena that involve conductive, convective, and radiative heat transfer
- Solve problems on the relationships between heat transfer, time, and rate of heat transfer
- Solve problems using the formulas for conduction and radiation

Just as interesting as the effects of heat transfer on a system are the methods by which it occurs. Whenever there is a temperature difference, heat transfer occurs. It may occur rapidly, as through a cooking pan, or slowly, as through the walls of a picnic ice chest. So many processes involve heat transfer that it is hard to imagine a situation where no heat transfer occurs. Yet every heat transfer takes place by only three methods:

1. **Conduction** is heat transfer through stationary matter by physical contact. (The matter is stationary on a macroscopic scale—we know that thermal motion of the atoms and molecules occurs at any temperature above absolute zero.) Heat transferred from the burner of a stove through the bottom of a pan to food in the pan is transferred by conduction.
2. **Convection** is the heat transfer by the macroscopic movement of a fluid. This type of transfer takes place in a forced-air furnace and in weather systems, for example.
3. Heat transfer by **radiation** occurs when microwaves, infrared radiation, visible light, or another form of electromagnetic radiation is emitted or absorbed. An obvious example is the warming of Earth by the Sun. A less obvious example is thermal radiation from the human body.

In the illustration at the beginning of this chapter, the fire warms the snowshoers' faces largely by radiation. Convection carries some heat to them, but most of the air flow from the fire is upward (creating the familiar shape of flames), carrying heat to the food being cooked and into the sky. The snowshoers wear clothes designed with low conductivity to prevent heat flow out of their bodies.

In this section, we examine these methods in some detail. Each method has unique and interesting characteristics, but all three have two things in common: They transfer heat solely because of a temperature difference, and the greater the temperature difference, the faster the heat transfer ([link](#)).



In a fireplace, heat transfer occurs by all three methods: conduction, convection, and radiation. Radiation is responsible for most of the heat transferred into the room. Heat transfer also occurs through conduction into the room, but much slower. Heat transfer by convection also occurs through cold air entering the room around windows and hot air leaving the room by rising up the chimney.

Note:

Exercise:

Problem:

Check Your Understanding Name an example from daily life (different from the text) for each mechanism of heat transfer.

Solution:

Conduction: Heat transfers into your hands as you hold a hot cup of coffee.

Convection: Heat transfers as the barista “steams” cold milk to make hot cocoa.

Radiation: Heat transfers from the Sun to a jar of water with tea leaves in it to make “Sun tea.” A great many other answers are possible.

Conduction

As you walk barefoot across the living room carpet in a cold house and then step onto the kitchen tile floor, your feet feel colder on the tile. This result is intriguing, since the carpet and tile floor are both at the same temperature. The different sensation is explained by the different rates of heat transfer: The heat loss is faster for skin in contact with the tiles than with the carpet, so the sensation of cold is more intense.

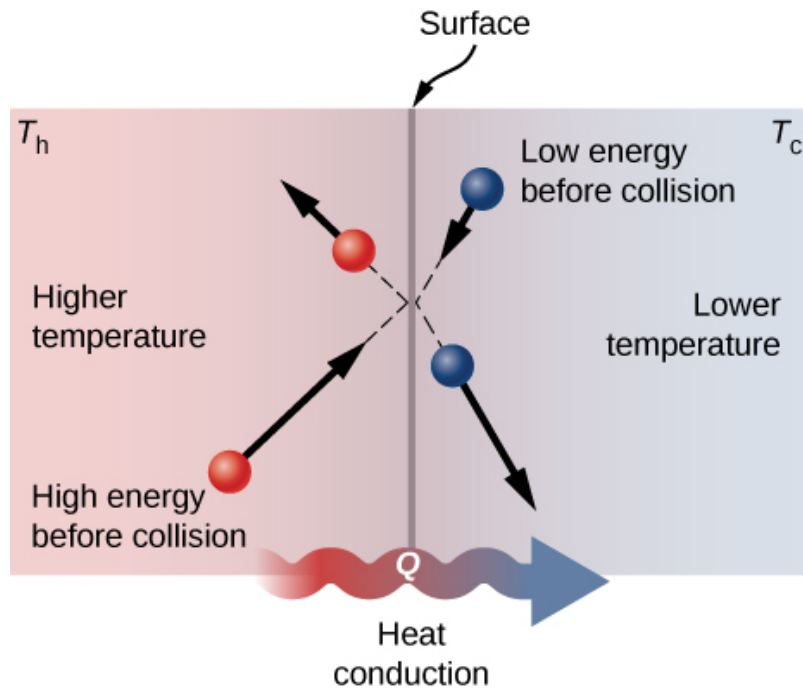
Some materials conduct thermal energy faster than others. [\[link\]](#) shows a material that conducts heat slowly—it is a good thermal insulator, or poor heat conductor—used to reduce heat flow into and out of a house.



Insulation is used to limit the conduction of heat from the inside to the outside (in winter) and from the outside to the inside (in summer). (credit: Giles Douglas)

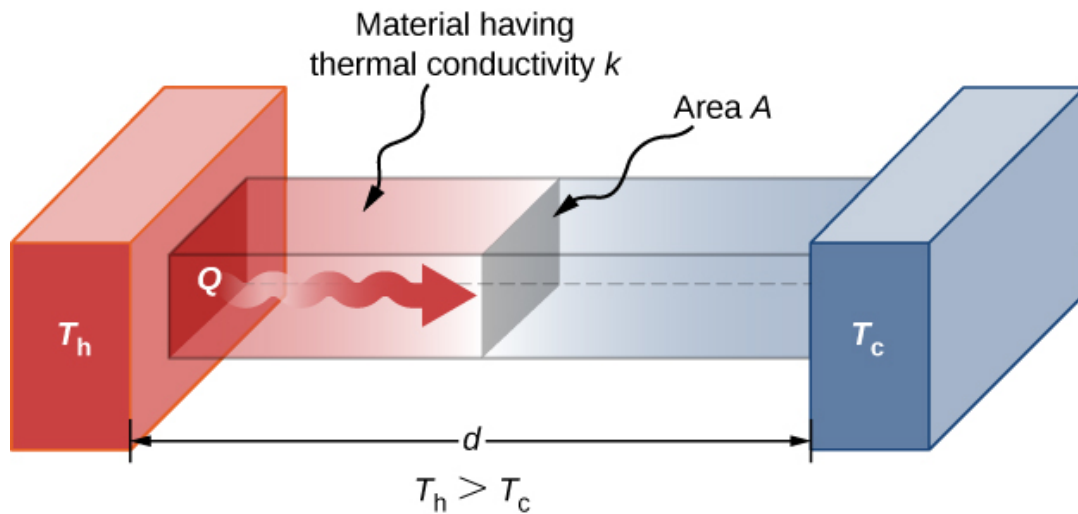
A molecular picture of heat conduction will help justify the equation that describes it. [\[link\]](#) shows molecules in two bodies at different temperatures, T_h and T_c , for “hot” and “cold.” The average kinetic energy of a molecule in the hot body is higher than in the colder body. If two molecules collide, energy transfers from the high-energy to the low-energy molecule. In a metal, the picture would also include free valence electrons colliding with each other and with atoms, likewise transferring energy. The cumulative effect of all collisions is a net flux of heat from the hotter body to the colder body. Thus, the rate of heat transfer increases with increasing temperature difference $\Delta T = T_h - T_c$. If the temperatures are the same, the net heat transfer rate is zero.

Because the number of collisions increases with increasing area, heat conduction is proportional to the cross-sectional area—a second factor in the equation.



Molecules in two bodies at different temperatures have different average kinetic energies. Collisions occurring at the contact surface tend to transfer energy from high-temperature regions to low-temperature regions. In this illustration, a molecule in the lower-temperature region (right side) has low energy before collision, but its energy increases after colliding with a high-energy molecule at the contact surface. In contrast, a molecule in the higher-temperature region (left side) has high energy before collision, but its energy decreases after colliding with a low-energy molecule at the contact surface.

A third quantity that affects the conduction rate is the thickness of the material through which heat transfers. [\[link\]](#) shows a slab of material with a higher temperature on the left than on the right. Heat transfers from the left to the right by a series of molecular collisions. The greater the distance between hot and cold, the more time the material takes to transfer the same amount of heat.



Heat conduction occurs through any material, represented here by a rectangular bar, whether window glass or walrus blubber.

All four of these quantities appear in a simple equation deduced from and confirmed by experiments. The **rate of conductive heat transfer** through a slab of material, such as the one in [\[link\]](#), is given by

Note:

Equation:

$$P = \frac{dQ}{dt} = \frac{kA(T_h - T_c)}{d}$$

where P is the power or rate of heat transfer in watts or in kilocalories per second, A and d are its surface area and thickness, as shown in [\[link\]](#), $T_h - T_c$ is the temperature difference across the slab, and k is the **thermal conductivity** of the material. [\[link\]](#) gives representative values of thermal conductivity.

More generally, we can write

Equation:

$$P = -kA \frac{dT}{dx},$$

where x is the coordinate in the direction of heat flow. Since in [\[link\]](#), the power and area are constant, dT/dx is constant, and the temperature decreases linearly from T_h to T_c .

Substance	Thermal Conductivity k (W/m · °C)
Diamond	2000
Silver	420
Copper	390
Gold	318
Aluminum	220
Steel iron	80
Steel (stainless)	14
Ice	2.2
Glass (average)	0.84
Concrete brick	0.84
Water	0.6
Fatty tissue (without blood)	0.2
Asbestos	0.16
Plasterboard	0.16
Wood	0.08–0.16
Snow (dry)	0.10

Substance	Thermal Conductivity k (W/m · °C)
Cork	0.042
Glass wool	0.042
Wool	0.04
Down feathers	0.025
Air	0.023
Polystyrene foam	0.010

Thermal Conductivities of Common Substances Values are given for temperatures near 0 °C.

Example:

Calculating Heat Transfer through Conduction

A polystyrene foam icebox has a total area of 0.950 m² and walls with an average thickness of 2.50 cm. The box contains ice, water, and canned beverages at 0 °C. The inside of the box is kept cold by melting ice. How much ice melts in one day if the icebox is kept in the trunk of a car at 35.0 °C?

Strategy

This question involves both heat for a phase change (melting of ice) and the transfer of heat by conduction. To find the amount of ice melted, we must find the net heat transferred. This value can be obtained by calculating the rate of heat transfer by conduction and multiplying by time.

Solution

First we identify the knowns.

$k = 0.010$ W/m · °C for polystyrene foam; $A = 0.950$ m²;

$d = 2.50$ cm = 0.0250 m;; $T_c = 0$ °C; $T_h = 35.0$ °C;

$t = 1$ day = 24 hours - 86,400 s.

Then we identify the unknowns. We need to solve for the mass of the ice, m . We also need to solve for the net heat transferred to melt the ice, Q . The rate of heat transfer by conduction is given by

Equation:

$$P = \frac{dQ}{dt} = \frac{kA(T_h - T_c)}{d}.$$

The heat used to melt the ice is $Q = mL_f$. We insert the known values:

Equation:

$$P = \frac{(0.010 \text{ W/m} \cdot ^\circ\text{C}) (0.950 \text{ m}^2) (35.0 ^\circ\text{C} - 0 ^\circ\text{C})}{0.0250 \text{ m}} = 13.3 \text{ W}.$$

Multiplying the rate of heat transfer by the time we obtain

Equation:

$$Q = Pt = (13.3 \text{ W}) (86.400 \text{ s}) = 1.15 \times 10^6 \text{ J}.$$

We set this equal to the heat transferred to melt the ice, $Q = mL_f$, and solve for the mass m :

Equation:

$$m = \frac{Q}{L_f} = \frac{1.15 \times 10^6 \text{ J}}{334 \times 10^3 \text{ J/kg}} = 3.44 \text{ kg}.$$

Significance

The result of 3.44 kg, or about 7.6 lb, seems about right, based on experience. You might expect to use about a 4 kg (7–10 lb) bag of ice per day. A little extra ice is required if you add any warm food or beverages.

[\[link\]](#) shows that polystyrene foam is a very poor conductor and thus a good insulator. Other good insulators include fiberglass, wool, and goose down feathers. Like polystyrene foam, these all contain many small pockets of air, taking advantage of air's poor thermal conductivity.

In developing insulation, the smaller the conductivity k and the larger the thickness d , the better. Thus, the ratio d/k , called the *R factor*, is large for a good insulator. The rate of conductive heat transfer is inversely proportional to R . R factors are most commonly quoted for household insulation, refrigerators, and the like. Unfortunately, in the United States, R is still in non-metric units of $\text{ft}^2 \cdot ^\circ\text{F} \cdot \text{h}/\text{Btu}$, although the unit usually goes unstated [1 British thermal unit (Btu) is the amount of energy needed to change the temperature of 1.0 lb of water by $1.0 ^\circ\text{F}$, which is 1055.1 J]. A couple of representative values are an R factor of 11 for 3.5-inch-thick fiberglass batts (pieces) of insulation and an R factor of 19 for 6.5-inch-thick fiberglass batts ([\[link\]](#)). In the US, walls are usually insulated with 3.5-inch batts, whereas ceilings are usually insulated with 6.5-inch batts. In cold climates, thicker batts may be used.



The fiberglass batt is used for insulation of walls and ceilings to prevent heat transfer between the inside of the building and the outside environment. (credit: Tracey Nicholls)

Note that in [\[link\]](#), most of the best thermal conductors—silver, copper, gold, and aluminum—are also the best electrical conductors, because they contain many free electrons that can transport thermal energy. (Diamond, an electrical insulator, conducts heat by atomic vibrations.) Cooking utensils are typically made from good conductors, but the handles of those used on the stove are made from good insulators (bad conductors).

Example:
Two Conductors End to End

A steel rod and an aluminum rod, each of diameter 1.00 cm and length 25.0 cm, are welded end to end. One end of the steel rod is placed in a large tank of boiling water at 100°C , while the far end of the aluminum rod is placed in a large tank of water at 20°C . The rods are insulated so that no heat escapes from their surfaces. What is the temperature at the joint, and what is the rate of heat conduction through this composite rod?

Strategy

The heat that enters the steel rod from the boiling water has no place to go but through the steel rod, then through the aluminum rod, to the cold water. Therefore, we can equate the rate of conduction through the steel to the rate of conduction through the aluminum.

We repeat the calculation with a second method, in which we use the thermal resistance R of the rod, since it simply adds when two rods are joined end to end. (We will use a similar method in the chapter on direct-current circuits.)

Solution

1. Identify the knowns and convert them to SI units.

The length of each rod is $L_{\text{Al}} = L_{\text{steel}} = 0.25\text{ m}$, the cross-sectional area of each rod is $A_{\text{Al}} = A_{\text{steel}} = 7.85 \times 10^{-5}\text{ m}^2$, the thermal conductivity of aluminum is $k_{\text{Al}} = 220\text{ W/m}\cdot^\circ\text{C}$, the thermal conductivity of steel is $k_{\text{steel}} = 80\text{ W/m}\cdot^\circ\text{C}$, the temperature at the hot end is $T = 100^\circ\text{C}$, and the temperature at the cold end is $T = 20^\circ\text{C}$.

2. Calculate the heat-conduction rate through the steel rod and the heat-conduction rate through the aluminum rod in terms of the unknown temperature T at the joint:

Equation:

$$\begin{aligned} P_{\text{steel}} &= \frac{k_{\text{steel}} A_{\text{steel}} \Delta T_{\text{steel}}}{L_{\text{steel}}} \\ &= \frac{(80\text{ W/m}\cdot^\circ\text{C})(7.85 \times 10^{-5}\text{ m}^2)(100^\circ\text{C} - T)}{0.25\text{ m}} \\ &= (0.0251\text{ W}/^\circ\text{C})(100^\circ\text{C} - T); \end{aligned}$$

Equation:

$$\begin{aligned} P_{\text{Al}} &= \frac{k_{\text{Al}} A_{\text{Al}} \Delta T_{\text{Al}}}{L_{\text{Al}}} \\ &= \frac{(220\text{ W/m}\cdot^\circ\text{C})(7.85 \times 10^{-5}\text{ m}^2)(T - 20^\circ\text{C})}{0.25\text{ m}} \\ &= (0.0691\text{ W}/^\circ\text{C})(T - 20^\circ\text{C}). \end{aligned}$$

3. Set the two rates equal and solve for the unknown temperature:

Equation:

$$\begin{aligned} (0.0691\text{ W}/^\circ\text{C})(T - 20^\circ\text{C}) &= (0.0251\text{ W}/^\circ\text{C})(100^\circ\text{C} - T) \\ T &= 41.3^\circ\text{C}. \end{aligned}$$

4. Calculate either rate:

Equation:

$$P_{\text{steel}} = (0.0251 \text{ W}/^{\circ}\text{C}) (100^{\circ}\text{C} - 41.3^{\circ}\text{C}) = 1.47 \text{ W}.$$

5. If desired, check your answer by calculating the other rate.

Solution

1. Recall that $R = L/k$. Now $P = A\Delta T/R$, or $\Delta T = PR/A$.
2. We know that $\Delta T_{\text{steel}} + \Delta T_{\text{Al}} = 100^{\circ}\text{C} - 20^{\circ}\text{C} = 80^{\circ}\text{C}$. We also know that $P_{\text{steel}} = P_{\text{Al}}$, and we denote that rate of heat flow by P . Combine the equations:

Equation:

$$\frac{PR_{\text{steel}}}{A} + \frac{PR_{\text{Al}}}{A} = 80^{\circ}\text{C}.$$

Thus, we can simply add R factors. Now, $P = \frac{80^{\circ}\text{C}}{A(R_{\text{steel}} + R_{\text{Al}})}$.

3. Find the R_s from the known quantities:

Equation:

$$R_{\text{steel}} = 3.13 \times 10^{-3} \text{ m}^2 \cdot ^{\circ}\text{C}/\text{W}$$

and

Equation:

$$R_{\text{Al}} = 1.14 \times 10^{-3} \text{ m}^2 \cdot ^{\circ}\text{C}/\text{W}.$$

4. Substitute these values in to find $P = 1.47 \text{ W}$ as before.
5. Determine ΔT for the aluminum rod (or for the steel rod) and use it to find T at the joint.

Equation:

$$\Delta T_{\text{Al}} = \frac{PR_{\text{Al}}}{A} = \frac{(1.47 \text{ W}) (1.14 \times 10^{-3} \text{ m}^2 \cdot ^{\circ}\text{C}/\text{W})}{7.85 \times 10^{-5} \text{ m}^2} = 21.3^{\circ}\text{C},$$

so $T = 20^{\circ}\text{C} + 21.3^{\circ}\text{C} = 41.3^{\circ}\text{C}$, as in Solution 1.

6. If desired, check by determining ΔT for the other rod.

Significance

In practice, adding R values is common, as in calculating the R value of an insulated wall. In the analogous situation in electronics, the resistance corresponds to AR in this problem and is additive even when the areas are unequal, as is common in electronics.

Our equation for heat conduction can be used only when the areas are equal; otherwise, we would have a problem in three-dimensional heat flow, which is beyond our scope.

Note:

Exercise:

Problem:

Check Your Understanding How does the rate of heat transfer by conduction change when all spatial dimensions are doubled?

Solution:

Because area is the product of two spatial dimensions, it increases by a factor of four when each dimension is doubled ($A_{\text{final}} = (2d)^2 = 4d^2 = 4A_{\text{initial}}$). The distance, however, simply doubles. Because the temperature difference and the coefficient of thermal conductivity are independent of the spatial dimensions, the rate of heat transfer by conduction increases by a factor of four divided by two, or two:

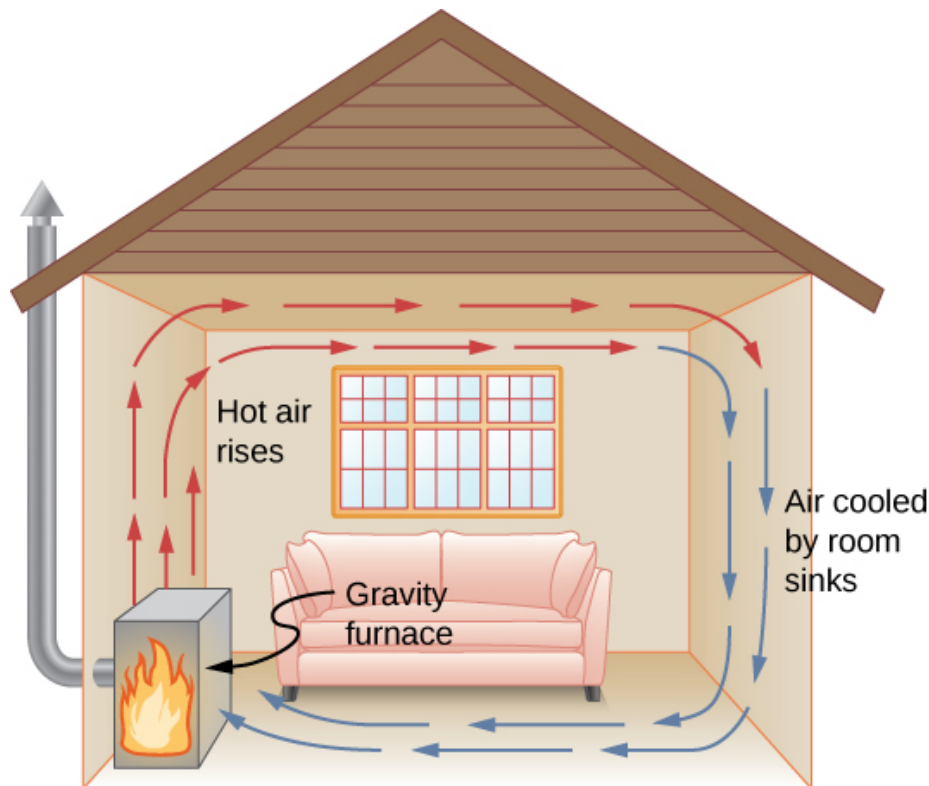
$$P_{\text{final}} = \frac{kA_{\text{final}}(T_h - T_c)}{d_{\text{final}}} = \frac{k(4A_{\text{final}})(T_h - T_c)}{2d_{\text{initial}}} = 2 \frac{kA_{\text{final}}(T_h - T_c)}{d_{\text{initial}}} = 2P_{\text{initial}}.$$

Conduction is caused by the random motion of atoms and molecules. As such, it is an ineffective mechanism for heat transport over macroscopic distances and short times. For example, the temperature on Earth would be unbearably cold during the night and extremely hot during the day if heat transport in the atmosphere were only through conduction. Also, car engines would overheat unless there was a more efficient way to remove excess heat from the pistons. The next module discusses the important heat-transfer mechanism in such situations.

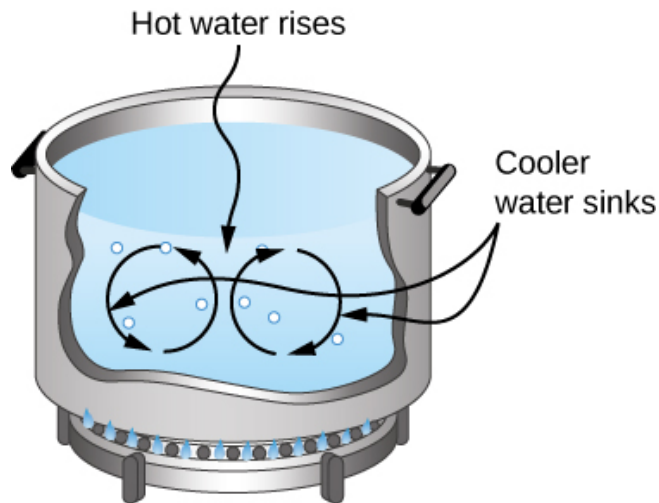
Convection

In convection, thermal energy is carried by the large-scale flow of matter. It can be divided into two types. In *forced convection*, the flow is driven by fans, pumps, and the like. A simple example is a fan that blows air past you in hot surroundings and cools you by replacing the air heated by your body with cooler air. A more complicated example is the cooling system of a typical car, in which a pump moves coolant through the radiator and engine to cool the engine and a fan blows air to cool the radiator.

In *free* or *natural convection*, the flow is driven by buoyant forces: hot fluid rises and cold fluid sinks because density decreases as temperature increases. The house in [\[link\]](#) is kept warm by natural convection, as is the pot of water on the stove in [\[link\]](#). Ocean currents and large-scale atmospheric circulation, which result from the buoyancy of warm air and water, transfer hot air from the tropics toward the poles and cold air from the poles toward the tropics. (Earth's rotation interacts with those flows, causing the observed eastward flow of air in the temperate zones.)



Air heated by a so-called gravity furnace expands and rises, forming a convective loop that transfers energy to other parts of the room. As the air is cooled at the ceiling and outside walls, it contracts, eventually becoming denser than room air and sinking to the floor. A properly designed heating system using natural convection, like this one, can heat a home quite efficiently.



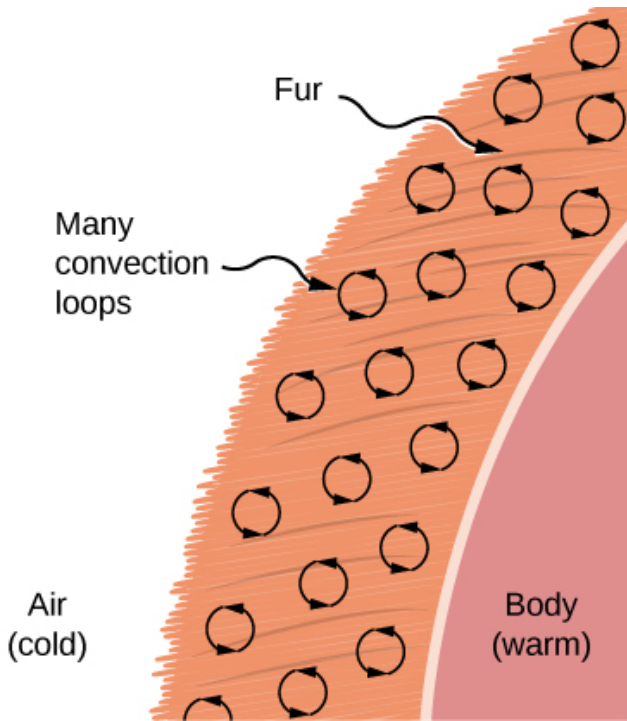
Natural convection plays an important role in heat transfer inside this pot of water. Once conducted to the inside, heat transfer to other parts of the pot is mostly by convection. The hotter water expands, decreases in density, and rises to transfer heat to other regions of the water, while colder water sinks to the bottom. This process keeps repeating.

Note:

Natural convection like that of [link](#) and [link](#), but acting on rock in Earth's mantle, drives [plate tectonics](#) that are the motions that have shaped Earth's surface.

Convection is usually more complicated than conduction. Beyond noting that the convection rate is often approximately proportional to the temperature difference, we will not do any quantitative work comparable to the formula for conduction. However, we can describe convection qualitatively and relate convection rates to heat and time. Air is a poor conductor, so convection dominates heat transfer by air. Therefore, the amount of available space for airflow determines whether air transfers heat rapidly or slowly. There is little heat transfer in a space filled with air with a small amount of other material that prevents flow. The space between the inside and outside walls of a typical American house, for example, is about 9 cm (3.5 in.)—large enough for convection to work effectively. The addition of wall insulation prevents airflow, so heat loss (or gain)

is decreased. On the other hand, the gap between the two panes of a double-paned window is about 1 cm, which largely prevents convection and takes advantage of air's low conductivity reduce heat loss. Fur, cloth, and fiberglass also take advantage of the low conductivity of air by trapping it in spaces too small to support convection ([link](#)).



Fur is filled with air, breaking it up into many small pockets. Convection is very slow here, because the loops are so small. The low conductivity of air makes fur a very good lightweight insulator.

Some interesting phenomena happen when convection is accompanied by a phase change. The combination allows us to cool off by sweating even if the temperature of the surrounding air exceeds body temperature. Heat from the skin is required for sweat to evaporate from the skin, but without air flow, the air becomes saturated and evaporation stops. Air flow caused by convection replaces the saturated air by dry air and evaporation continues.

Example:**Calculating the Flow of Mass during Convection**

The average person produces heat at the rate of about 120 W when at rest. At what rate must water evaporate from the body to get rid of all this energy? (For simplicity, we assume this evaporation occurs when a person is sitting in the shade and surrounding temperatures are the same as skin temperature, eliminating heat transfer by other methods.)

Strategy

Energy is needed for this phase change ($Q = mL_v$). Thus, the energy loss per unit time is

Equation:

$$\frac{Q}{t} = \frac{mL_v}{t} = 120 \text{ W} = 120 \text{ J/s}.$$

We divide both sides of the equation by L_v to find that the mass evaporated per unit time is

Equation:

$$\frac{m}{t} = \frac{120 \text{ J/s}}{L_v}.$$

Solution

Insert the value of the latent heat from [\[link\]](#), $L_v = 2430 \text{ kJ/kg} = 2430 \text{ J/g}$. This yields

Equation:

$$\frac{m}{t} = \frac{120 \text{ J/s}}{2430 \text{ J/g}} = 0.0494 \text{ g/s} = 2.96 \text{ g/min}.$$

Significance

Evaporating about 3 g/min seems reasonable. This would be about 180 g (about 7 oz.) per hour. If the air is very dry, the sweat may evaporate without even being noticed. A significant amount of evaporation also takes place in the lungs and breathing passages.

Another important example of the combination of phase change and convection occurs when water evaporates from the oceans. Heat is removed from the ocean when water evaporates. If the water vapor condenses in liquid droplets as clouds form, possibly far from the ocean, heat is released in the atmosphere. Thus, there is an overall transfer of heat from the ocean to the atmosphere. This process is the driving power behind thunderheads, those great cumulus clouds that rise as much as 20.0 km into the stratosphere ([\[link\]](#)). Water vapor carried in by convection condenses, releasing

tremendous amounts of energy. This energy causes the air to expand and rise to colder altitudes. More condensation occurs in these regions, which in turn drives the cloud even higher. This mechanism is an example of positive feedback, since the process reinforces and accelerates itself. It sometimes produces violent storms, with lightning and hail. The same mechanism drives hurricanes.

Note:

This [time-lapse video](#) shows convection currents in a thunderstorm, including “rolling” motion similar to that of boiling water.



Cumulus clouds are caused by water vapor that rises because of convection. The rise of clouds is driven by a positive feedback mechanism. (credit: “Amada44”/Wikimedia Commons)

Note:

Exercise:

Problem:

Check Your Understanding Explain why using a fan in the summer feels refreshing.

Solution:

Using a fan increases the flow of air: Warm air near your body is replaced by cooler air from elsewhere. Convection increases the rate of heat transfer so that moving air “feels” cooler than still air.

Radiation

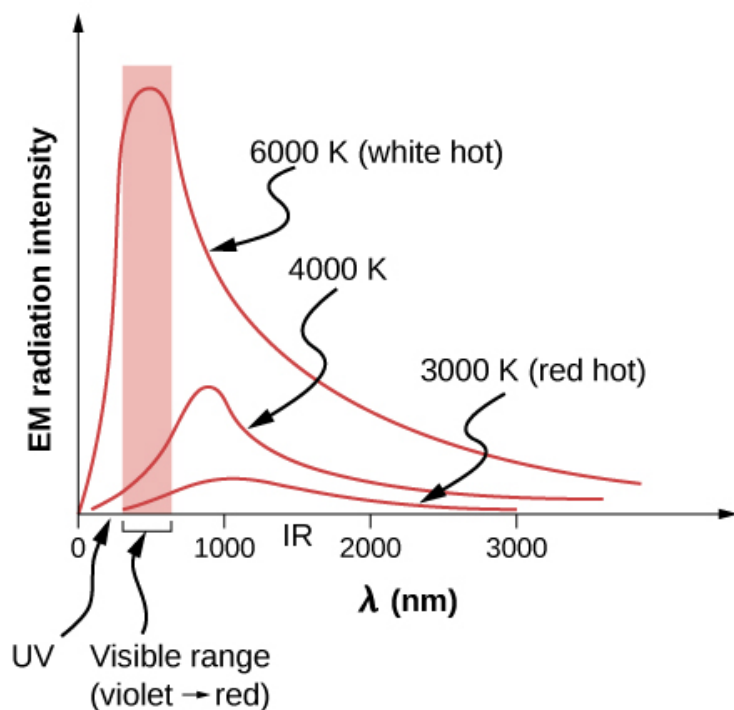
You can feel the heat transfer from the Sun. The space between Earth and the Sun is largely empty, so the Sun warms us without any possibility of heat transfer by convection or conduction. Similarly, you can sometimes tell that the oven is hot without touching its door or looking inside—it may just warm you as you walk by. In these examples, heat is transferred by radiation ([link](#)). That is, the hot body emits electromagnetic waves that are absorbed by the skin. No medium is required for electromagnetic waves to propagate. Different names are used for electromagnetic waves of different wavelengths: radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays.



Most of the heat transfer from this fire to the observers

occurs through infrared radiation. The visible light, although dramatic, transfers relatively little thermal energy. Convection transfers energy away from the observers as hot air rises, while conduction is negligibly slow here. Skin is very sensitive to infrared radiation, so you can sense the presence of a fire without looking at it directly. (credit: Daniel O'Neil)

The energy of electromagnetic radiation varies over a wide range, depending on the wavelength: A shorter wavelength (or higher frequency) corresponds to a higher energy. Because more heat is radiated at higher temperatures, higher temperatures produce more intensity at every wavelength but especially at shorter wavelengths. In visible light, wavelength determines color—red has the longest wavelength and violet the shortest—so a temperature change is accompanied by a color change. For example, an electric heating element on a stove glows from red to orange, while the higher-temperature steel in a blast furnace glows from yellow to white. Infrared radiation is the predominant form radiated by objects cooler than the electric element and the steel. The radiated energy as a function of wavelength depends on its intensity, which is represented in [\[link\]](#) by the height of the distribution. ([Electromagnetic Waves](#) explains more about the electromagnetic spectrum, and [Photons and Matter Waves](#) discusses why the decrease in wavelength corresponds to an increase in energy.)



(a)

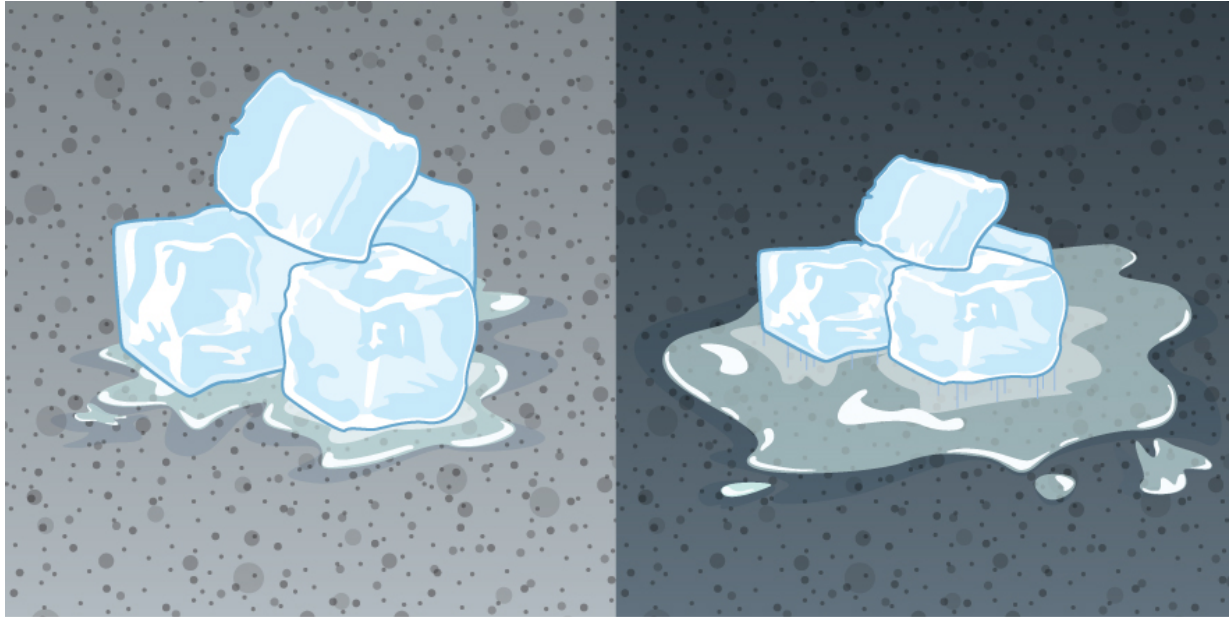


(b)

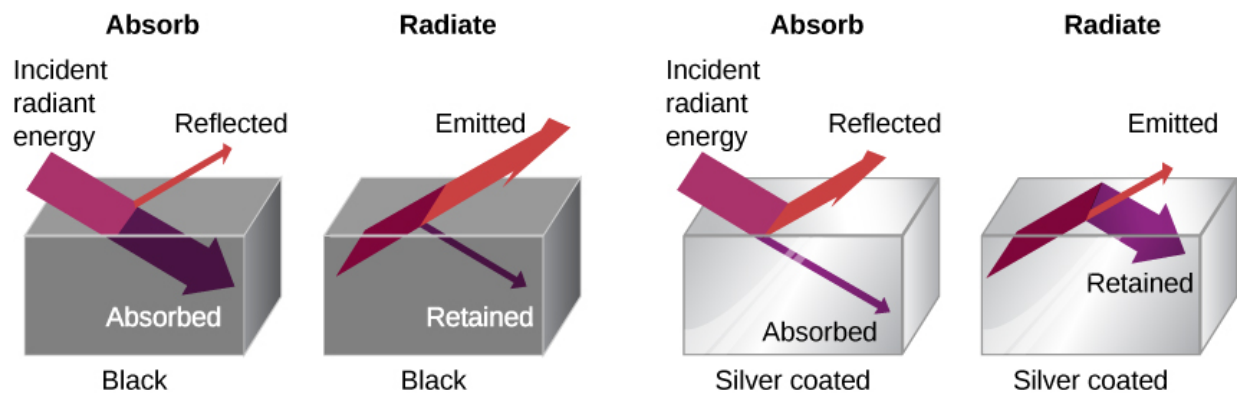
(a) A graph of the spectrum of electromagnetic waves emitted from an ideal radiator at three different temperatures. The intensity or rate of radiation emission increases dramatically with temperature, and the spectrum shifts down in wavelength toward the visible and ultraviolet parts of the spectrum. The shaded portion denotes the visible part of the spectrum. It is apparent that the shift toward the ultraviolet with temperature makes the visible appearance shift from red to white to blue as temperature increases. (b) Note the variations in color corresponding to variations in flame temperature.

The rate of heat transfer by radiation also depends on the object's color. Black is the most effective, and white is the least effective. On a clear summer day, black asphalt in a parking lot is hotter than adjacent gray sidewalk, because black absorbs better than gray ([link](#)). The reverse is also true—black radiates better than gray. Thus, on a clear summer night, the asphalt is colder than the gray sidewalk, because black radiates the energy more rapidly than gray. A perfectly black object would be an *ideal radiator* and an *ideal absorber*, as it would capture all the radiation that falls on it. In contrast, a perfectly white object or a perfect mirror would reflect all radiation, and a perfectly transparent object would transmit it all ([link](#)). Such objects would not emit any radiation. Mathematically, the color is represented by the **emissivity** e . A “blackbody” radiator would have an $e = 1$, whereas a perfect reflector or transmitter would have

$e = 0$. For real examples, tungsten light bulb filaments have an e of about 0.5, and carbon black (a material used in printer toner) has an emissivity of about 0.95.



The darker pavement is hotter than the lighter pavement (much more of the ice on the right has melted), although both have been in the sunlight for the same time.
The thermal conductivities of the pavements are the same.



A black object is a good absorber and a good radiator, whereas a white, clear, or silver object is a poor absorber and a poor radiator.

To see that, consider a silver object and a black object that can exchange heat by radiation and are in thermal equilibrium. We know from experience that they will stay in equilibrium (the result of a principle that will be discussed at length in [Second Law of Thermodynamics](#)). For the black object's temperature to stay constant, it must emit as much radiation as it absorbs, so it must be as good at radiating as absorbing. Similar considerations show that the silver object must radiate as little as it absorbs. Thus, one property, emissivity, controls both radiation and absorption.

Finally, the radiated heat is proportional to the object's surface area, since every part of the surface radiates. If you knock apart the coals of a fire, the radiation increases noticeably due to an increase in radiating surface area.

The rate of heat transfer by emitted radiation is described by the **Stefan-Boltzmann law of radiation**:

Equation:

$$P = \sigma A e T^4,$$

where $\sigma = 5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant, a combination of fundamental constants of nature; A is the surface area of the object; and T is its temperature in kelvins.

The proportionality to the *fourth power* of the absolute temperature is a remarkably strong temperature dependence. It allows the detection of even small temperature variations. Images called *thermographs* can be used medically to detect regions of abnormally high temperature in the body, perhaps indicative of disease. Similar techniques can be used to detect heat leaks in homes ([\[link\]](#)), optimize performance of blast furnaces, improve comfort levels in work environments, and even remotely map Earth's temperature profile.



A thermograph of part of a building shows temperature variations, indicating where heat transfer to the outside is most severe. Windows are a major region of heat transfer to the outside of homes. (credit: US Army)

The Stefan-Boltzmann equation needs only slight refinement to deal with a simple case of an object's absorption of radiation from its surroundings. Assuming that an object with a temperature T_1 is surrounded by an environment with uniform temperature T_2 , the **net rate of heat transfer by radiation** is

Note:

Equation:

$$P_{\text{net}} = \sigma e A (T_2^4 - T_1^4),$$

where e is the emissivity of the object alone. In other words, it does not matter whether the surroundings are white, gray, or black: The balance of radiation into and out of the object depends on how well it emits and absorbs radiation. When $T_2 > T_1$, the quantity P_{net} is positive, that is, the net heat transfer is from hot to cold.

Before doing an example, we have a complication to discuss: different emissivities at different wavelengths. If the fraction of incident radiation an object reflects is the same at all visible wavelengths, the object is gray; if the fraction depends on the wavelength, the object has some other color. For instance, a red or reddish object reflects red light more strongly than other visible wavelengths. Because it absorbs less red, it radiates less red when hot. Differential reflection and absorption of wavelengths outside the visible range have no effect on what we see, but they may have physically important effects. Skin is a very good absorber and emitter of infrared radiation, having an emissivity of 0.97 in the infrared spectrum. Thus, in spite of the obvious variations in skin color, we are all nearly black in the infrared. This high infrared emissivity is why we can so easily feel radiation on our skin. It is also the basis for the effectiveness of night-vision scopes used by law enforcement and the military to detect human beings.

Example:**Calculating the Net Heat Transfer of a Person**

What is the rate of heat transfer by radiation of an unclothed person standing in a dark room whose ambient temperature is 22.0°C ? The person has a normal skin temperature of 33.0°C and a surface area of 1.50 m^2 . The emissivity of skin is 0.97 in the infrared, the part of the spectrum where the radiation takes place.

Strategy

We can solve this by using the equation for the rate of radiative heat transfer.

Solution

Insert the temperature values $T_2 = 295\text{ K}$ and $T_1 = 306\text{ K}$, so that

Equation:

$$\begin{aligned}\frac{Q}{t} &= \sigma e A (T_2^4 - T_1^4) \\ &= (5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4) (0.97) (1.50 \text{ m}^2) [(295 \text{ K})^4 - (306 \text{ K})^4] \\ &= -99 \text{ J/s} = -99 \text{ W}.\end{aligned}$$

Significance

This value is a significant rate of heat transfer to the environment (note the minus sign), considering that a person at rest may produce energy at the rate of 125 W and that conduction and convection are also transferring energy to the environment. Indeed, we would probably expect this person to feel cold. Clothing significantly reduces heat transfer to the environment by all mechanisms, because clothing slows down both

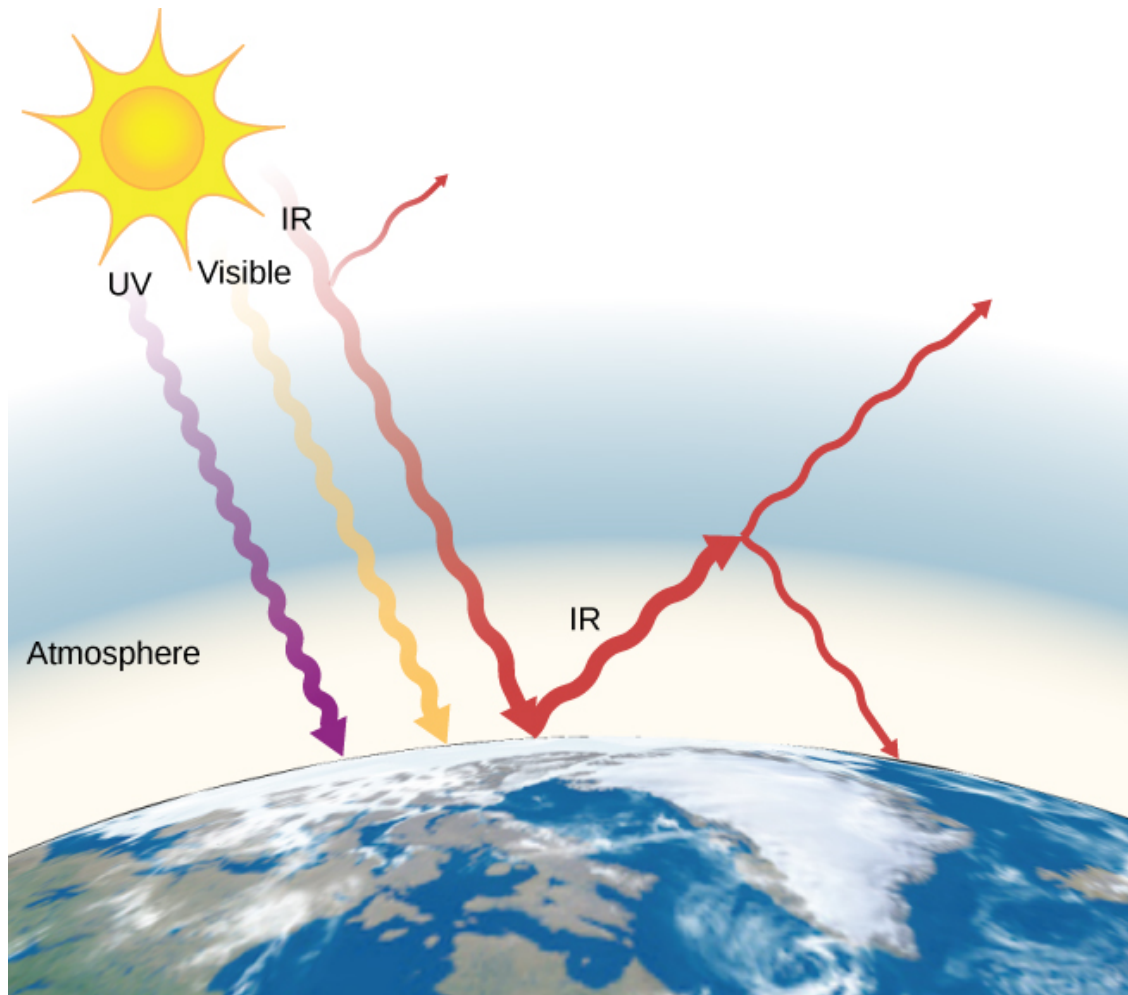
conduction and convection, and has a lower emissivity (especially if it is light-colored) than skin.

The average temperature of Earth is the subject of much current discussion. Earth is in radiative contact with both the Sun and dark space, so we cannot use the equation for an environment at a uniform temperature. Earth receives almost all its energy from radiation of the Sun and reflects some of it back into outer space. Conversely, dark space is very cold, about 3 K, so that Earth radiates energy into the dark sky. The rate of heat transfer from soil and grasses can be so rapid that frost may occur on clear summer evenings, even in warm latitudes.

The average temperature of Earth is determined by its energy balance. To a first approximation, it is the temperature at which Earth radiates heat to space as fast as it receives energy from the Sun.

An important parameter in calculating the temperature of Earth is its emissivity (e). On average, it is about 0.65, but calculation of this value is complicated by the great day-to-day variation in the highly reflective cloud coverage. Because clouds have lower emissivity than either oceans or land masses, they emit some of the radiation back to the surface, greatly reducing heat transfer into dark space, just as they greatly reduce heat transfer into the atmosphere during the day. There is negative feedback (in which a change produces an effect that opposes that change) between clouds and heat transfer; higher temperatures evaporate more water to form more clouds, which reflect more radiation back into space, reducing the temperature.

The often-mentioned **greenhouse effect** is directly related to the variation of Earth's emissivity with wavelength ([link](#)). The greenhouse effect is a natural phenomenon responsible for providing temperatures suitable for life on Earth and for making Venus unsuitable for human life. Most of the infrared radiation emitted from Earth is absorbed by carbon dioxide (CO₂) and water (H₂O) in the atmosphere and then re-radiated into outer space or back to Earth. Re-radiation back to Earth maintains its surface temperature about 40 °C higher than it would be if there were no atmosphere. (The glass walls and roof of a greenhouse increase the temperature inside by blocking convective heat losses, not radiative losses.)



The greenhouse effect is the name given to the increase of Earth's temperature due to absorption of radiation in the atmosphere. The atmosphere is transparent to incoming visible radiation and most of the Sun's infrared. The Earth absorbs that energy and re-emits it. Since Earth's temperature is much lower than the Sun's, it re-emits the energy at much longer wavelengths, in the infrared. The atmosphere absorbs much of that infrared radiation and radiates about half of the energy back down, keeping Earth warmer than it would otherwise be. The amount of trapping depends on concentrations of trace gases such as carbon dioxide, and an increase in the concentration of these gases increases Earth's surface temperature.

The greenhouse effect is central to the discussion of global warming due to emission of carbon dioxide and methane (and other greenhouse gases) into Earth's atmosphere from industry, transportation, and farming. Changes in global climate could lead to more

intense storms, precipitation changes (affecting agriculture), reduction in rain forest biodiversity, and rising sea levels.

Note:

You can explore [a simulation of the greenhouse effect](#) that takes the point of view that the atmosphere scatters (redirects) infrared radiation rather than absorbing it and reradiating it. You may want to run the simulation first with no greenhouse gases in the atmosphere and then look at how adding greenhouse gases affects the infrared radiation from the Earth and the Earth's temperature.

Note:

Effects of Heat Transfer

1. Examine the situation to determine what type of heat transfer is involved.
2. Identify the type(s) of heat transfer—conduction, convection, or radiation.
3. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.
4. Make a list of what is given or what can be inferred from the problem as stated (identify the knowns).
5. Solve the appropriate equation for the quantity to be determined (the unknown).
6. For conduction, use the equation $P = \frac{kA\Delta T}{d}$. [\[link\]](#) lists thermal conductivities. For convection, determine the amount of matter moved and the equation $Q = mc\Delta T$, along with $Q = mL_f$ or $Q = mL_v$ if a substance changes phase. For radiation, the equation $P_{\text{net}} = \sigma eA (T_2^4 - T_1^4)$ gives the net heat transfer rate.
7. Substitute the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.
8. Check the answer to see if it is reasonable. Does it make sense?

Note:

Exercise:

Problem:

Check Your Understanding How much greater is the rate of heat radiation when a body is at the temperature 40°C than when it is at the temperature 20°C ?

Solution:

The radiated heat is proportional to the fourth power of the *absolute temperature*. Because $T_1 = 293\text{ K}$ and $T_2 = 313\text{ K}$, the rate of heat transfer increases by about 30% of the original rate.

Summary

- Heat is transferred by three different methods: conduction, convection, and radiation.
- Heat conduction is the transfer of heat between two objects in direct contact with each other.
- The rate of heat transfer P (energy per unit time) is proportional to the temperature difference $T_h - T_c$ and the contact area A and inversely proportional to the distance d between the objects.
- Convection is heat transfer by the macroscopic movement of mass. Convection can be natural or forced, and generally transfers thermal energy faster than conduction. Convection that occurs along with a phase change can transfer energy from cold regions to warm ones.
- Radiation is heat transfer through the emission or absorption of electromagnetic waves.
- The rate of radiative heat transfer is proportional to the emissivity e . For a perfect blackbody, $e = 1$, whereas a perfectly white, clear, or reflective body has $e = 0$, with real objects having values of e between 1 and 0.
- The rate of heat transfer depends on the surface area and the fourth power of the absolute temperature:

Equation:

$$P = \sigma e A T^4,$$

where $\sigma = 5.67 \times 10^{-8}\text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant and e is the emissivity of the body. The net rate of heat transfer from an object by radiation is

Equation:

$$\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4),$$

where T_1 is the temperature of the object surrounded by an environment with uniform temperature T_2 and e is the emissivity of the object.

Key Equations

Linear thermal expansion	$\Delta L = \alpha L \Delta T$
Thermal expansion in two dimensions	$\Delta A = 2\alpha A \Delta T$
Thermal expansion in three dimensions	$\Delta V = \beta V \Delta T$
Heat transfer	$Q = mc\Delta T$
Transfer of heat in a calorimeter	$Q_{\text{cold}} + Q_{\text{hot}} = 0$
Heat due to phase change (melting and freezing)	$Q = mL_f$
Heat due to phase change (evaporation and condensation)	$Q = mL_v$
Rate of conductive heat transfer	$P = \frac{kA(T_h - T_c)}{d}$
Net rate of heat transfer by radiation	$P_{\text{net}} = \sigma eA (T_2^4 - T_1^4)$

Conceptual Questions

Exercise:

Problem:

What are the main methods of heat transfer from the hot core of Earth to its surface? From Earth's surface to outer space?

Exercise:

Problem:

When our bodies get too warm, they respond by sweating and increasing blood circulation to the surface to transfer thermal energy away from the core. What effect will those processes have on a person in a 40.0-°C hot tub?

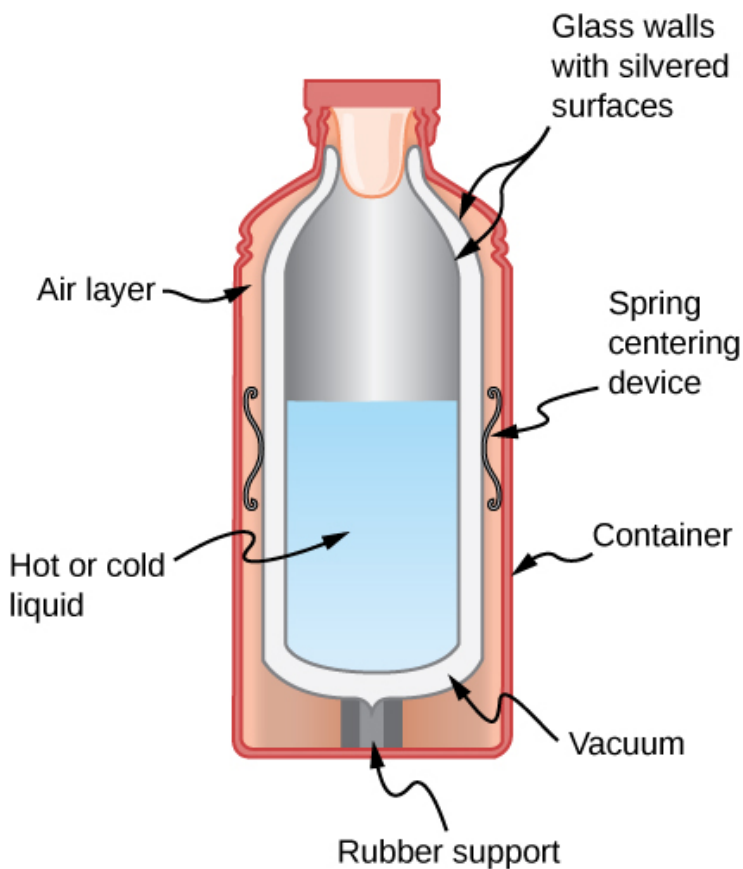
Solution:

Increasing circulation to the surface will warm the person, as the temperature of the water is warmer than human body temperature. Sweating will cause no evaporative cooling under water or in the humid air immediately above the tub.

Exercise:

Problem:

Shown below is a cut-away drawing of a thermos bottle (also known as a Dewar flask), which is a device designed specifically to slow down all forms of heat transfer. Explain the functions of the various parts, such as the vacuum, the silvering of the walls, the thin-walled long glass neck, the rubber support, the air layer, and the stopper.



Exercise:

Problem:

Some electric stoves have a flat ceramic surface with heating elements hidden beneath. A pot placed over a heating element will be heated, while the surface only a few centimeters away is safe to touch. Why is ceramic, with a conductivity less than that of a metal but greater than that of a good insulator, an ideal choice for the stove top?

Solution:

It spread the heat over the area above the heating elements, evening the temperature there, but does not spread the heat much beyond the heating elements.

Exercise:**Problem:**

Loose-fitting white clothing covering most of the body, shown below, is ideal for desert dwellers, both in the hot Sun and during cold evenings. Explain how such clothing is advantageous during both day and night.



Exercise:

Problem:

One way to make a fireplace more energy-efficient is to have room air circulate around the outside of the fire box and back into the room. Detail the methods of heat transfer involved.

Solution:

Heat is conducted from the fire through the fire box to the circulating air and then convected by the air into the room (forced convection).

Exercise:

Problem:

On cold, clear nights horses will sleep under the cover of large trees. How does this help them keep warm?

Exercise:**Problem:**

When watching a circus during the day in a large, dark-colored tent, you sense significant heat transfer from the tent. Explain why this occurs.

Solution:

The tent is heated by the Sun and transfers heat to you by all three processes, especially radiation.

Exercise:**Problem:**

Satellites designed to observe the radiation from cold (3 K) dark space have sensors that are shaded from the Sun, Earth, and the Moon and are cooled to very low temperatures. Why must the sensors be at low temperature?

Exercise:**Problem:**

Why are thermometers that are used in weather stations shielded from the sunshine? What does a thermometer measure if it is shielded from the sunshine? What does it measure if it is not?

Solution:

If shielded, it measures the air temperature. If not, it measures the combined effect of air temperature and net radiative heat gain from the Sun.

Exercise:**Problem:**

Putting a lid on a boiling pot greatly reduces the heat transfer necessary to keep it boiling. Explain why.

Exercise:

Problem:

Your house will be empty for a while in cold weather, and you want to save energy and money. Should you turn the thermostat down to the lowest level that will protect the house from damage such as freezing pipes, or leave it at the normal temperature? (If you don't like coming back to a cold house, imagine that a timer controls the heating system so the house will be warm when you get back.) Explain your answer.

Solution:

Turn the thermostat down. To have the house at the normal temperature, the heating system must replace all the heat that was lost. For all three mechanisms of heat transfer, the greater the temperature difference between inside and outside, the more heat is lost and must be replaced. So the house should be at the lowest temperature that does not allow freezing damage.

Exercise:**Problem:**

You pour coffee into an unlidged cup, intending to drink it 5 minutes later. You can add cream when you pour the cup or right before you drink it. (The cream is at the same temperature either way. Assume that the cream and coffee come into thermal equilibrium with each other very quickly.) Which way will give you hotter coffee? What feature of this question is different from the previous one?

Exercise:**Problem:**

Broiling is a method of cooking by radiation, which produces somewhat different results from cooking by conduction or convection. A gas flame or electric heating element produces a very high temperature close to the food and *above* it. Why is radiation the dominant heat-transfer method in this situation?

Solution:

Air is a good insulator, so there is little conduction, and the heated air rises, so there is little convection downward.

Exercise:**Problem:**

On a cold winter morning, why does the metal of a bike feel colder than the wood of a porch?

Problems

Exercise:

Problem:

(a) Calculate the rate of heat conduction through house walls that are 13.0 cm thick and have an average thermal conductivity twice that of glass wool. Assume there are no windows or doors. The walls' surface area is 120 m^2 and their inside surface is at 18.0°C , while their outside surface is at 5.00°C . (b) How many 1-kW room heaters would be needed to balance the heat transfer due to conduction?

Solution:

a. $1.01 \times 10^3 \text{ W}$; b. One 1-kilowatt room heater is needed.

Exercise:

Problem:

The rate of heat conduction out of a window on a winter day is rapid enough to chill the air next to it. To see just how rapidly the windows transfer heat by conduction, calculate the rate of conduction in watts through a 3.00-m^2 window that is 0.634 cm thick ($1/4$ in.) if the temperatures of the inner and outer surfaces are 5.00°C and -10.0°C , respectively. (This rapid rate will not be maintained—the inner surface will cool, even to the point of frost formation.)

Exercise:

Problem:

Calculate the rate of heat conduction out of the human body, assuming that the core internal temperature is 37.0°C , the skin temperature is 34.0°C , the thickness of the fatty tissues between the core and the skin averages 1.00 cm, and the surface area is 1.40 m^2 .

Solution:

84.0 W

Exercise:

Problem:

Suppose you stand with one foot on ceramic flooring and one foot on a wool carpet, making contact over an area of 80.0 cm^2 with each foot. Both the ceramic and the carpet are 2.00 cm thick and are 10.0°C on their bottom sides. At what rate must heat transfer occur from each foot to keep the top of the ceramic and carpet at 33.0°C ?

Exercise:**Problem:**

A man consumes 3000 kcal of food in one day, converting most of it to thermal energy to maintain body temperature. If he loses half this energy by evaporating water (through breathing and sweating), how many kilograms of water evaporate?

Solution:

2.59 kg

Exercise:**Problem:**

A firewalker runs across a bed of hot coals without sustaining burns. Calculate the heat transferred by conduction into the sole of one foot of a firewalker given that the bottom of the foot is a 3.00-mm -thick callus with a conductivity at the low end of the range for wood and its density is 300 kg/m^3 . The area of contact is 25.0 cm^2 , the temperature of the coals is 700°C , and the time in contact is 1.00 s . Ignore the evaporative cooling of sweat.

Exercise:**Problem:**

(a) What is the rate of heat conduction through the 3.00-cm -thick fur of a large animal having a 1.40-m^2 surface area? Assume that the animal's skin temperature is 32.0°C , that the air temperature is -5.00°C , and that fur has the same thermal conductivity as air. (b) What food intake will the animal need in one day to replace this heat transfer?

Solution:

a. 39.7 W ; b. 820 kcal

Exercise:

Problem:

A walrus transfers energy by conduction through its blubber at the rate of 150 W when immersed in $-1.00\text{ }^{\circ}\text{C}$ water. The walrus's internal core temperature is $37.0\text{ }^{\circ}\text{C}$, and it has a surface area of 2.00 m^2 . What is the average thickness of its blubber, which has the conductivity of fatty tissues without blood?

Exercise:**Problem:**

Compare the rate of heat conduction through a 13.0-cm-thick wall that has an area of 10.0 m^2 and a thermal conductivity twice that of glass wool with the rate of heat conduction through a 0.750-cm-thick window that has an area of 2.00 m^2 , assuming the same temperature difference across each.

Solution:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}, \text{ so that}$$

$$\frac{(Q/t)_{\text{wall}}}{(Q/t)_{\text{window}}} = \frac{k_{\text{wall}}A_{\text{wall}}d_{\text{window}}}{k_{\text{window}}A_{\text{window}}d_{\text{wall}}} = \frac{(2 \times 0.042\text{ J/s}\cdot\text{m}\cdot^{\circ}\text{C})(10.0\text{ m}^2)(0.750 \times 10^{-2}\text{ m})}{(0.84\text{ J/s}\cdot\text{m}\cdot^{\circ}\text{C})(2.00\text{ m}^2)(13.0 \times 10^{-2}\text{ m})}$$

This gives 0.0288 wall: window, or 35:1 window: wall

Exercise:**Problem:**

Suppose a person is covered head to foot by wool clothing with average thickness of 2.00 cm and is transferring energy by conduction through the clothing at the rate of 50.0 W. What is the temperature difference across the clothing, given the surface area is 1.40 m^2 ?

Exercise:**Problem:**

Some stove tops are smooth ceramic for easy cleaning. If the ceramic is 0.600 cm thick and heat conduction occurs through the same area and at the same rate as computed in [\[link\]](#), what is the temperature difference across it? Ceramic has the same thermal conductivity as glass and brick.

Solution:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d} = \frac{kA\Delta T}{d} \Rightarrow$$

$$\Delta T = \frac{d(Q/t)}{kA} = \frac{(6.00 \times 10^{-3}\text{ m})(2256\text{ W})}{(0.84\text{ J/s}\cdot\text{m}\cdot^{\circ}\text{C})(1.54 \times 10^{-2}\text{ m}^2)} = 1046\text{ }^{\circ}\text{C} = 1.05 \times 10^3\text{ K}$$

Exercise:**Problem:**

One easy way to reduce heating (and cooling) costs is to add extra insulation in the attic of a house. Suppose a single-story cubical house already had 15 cm of fiberglass insulation in the attic and in all the exterior surfaces. If you added an extra 8.0 cm of fiberglass to the attic, by what percentage would the heating cost of the house drop? Take the house to have dimensions 10 m by 15 m by 3.0 m. Ignore air infiltration and heat loss through windows and doors, and assume that the interior is uniformly at one temperature and the exterior is uniformly at another.

Exercise:**Problem:**

Many decisions are made on the basis of the payback period: the time it will take through savings to equal the capital cost of an investment. Acceptable payback times depend upon the business or philosophy one has. (For some industries, a payback period is as small as 2 years.) Suppose you wish to install the extra insulation in the preceding problem. If energy cost \$1.00 per million joules and the insulation was \$4.00 per square meter, then calculate the simple payback time. Take the average ΔT for the 120-day heating season to be 15.0 °C.

Solution:

We found in the preceding problem that $P = 126\Delta T \text{ W} \cdot ^\circ\text{C}$ as baseline energy use. So the total heat loss during this period is
 $Q = (126 \text{ J/s} \cdot ^\circ\text{C}) (15.0 ^\circ\text{C}) (120 \text{ days}) (86.4 \times 10^3 \text{ s/day}) = 1960 \times 10^6 \text{ J}$
. At the cost of \$1/MJ, the cost is \$1960. From an earlier problem, the savings is 12% or \$235/y. We need 150 m² of insulation in the attic. At \$4/m², this is a \$500 cost. So the payback period is $\$600 / (\$235/\text{y}) = 2.6 \text{ years}$ (excluding labor costs).

Additional Problems**Exercise:****Problem:**

In 1701, the Danish astronomer Ole Rømer proposed a temperature scale with two fixed points, freezing water at 7.5 degrees, and boiling water at 60.0 degrees. What is the boiling point of oxygen, 90.2 K, on the Rømer scale?

Exercise:

Problem:

What is the percent error of thinking the melting point of tungsten is 3695°C instead of the correct value of 3695 K ?

Solution:

$$7.39\%$$

Exercise:**Problem:**

An engineer wants to design a structure in which the difference in length between a steel beam and an aluminum beam remains at 0.500 m regardless of temperature, for ordinary temperatures. What must the lengths of the beams be?

Exercise:**Problem:**

How much stress is created in a steel beam if its temperature changes from -15°C to 40°C but it cannot expand? For steel, the Young's modulus $Y = 210 \times 10^9\text{ N/m}^2$ from [Stress, Strain, and Elastic Modulus](#). (Ignore the change in area resulting from the expansion.)

Solution:

$$\frac{F}{A} = (210 \times 10^9\text{ Pa}) (12 \times 10^{-6}/^{\circ}\text{C}) (40^{\circ}\text{C} - (-15^{\circ}\text{C})) = 1.4 \times 10^8\text{ N/m}^2$$

Exercise:**Problem:**

A brass rod ($Y = 90 \times 10^9\text{ N/m}^2$), with a diameter of 0.800 cm and a length of 1.20 m when the temperature is 25°C , is fixed at both ends. At what temperature is the force in it at $36,000\text{ N}$?

Exercise:

Problem:

A mercury thermometer still in use for meteorology has a bulb with a volume of 0.780 cm^3 and a tube for the mercury to expand into of inside diameter 0.130 mm . (a) Neglecting the thermal expansion of the glass, what is the spacing between marks 1°C apart? (b) If the thermometer is made of ordinary glass (not a good idea), what is the spacing?

Solution:

a. 1.06 cm ; b. 1.11 cm

Exercise:**Problem:**

Even when shut down after a period of normal use, a large commercial nuclear reactor transfers thermal energy at the rate of 150 MW by the radioactive decay of fission products. This heat transfer causes a rapid increase in temperature if the cooling system fails ($1 \text{ watt} = 1 \text{ joule/second}$ or $1 \text{ W} = 1 \text{ J/s}$ and $1 \text{ MW} = 1 \text{ megawatt}$). (a) Calculate the rate of temperature increase in degrees Celsius per second ($^\circ \text{C/s}$) if the mass of the reactor core is $1.60 \times 10^5 \text{ kg}$ and it has an average specific heat of $0.3349 \text{ kJ/kg} \cdot ^\circ \text{C}$. (b) How long would it take to obtain a temperature increase of 2000°C , which could cause some metals holding the radioactive materials to melt? (The initial rate of temperature increase would be greater than that calculated here because the heat transfer is concentrated in a smaller mass. Later, however, the temperature increase would slow down because the $500,000\text{-kg}$ steel containment vessel would also begin to heat up.)

Exercise:

Problem:

You leave a pastry in the refrigerator on a plate and ask your roommate to take it out before you get home so you can eat it at room temperature, the way you like it. Instead, your roommate plays video games for hours. When you return, you notice that the pastry is still cold, but the game console has become hot. Annoyed, and knowing that the pastry will not be good if it is microwaved, you warm up the pastry by unplugging the console and putting it in a clean trash bag (which acts as a perfect calorimeter) with the pastry on the plate. After a while, you find that the equilibrium temperature is a nice, warm 38.3°C . You know that the game console has a mass of 2.1 kg . Approximate it as having a uniform initial temperature of 45°C . The pastry has a mass of 0.16 kg and a specific heat of $3.0\text{ kJ}/(\text{kg} \cdot ^\circ\text{C})$, and is at a uniform initial temperature of 4.0°C . The plate is at the same temperature and has a mass of 0.24 kg and a specific heat of $0.90\text{ kJ}/(\text{kg} \cdot ^\circ\text{C})$. What is the specific heat of the console?

Solution:

$$1.7\text{ kJ}/(\text{kg} \cdot ^\circ\text{C})$$

Exercise:**Problem:**

Two solid spheres, A and B , made of the same material, are at temperatures of 0°C and 100°C , respectively. The spheres are placed in thermal contact in an ideal calorimeter, and they reach an equilibrium temperature of 20°C . Which is the bigger sphere? What is the ratio of their diameters?

Exercise:**Problem:**

In some countries, liquid nitrogen is used on dairy trucks instead of mechanical refrigerators. A 3.00-hour delivery trip requires 200 L of liquid nitrogen, which has a density of $808\text{ kg}/\text{m}^3$. (a) Calculate the heat transfer necessary to evaporate this amount of liquid nitrogen and raise its temperature to 3.00°C . (Use c_p and assume it is constant over the temperature range.) This value is the amount of cooling the liquid nitrogen supplies. (b) What is this heat transfer rate in kilowatt-hours? (c) Compare the amount of cooling obtained from melting an identical mass of 0°C ice with that from evaporating the liquid nitrogen.

Solution:

$$\text{a. } 1.57 \times 10^4\text{ kcal; b. } 18.3\text{ kW} \cdot \text{h; c. } 1.29 \times 10^4\text{ kcal}$$

Exercise:

Problem:

Some gun fanciers make their own bullets, which involves melting lead and casting it into lead slugs. How much heat transfer is needed to raise the temperature and melt 0.500 kg of lead, starting from 25.0 °C?

Exercise:

Problem:

A 0.800-kg iron cylinder at a temperature of 1.00×10^3 °C is dropped into an insulated chest of 1.00 kg of ice at its melting point. What is the final temperature, and how much ice has melted?

Solution:

6.3 °C. All of the ice melted.

Exercise:

Problem: Repeat the preceding problem with 2.00 kg of ice instead of 1.00 kg.

Exercise:

Problem:

Repeat the preceding problem with 0.500 kg of ice, assuming that the ice is initially in a copper container of mass 1.50 kg in equilibrium with the ice.

Solution:

63.9 °C, all the ice melted

Exercise:

Problem:

A 30.0-g ice cube at its melting point is dropped into an aluminum calorimeter of mass 100.0 g in equilibrium at 24.0 °C with 300.0 g of an unknown liquid. The final temperature is 4.0 °C. What is the heat capacity of the liquid?

Exercise:

Problem:

(a) Calculate the rate of heat conduction through a double-paned window that has a 1.50-m^2 area and is made of two panes of 0.800-cm -thick glass separated by a 1.00-cm air gap. The inside surface temperature is 15.0°C , while that on the outside is -10.0°C . (*Hint:* There are identical temperature drops across the two glass panes. First find these and then the temperature drop across the air gap. This problem ignores the increased heat transfer in the air gap due to convection.) (b) Calculate the rate of heat conduction through a 1.60-cm -thick window of the same area and with the same temperatures. Compare your answer with that for part (a).

Solution:

a. 83 W ; b. $1.97 \times 10^3\text{ W}$; The single-pane window has a rate of heat conduction equal to $1969/83$, or 24 times that of a double-pane window.

Exercise:**Problem:**

(a) An exterior wall of a house is 3 m tall and 10 m wide. It consists of a layer of drywall with an R factor of 0.56 , a layer 3.5 inches thick filled with fiberglass batts, and a layer of insulated siding with an R factor of 2.6 . The wall is built so well that there are no leaks of air through it. When the inside of the wall is at 22°C and the outside is at -2°C , what is the rate of heat flow through the wall? (b) More realistically, the 3.5 -inch space also contains 2-by-4 studs—wooden boards 1.5 inches by 3.5 inches oriented so that 3.5 -inch dimension extends from the drywall to the siding. They are “on 16-inch centers,” that is, the centers of the studs are 16 inches apart. What is the heat current in this situation? Don’t worry about one stud more or less.

Exercise:**Problem:**

For the human body, what is the rate of heat transfer by conduction through the body’s tissue with the following conditions: the tissue thickness is 3.00 cm , the difference in temperature is 2.00°C , and the skin area is 1.50 m^2 . How does this compare with the average heat transfer rate to the body resulting from an energy intake of about 2400 kcal per day? (No exercise is included.)

Solution:

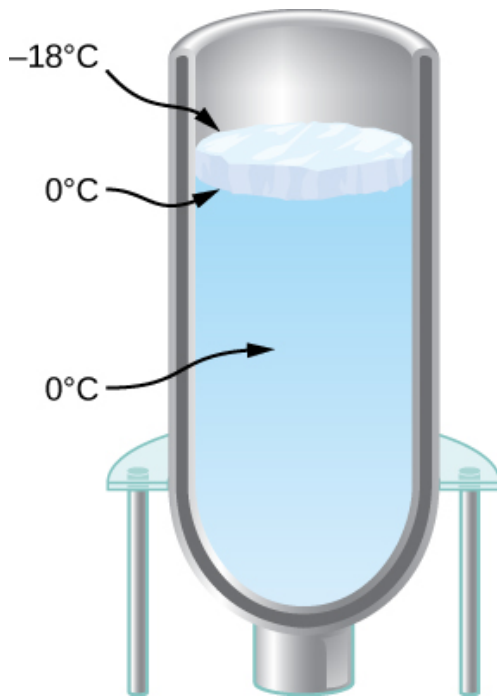
The rate of heat transfer by conduction is 20.0 W . On a daily basis, this is $1,728\text{ kJ/day}$. Daily food intake is $2400\text{ kcal/d} \times 4186\text{ J/kcal} = 10,050\text{ kJ/day}$. So

only 17.2% of energy intake goes as heat transfer by conduction to the environment at this ΔT .

Exercise:

Problem:

You have a Dewar flask (a laboratory vacuum flask) that has an open top and straight sides, as shown below. You fill it with water and put it into the freezer. It is effectively a perfect insulator, blocking all heat transfer, except on the top. After a time, ice forms on the surface of the water. The liquid water and the bottom surface of the ice, in contact with the liquid water, are at 0°C . The top surface of the ice is at the same temperature as the air in the freezer, -18°C . Set the rate of heat flow through the ice equal to the rate of loss of heat of fusion as the water freezes. When the ice layer is 0.700 cm thick, find the rate in m/s at which the ice is thickening.



Exercise:

Problem:

An infrared heater for a sauna has a surface area of 0.050 m^2 and an emissivity of 0.84. What temperature must it run at if the required power is 360 W? Neglect the temperature of the environment.

Solution:

620 K

Exercise:**Problem:**

- (a) Determine the power of radiation from the Sun by noting that the intensity of the radiation at the distance of Earth is 1370 W/m^2 . *Hint:* That intensity will be found everywhere on a spherical surface with radius equal to that of Earth's orbit.
- (b) Assuming that the Sun's temperature is 5780 K and that its emissivity is 1, find its radius.

Challenge Problems**Exercise:****Problem:**

A pendulum is made of a rod of length L and negligible mass, but capable of thermal expansion, and a weight of negligible size. (a) Show that when the temperature increases by dT , the period of the pendulum increases by a fraction $\alpha L dT / 2$. (b) A clock controlled by a brass pendulum keeps time correctly at 10°C . If the room temperature is 30°C , does the clock run faster or slower? What is its error in seconds per day?

Solution:

Denoting the period by P , we know $P = 2\pi\sqrt{L/g}$. When the temperature increases by dT , the length increases by $\alpha L dT$. Then the new length is a.

$$P = 2\pi\sqrt{\frac{L+\alpha L dT}{g}} = 2\pi\sqrt{\frac{L}{g}(1 + \alpha dT)} = 2\pi\sqrt{\frac{L}{g}} \left(1 + \frac{1}{2}\alpha dT\right) = P \left(1 + \frac{1}{2}\alpha dT\right)$$

by the binomial expansion. b. The clock runs slower, as its new period is 1.00019 s. It loses 16.4 s per day.

Exercise:**Problem:**

At temperatures of a few hundred kelvins the specific heat capacity of copper approximately follows the empirical formula $c = \alpha + \beta T + \delta T^{-2}$, where $\alpha = 349 \text{ J/kg} \cdot \text{K}$, $\beta = 0.107 \text{ J/kg} \cdot \text{K}^2$, and $\delta = 4.58 \times 10^5 \text{ J} \cdot \text{kg} \cdot \text{K}$. How much heat is needed to raise the temperature of a 2.00-kg piece of copper from 20°C to 250°C ?

Exercise:

Problem:

In a calorimeter of negligible heat capacity, 200 g of steam at $150\text{ }^{\circ}\text{C}$ and 100 g of ice at $-40\text{ }^{\circ}\text{C}$ are mixed. The pressure is maintained at 1 atm. What is the final temperature, and how much steam, ice, and water are present?

Solution:

The amount of heat to melt the ice and raise it to $100\text{ }^{\circ}\text{C}$ is not enough to condense the steam, but it is more than enough to lower the steam's temperature by $50\text{ }^{\circ}\text{C}$, so the final state will consist of steam and liquid water in equilibrium, and the final temperature is $100\text{ }^{\circ}\text{C}$; 9.5 g of steam condenses, so the final state contains 49.5 g of steam and 40.5 g of liquid water.

Exercise:**Problem:**

An astronaut performing an extra-vehicular activity (space walk) shaded from the Sun is wearing a spacesuit that can be approximated as perfectly white ($e = 0$) except for a $5\text{ cm} \times 8\text{ cm}$ patch in the form of the astronaut's national flag. The patch has emissivity 0.300. The spacesuit under the patch is 0.500 cm thick, with a thermal conductivity $k = 0.0600\text{ W/m }^{\circ}\text{C}$, and its inner surface is at a temperature of $20.0\text{ }^{\circ}\text{C}$. What is the temperature of the patch, and what is the rate of heat loss through it? Assume the patch is so thin that its outer surface is at the same temperature as the outer surface of the spacesuit under it. Also assume the temperature of outer space is 0 K. You will get an equation that is very hard to solve in closed form, so you can solve it numerically with a graphing calculator, with software, or even by trial and error with a calculator.

Exercise:**Problem:**

Find the growth of an ice layer as a function of time in a Dewar flask as seen in [\[link\]](#). Call the thickness of the ice layer L . (a) Derive an equation for dL/dt in terms of L , the temperature T above the ice, and the properties of ice (which you can leave in symbolic form instead of substituting the numbers). (b) Solve this differential equation assuming that at $t = 0$, you have $L = 0$. If you have studied differential equations, you will know a technique for solving equations of this type: manipulate the equation to get dL/dt multiplied by a (very simple) function of L on one side, and integrate both sides with respect to time. Alternatively, you may be able to use your knowledge of the derivatives of various functions to guess the solution, which has a simple dependence on t . (c) Will the water eventually freeze to the bottom of the flask?

Solution:

a. $dL/dT = kT/\rho L$; b. $L = \sqrt{2kTt/\rho L_f}$; c. yes

Exercise:**Problem:**

As the very first rudiment of climatology, estimate the temperature of Earth. Assume it is a perfect sphere and its temperature is uniform. Ignore the greenhouse effect. Thermal radiation from the Sun has an intensity (the “solar constant” S) of about 1370 W/m^2 at the radius of Earth’s orbit. (a) Assuming the Sun’s rays are parallel, what area must S be multiplied by to get the total radiation intercepted by Earth? It will be easiest to answer in terms of Earth’s radius, R . (b) Assume that Earth reflects about 30% of the solar energy it intercepts. In other words, Earth has an albedo with a value of $A = 0.3$. In terms of S , A , and R , what is the rate at which Earth absorbs energy from the Sun? (c) Find the temperature at which Earth radiates energy at the same rate. Assume that at the infrared wavelengths where it radiates, the emissivity e is 1. Does your result show that the greenhouse effect is important? (d) How does your answer depend on the the area of Earth?

Exercise:**Problem:**

Let’s stop ignoring the greenhouse effect and incorporate it into the previous problem in a very rough way. Assume the atmosphere is a single layer, a spherical shell around Earth, with an emissivity $e = 0.77$ (chosen simply to give the right answer) at infrared wavelengths emitted by Earth and by the atmosphere. However, the atmosphere is transparent to the Sun’s radiation (that is, assume the radiation is at visible wavelengths with no infrared), so the Sun’s radiation reaches the surface. The greenhouse effect comes from the difference between the atmosphere’s transmission of visible light and its rather strong absorption of infrared. Note that the atmosphere’s radius is not significantly different from Earth’s, but since the atmosphere is a layer above Earth, it emits radiation both upward and downward, so it has twice Earth’s area. There are three radiative energy transfers in this problem: solar radiation absorbed by Earth’s surface; infrared radiation from the surface, which is absorbed by the atmosphere according to its emissivity; and infrared radiation from the atmosphere, half of which is absorbed by Earth and half of which goes out into space. Apply the method of the previous problem to get an equation for Earth’s surface and one for the atmosphere, and solve them for the two unknown temperatures, surface and atmosphere.

- a. In terms of Earth’s radius, the constant σ , and the unknown temperature T_s of the surface, what is the power of the infrared radiation from the surface?

- b. What is the power of Earth's radiation absorbed by the atmosphere?
- c. In terms of the unknown temperature T_e of the atmosphere, what is the power radiated from the atmosphere?
- d. Write an equation that says the power of the radiation the atmosphere absorbs from Earth equals the power of the radiation it emits.
- e. Half of the power radiated by the atmosphere hits Earth. Write an equation that says that the power Earth absorbs from the atmosphere and the Sun equals the power that it emits.
- f. Solve your two equations for the unknown temperature of Earth.
For steps that make this model less crude, see for example the [lectures](#) by Paul O'Gorman.

Solution:

a. $4(\pi R^2)T_s^4$; b. $4e\sigma\pi R^2T_s^4$; c. $8e\sigma\pi R^2T_e^4$; d. $T_s^4 = 2T_e^4$; e. $e\sigma T_s^4 + \frac{1}{4}(1 - A)S = \sigma T_s^4$; f. $288K$

Glossary

conduction

heat transfer through stationary matter by physical contact

convection

heat transfer by the macroscopic movement of fluid

emissivity

measure of how well an object radiates

greenhouse effect

warming of the earth that is due to gases such as carbon dioxide and methane that absorb infrared radiation from Earth's surface and reradiate it in all directions, thus sending some of it back toward Earth

net rate of heat transfer by radiation

$$P_{\text{net}} = \sigma e A (T_2^4 - T_1^4)$$

radiation

energy transferred by electromagnetic waves directly as a result of a temperature difference

rate of conductive heat transfer

rate of heat transfer from one material to another

Stefan-Boltzmann law of radiation

$P = \sigma A e T^4$, where $\sigma = 5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant, A is the surface area of the object, T is the absolute temperature, and e is the emissivity

thermal conductivity

property of a material describing its ability to conduct heat

Introduction

class="introduction"

A volcanic eruption releases tons of gas and dust into the atmosphere. Most of the gas is water vapor, but several other gases are common, including greenhouse gases such as carbon dioxide and acidic pollutants such as sulfur dioxide. However, the emission of volcanic gas is not all bad: Many geologists believe that in the earliest stages of Earth's formation, volcanic emissions formed the early atmosphere. (credit: modification of work by "Boaworm"/Wikimedia Commons)



Gases are literally all around us—the air that we breathe is a mixture of gases. Other gases include those that make breads and cakes soft, those that make drinks fizzy, and those that burn to heat many homes. Engines and refrigerators depend on the behaviors of gases, as we will see in later chapters.

As we discussed in the preceding chapter, the study of heat and temperature is part of an area of physics known as thermodynamics, in which we require a system to be *macroscopic*, that is, to consist of a huge number (such as 10^{23}) of molecules. We begin by considering some macroscopic properties of gases: volume, pressure, and temperature. The simple model of a hypothetical “ideal gas” describes these properties of a gas very accurately under many conditions. We move from the ideal gas model to a more widely applicable approximation, called the Van der Waals model.

To understand gases even better, we must also look at them on the *microscopic* scale of molecules. In gases, the molecules interact weakly, so the microscopic behavior of gases is relatively simple, and they serve as a good introduction to systems of many molecules. The molecular model of gases is called the kinetic theory of gases and is one of the classic examples of a molecular model that explains everyday behavior.

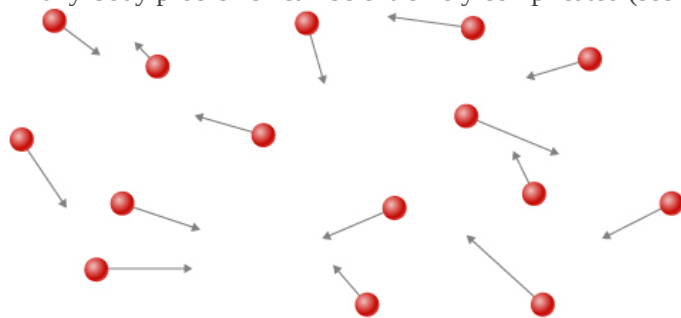
Molecular Model of an Ideal Gas

By the end of this section, you will be able to:

- Apply the ideal gas law to situations involving the pressure, volume, temperature, and the number of molecules of a gas
- Use the unit of moles in relation to numbers of molecules, and molecular and macroscopic masses
- Explain the ideal gas law in terms of moles rather than numbers of molecules
- Apply the van der Waals gas law to situations where the ideal gas law is inadequate

In this section, we explore the thermal behavior of gases. Our word “gas” comes from the Flemish word meaning “chaos,” first used for vapors by the seventeenth-century chemist J. B. van Helmont. The term was more appropriate than he knew, because gases consist of molecules moving and colliding with each other at random. This randomness makes the connection between the microscopic and macroscopic domains simpler for gases than for liquids or solids.

How do gases differ from solids and liquids? Under ordinary conditions, such as those of the air around us, the difference is that the molecules of gases are much farther apart than those of solids and liquids. Because the typical distances between molecules are large compared to the size of a molecule, as illustrated in [\[link\]](#), the forces between them are considered negligible, except when they come into contact with each other during collisions. Also, at temperatures well above the boiling temperature, the motion of molecules is fast, and the gases expand rapidly to occupy all of the accessible volume. In contrast, in liquids and solids, molecules are closer together, and the behavior of molecules in liquids and solids is highly constrained by the molecules’ interactions with one another. The macroscopic properties of such substances depend strongly on the forces between the molecules, and since many molecules are interacting, the resulting “many-body problems” can be extremely complicated (see [Condensed Matter Physics](#)).

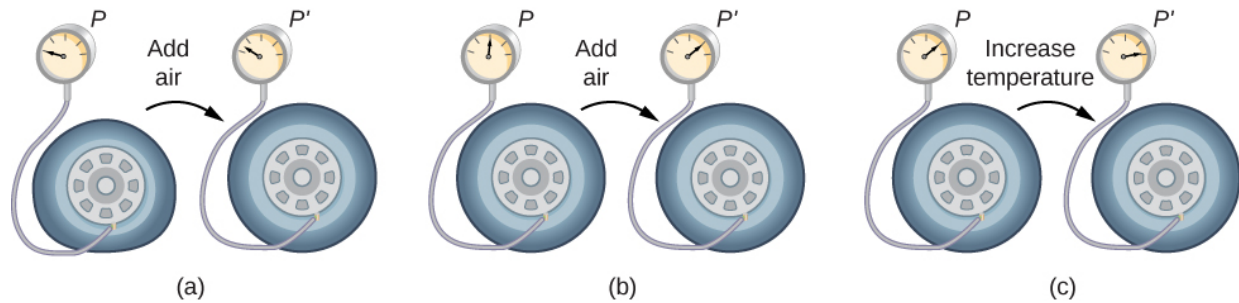


Atoms and molecules in a gas are typically widely separated. Because the forces between them are quite weak at these distances, the properties of a gas depend more on the number of atoms per unit volume and on temperature than on the type of atom.

The Gas Laws

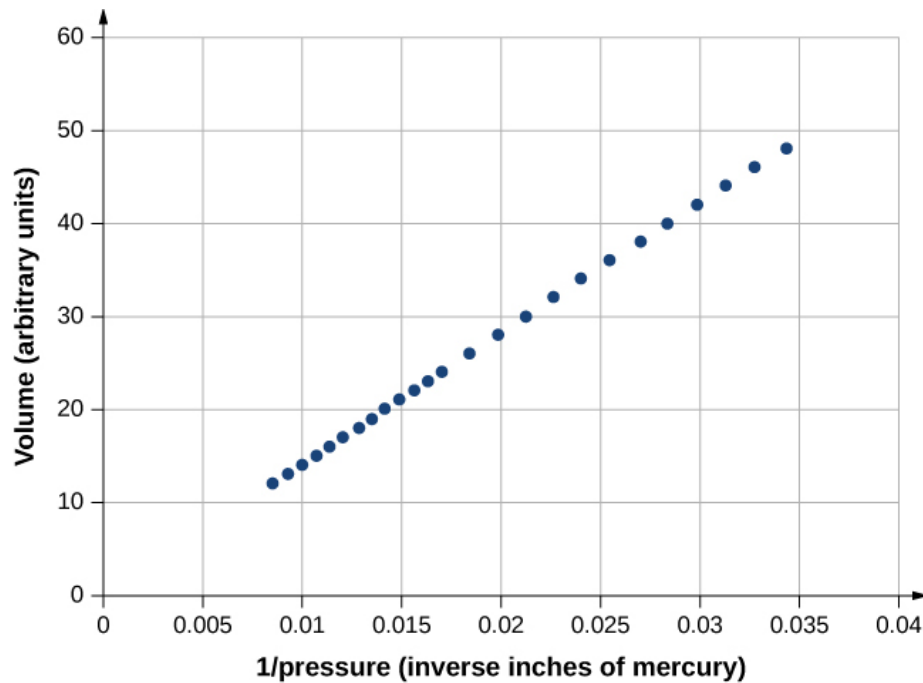
In the previous chapter, we saw one consequence of the large intermolecular spacing in gases: Gases are easily compressed. [\[link\]](#) shows that gases have larger coefficients of volume expansion than either solids or liquids. These large coefficients mean that gases expand and contract very rapidly with temperature changes. We also saw (in the section on thermal expansion) that most gases expand at the same rate or have the same coefficient of volume expansion, β . This raises a question: Why do all gases act in nearly the same way, when all the various liquids and solids have widely varying expansion rates?

To study how the pressure, temperature, and volume of a gas relate to one another, consider what happens when you pump air into a deflated car tire. The tire's volume first increases in direct proportion to the amount of air injected, without much increase in the tire pressure. Once the tire has expanded to nearly its full size, the tire's walls limit its volume expansion. If we continue to pump air into the tire, the pressure increases. When the car is driven and the tires flex, their temperature increases, and therefore the pressure increases even further ([link](#)).



(a) When air is pumped into a deflated tire, its volume first increases without much increase in pressure. (b) When the tire is filled to a certain point, the tire walls resist further expansion, and the pressure increases with more air. (c) Once the tire is inflated, its pressure increases with temperature.

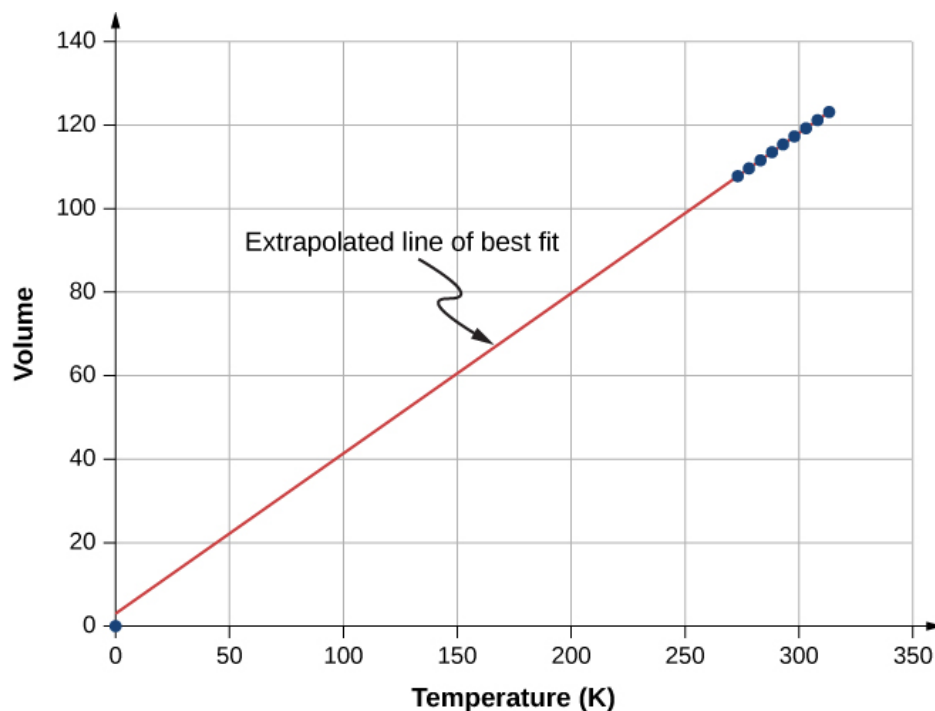
[link](#) shows data from the experiments of Robert Boyle (1627–1691), illustrating what is now called Boyle's law: At constant temperature and number of molecules, the absolute pressure of a gas and its volume are inversely proportional. (Recall from [Fluid Mechanics](#) that the absolute pressure is the true pressure and the gauge pressure is the absolute pressure minus the ambient pressure, typically atmospheric pressure.) The graph in [link](#) displays this relationship as an inverse proportionality of volume to pressure.



Robert Boyle and his assistant found that volume and pressure are inversely proportional. Here their data are plotted as V versus $1/p$; the linearity of the graph shows the inverse proportionality. The number shown as the volume is actually the height in inches of air in a cylindrical glass tube. The actual volume was that height multiplied by the cross-sectional area of the tube, which Boyle did not publish. The data are from Boyle's book *A Defence of the Doctrine Touching the Spring and Weight of the Air...*, p. 60.[\[footnote\]](#)

<http://bvpb.mcu.es/en/consulta/registro.cmd?id=406806>

[\[link\]](#) shows experimental data illustrating what is called Charles's law, after Jacques Charles (1746–1823). Charles's law states that at constant pressure and number of molecules, the volume of a gas is proportional to its absolute temperature.



Experimental data showing that at constant pressure, volume is approximately proportional to temperature. The best-fit line passes approximately through the origin. [\[footnote\]](http://chemed.chem.purdue.edu/genchem/history/charles.html)
<http://chemed.chem.purdue.edu/genchem/history/charles.html>

Similar is Amonton's or Gay-Lussac's law, which states that at constant volume and number of molecules, the pressure is proportional to the temperature. That law is the basis of the constant-volume gas thermometer, discussed in the previous chapter. (The histories of these laws and the appropriate credit for them are more complicated than can be discussed here.)

It is known experimentally that for gases at low density (such that their molecules occupy a negligible fraction of the total volume) and at temperatures well above the boiling point, these proportionalities hold to a good approximation. Not surprisingly, with the other quantities held constant, either pressure or volume is proportional to the number of molecules. More surprisingly, when the proportionalities are combined into a single equation, the constant of proportionality is independent of the composition of the gas. The resulting equation for all gases applies in the limit of low density and high temperature; it's the same for oxygen as for helium or uranium hexafluoride. A gas at that limit is called an **ideal gas**; it obeys the **ideal gas law**, which is also called the equation of state of an ideal gas.

Note:

Ideal Gas Law

The ideal gas law states that

Equation:

$$pV = Nk_{\text{B}}T,$$

where p is the absolute pressure of a gas, V is the volume it occupies, N is the number of molecules in the gas, and T is its absolute temperature.

The constant k_B is called the **Boltzmann constant** in honor of the Austrian physicist Ludwig Boltzmann (1844–1906) and has the value

Equation:

$$k_B = 1.38 \times 10^{-23} \text{ J/K}.$$

The ideal gas law describes the behavior of any real gas when its density is low enough or its temperature high enough that it is far from liquefaction. This encompasses many practical situations. In the next section, we'll see why it's independent of the type of gas.

In many situations, the ideal gas law is applied to a sample of gas with a constant number of molecules; for instance, the gas may be in a sealed container. If N is constant, then solving for N shows that pV/T is constant. We can write that fact in a convenient form:

Note:

Equation:

$$\frac{p_1 V_1}{T_1} = \frac{p_2 V_2}{T_2},$$

where the subscripts 1 and 2 refer to any two states of the gas at different times. Again, the temperature must be expressed in kelvin and the pressure must be absolute pressure, which is the sum of gauge pressure and atmospheric pressure.

Example:

Calculating Pressure Changes Due to Temperature Changes

Suppose your bicycle tire is fully inflated, with an absolute pressure of $7.00 \times 10^5 \text{ Pa}$ (a gauge pressure of just under 90.0 lb/in.^2) at a temperature of 18.0°C . What is the pressure after its temperature has risen to 35.0°C on a hot day? Assume there are no appreciable leaks or changes in volume.

Strategy

The pressure in the tire is changing only because of changes in temperature. We know the initial pressure $p_0 = 7.00 \times 10^5 \text{ Pa}$, the initial temperature $T_0 = 18.0^\circ \text{C}$, and the final temperature $T_f = 35.0^\circ \text{C}$. We must find the final pressure p_f . Since the number of molecules is constant, we can use the equation

Equation:

$$\frac{p_f V_f}{T_f} = \frac{p_0 V_0}{T_0}.$$

Since the volume is constant, V_f and V_0 are the same and they divide out. Therefore,

Equation:

$$\frac{p_f}{T_f} = \frac{p_0}{T_0}.$$

We can then rearrange this to solve for p_f :

Equation:

$$p_f = p_0 \frac{T_f}{T_0},$$

where the temperature must be in kelvin.

Solution

1. Convert temperatures from degrees Celsius to kelvin

Equation:

$$T_0 = (18.0 + 273)\text{K} = 291\text{ K},$$

Equation:

$$T_f = (35.0 + 273)\text{K} = 308\text{ K}.$$

2. Substitute the known values into the equation,

Equation:

$$p_f = p_0 \frac{T_f}{T_0} = 7.00 \times 10^5 \text{ Pa} \left(\frac{308 \text{ K}}{291 \text{ K}} \right) = 7.41 \times 10^5 \text{ Pa}.$$

Significance

The final temperature is about 6 % greater than the original temperature, so the final pressure is about 6 % greater as well. Note that *absolute pressure* (see [Fluid Mechanics](#)) and *absolute temperature* (see [Temperature and Heat](#)) must be used in the ideal gas law.

Example:

Calculating the Number of Molecules in a Cubic Meter of Gas

How many molecules are in a typical object, such as gas in a tire or water in a glass? This calculation can give us an idea of how large N typically is. Let's calculate the number of molecules in the air that a typical healthy young adult inhales in one breath, with a volume of 500 mL, at *standard temperature and pressure* (STP), which is defined as 0 °C and atmospheric pressure. (Our young adult is apparently outside in winter.)

Strategy

Because pressure, volume, and temperature are all specified, we can use the ideal gas law, $pV = Nk_B T$, to find N .

Solution

1. Identify the knowns.

Equation:

$$T = 0^\circ \text{C} = 273 \text{ K}, p = 1.01 \times 10^5 \text{ Pa}, V = 500 \text{ mL} = 5 \times 10^{-4} \text{ m}^3, k_B = 1.38 \times 10^{-23} \text{ J/K}$$

2. Substitute the known values into the equation and solve for N .

Equation:

$$N = \frac{pV}{k_B T} = \frac{(1.01 \times 10^5 \text{ Pa}) (5 \times 10^{-4} \text{ m}^3)}{(1.38 \times 10^{-23} \text{ J/K}) (273 \text{ K})} = 1.34 \times 10^{22} \text{ molecules}$$

Significance

N is huge, even in small volumes. For example, 1 cm^3 of a gas at STP contains 2.68×10^{19} molecules. Once again, note that our result for N is the same for all types of gases, including mixtures.

As we observed in the chapter on fluid mechanics, pascals are N/m^2 , so $\text{Pa} \cdot \text{m}^3 = \text{N} \cdot \text{m} = \text{J}$. Thus, our result for N is dimensionless, a pure number that could be obtained by counting (in principle) rather than measuring. As it is the number of molecules, we put “molecules” after the number, keeping in mind that it is an aid to communication rather than a unit.

Moles and Avogadro’s Number

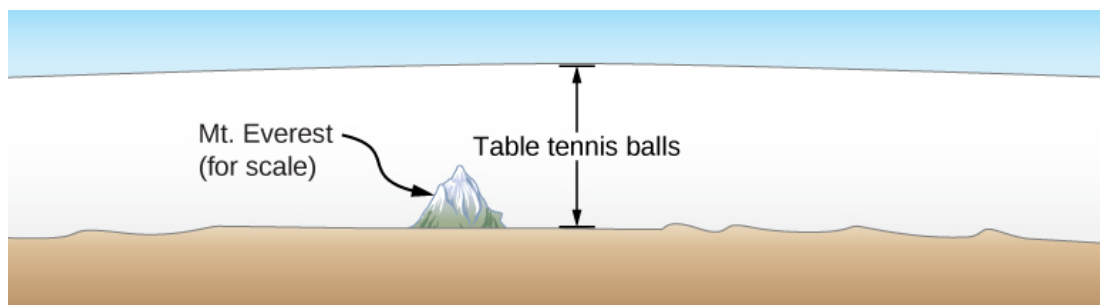
It is often convenient to measure the amount of substance with a unit on a more human scale than molecules. The SI unit for this purpose was developed by the Italian scientist Amedeo Avogadro (1776–1856). (He worked from the hypothesis that equal volumes of gas at equal pressure and temperature contain equal numbers of molecules, independent of the type of gas. As mentioned above, this hypothesis has been confirmed when the ideal gas approximation applies.) A **mole** (abbreviated mol) is defined as the amount of any substance that contains as many molecules as there are atoms in exactly 12 grams (0.012 kg) of carbon-12. (Technically, we should say “formula units,” not “molecules,” but this distinction is irrelevant for our purposes.) The number of molecules in one mole is called **Avogadro’s number** (N_A), and the value of Avogadro’s number is now known to be

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}.$$

We can now write $N = N_A n$, where n represents the number of moles of a substance.

Avogadro’s number relates the mass of an amount of substance in grams to the number of protons and neutrons in an atom or molecule (12 for a carbon-12 atom), which roughly determine its mass. It’s natural to define a unit of mass such that the mass of an atom is approximately equal to its number of neutrons and protons. The unit of that kind accepted for use with the SI is the *unified atomic mass unit* (u), also called the *dalton*. Specifically, a carbon-12 atom has a mass of exactly 12 u, so that its molar mass M in grams per mole is numerically equal to the mass of one carbon-12 atom in u. That equality holds for any substance. In other words, N_A is not only the conversion from numbers of molecules to moles, but it is also the conversion from u to grams: $6.02 \times 10^{23} \text{ u} = 1 \text{ g}$. See [\[link\]](#).



How big is a mole? On a macroscopic level, Avogadro's number of table tennis balls would cover Earth to a depth of about 40 km.

Now letting m_s stand for the mass of a sample of a substance, we have $m_s = nM$. Letting m stand for the mass of a molecule, we have $M = N_A m$.

Note:

Exercise:

Problem:

Check Your Understanding The recommended daily amount of vitamin B₃ or niacin, C₆NH₅O₂, for women who are not pregnant or nursing, is 14 mg. Find the number of molecules of niacin in that amount.

Solution:

We first need to calculate the molar mass (the mass of one mole) of niacin. To do this, we must multiply the number of atoms of each element in the molecule by the element's molar mass.

$$(6 \text{ mol of carbon}) (12.0 \text{ g/mol}) + (5 \text{ mol hydrogen}) (1.0 \text{ g/mol})$$

$$+ (1 \text{ mol nitrogen}) (14 \text{ g/mol}) + (2 \text{ mol oxygen}) (16.0 \text{ g/mol}) = 123 \text{ g/mol}$$

Then we need to calculate the number of moles in 14 mg.

$$\left(\frac{14 \text{ mg}}{123 \text{ g/mol}} \right) \left(\frac{1 \text{ g}}{1000 \text{ mg}} \right) = 1.14 \times 10^{-4} \text{ mol.}$$

Then, we use Avogadro's number to calculate the number of molecules:

$$N = nN_A = (1.14 \times 10^{-4} \text{ mol}) (6.02 \times 10^{23} \text{ molecules/mol}) = 6.85 \times 10^{19} \text{ molecules.}$$

Note:

Exercise:

Problem:

Check Your Understanding The density of air in a classroom ($p = 1.00 \text{ atm}$ and $T = 20^\circ \text{C}$) is 1.28 kg/m^3 . At what pressure is the density 0.600 kg/m^3 if the temperature is kept constant?

Solution:

The density of a gas is equal to a constant, the average molecular mass, times the number density N/V . From the ideal gas law, $pV = Nk_B T$, we see that $N/V = p/k_B T$. Therefore, at constant temperature, if the density and, consequently, the number density are reduced by half, the pressure must also be reduced by half, and $p_f = 0.500 \text{ atm}$.

The Ideal Gas Law Restated using Moles

A very common expression of the ideal gas law uses the number of moles in a sample, n , rather than the number of molecules, N . We start from the ideal gas law,

Equation:

$$pV = Nk_{\text{B}}T,$$

and multiply and divide the right-hand side of the equation by Avogadro's number N_{A} . This gives us

Equation:

$$pV = \frac{N}{N_{\text{A}}} N_{\text{A}} k_{\text{B}} T.$$

Note that $n = N/N_{\text{A}}$ is the number of moles. We define the **universal gas constant** as $R = N_{\text{A}}k_{\text{B}}$, and obtain the ideal gas law in terms of moles.

Note:

Ideal Gas Law (in terms of moles)

In terms of number of moles n , the ideal gas law is written as

Equation:

$$pV = nRT.$$

In SI units,

Equation:

$$R = N_{\text{A}}k_{\text{B}} = (6.02 \times 10^{23} \text{ mol}^{-1}) \left(1.38 \times 10^{-23} \frac{\text{J}}{\text{K}} \right) = 8.31 \frac{\text{J}}{\text{mol} \cdot \text{K}}.$$

In other units,

Equation:

$$R = 1.99 \frac{\text{cal}}{\text{mol} \cdot \text{K}} = 0.0821 \frac{\text{L} \cdot \text{atm}}{\text{mol} \cdot \text{K}}.$$

You can use whichever value of R is most convenient for a particular problem.

Example:

Density of Air at STP and in a Hot Air Balloon

Calculate the density of dry air (a) under standard conditions and (b) in a hot air balloon at a temperature of 120 °C. Dry air is approximately 78 % N₂, 21 % O₂, and 1 % Ar.

Strategy and Solution

- a. We are asked to find the density, or mass per cubic meter. We can begin by finding the molar mass. If we have a hundred molecules, of which 78 are nitrogen, 21 are oxygen, and 1 is argon, the average molecular mass is $\frac{78 m_{\text{N}_2} + 21 m_{\text{O}_2} + m_{\text{Ar}}}{100}$, or the mass of each constituent multiplied by its percentage. The same applies to the molar mass, which therefore is

Equation:

$$M = 0.78 M_{\text{N}_2} + 0.21 M_{\text{O}_2} + 0.01 M_{\text{Ar}} = 29.0 \text{ g/mol}.$$

Now we can find the number of moles per cubic meter. We use the ideal gas law in terms of moles, $pV = nRT$, with $p = 1.00 \text{ atm}$, $T = 273 \text{ K}$, $V = 1 \text{ m}^3$, and $R = 8.31 \text{ J/mol} \cdot \text{K}$. The most convenient choice for R in this case is $R = 8.31 \text{ J/mol} \cdot \text{K}$ because the known quantities are in SI units:

Equation:

$$n = \frac{pV}{RT} = \frac{(1.00 \times 10^5 \text{ Pa})(1 \text{ m}^3)}{(8.31 \text{ J/mol} \cdot \text{K})(273 \text{ K})} = 44.1 \text{ mol}.$$

Then, the mass m_s of that air is

Equation:

$$m_s = nM = (44.1 \text{ mol})(29.0 \text{ g/mol}) = 1290 \text{ g} = 1.28 \text{ kg}.$$

Finally the density of air at STP is

Equation:

$$\rho = \frac{m_s}{V} = \frac{1.28 \text{ kg}}{1 \text{ m}^3} = 1.28 \text{ kg/m}^3.$$

- b. The air pressure inside the balloon is still 1 atm because the bottom of the balloon is open to the atmosphere. The calculation is the same except that we use a temperature of 120°C , which is 393 K . We can repeat the calculation in (a), or simply observe that the density is proportional to the number of moles, which is inversely proportional to the temperature. Then using the subscripts 1 for air at STP and 2 for the hot air, we have

Equation:

$$\rho_2 = \frac{T_1}{T_2} \rho_1 = \frac{273 \text{ K}}{393 \text{ K}} (1.28 \text{ kg/m}^3) = 0.889 \text{ kg/m}^3.$$

Significance

Using the methods of [Archimedes' Principle and Buoyancy](#), we can find that the net force on 2200 m^3 of air at 120°C is $F_b - F_g = \rho_{\text{atmosphere}} Vg - \rho_{\text{hot air}} Vg = 8.49 \times 10^3 \text{ N}$, or enough to lift about 867 kg. The mass density and molar density of air at STP, found above, are often useful numbers. From the molar density, we can easily determine another useful number, the volume of a mole of any ideal gas at STP, which is 22.4 L .

Note:

Exercise:

Problem:

Check Your Understanding Liquids and solids have densities on the order of 1000 times greater than gases. Explain how this implies that the distances between molecules in gases are on the order of 10 times greater than the size of their molecules.

Solution:

Density is mass per unit volume, and volume is proportional to the size of a body (such as the radius of a sphere) cubed. So if the distance between molecules increases by a factor of 10, then the volume occupied increases by a factor of 1000, and the density decreases by a factor of 1000. Since we assume molecules are in contact in liquids and solids, the distance between their centers is on the order of their typical size, so the distance in gases is on the order of 10 times as great.

The ideal gas law is closely related to energy: The units on both sides of the equation are joules. The right-hand side of the ideal gas law equation is $Nk_{\text{B}}T$. This term is roughly the total translational kinetic energy (which, when discussing gases, refers to the energy of translation of a molecule, not that of vibration of its atoms or rotation) of N molecules at an absolute temperature T , as we will see formally in the next section. The left-hand side of the ideal gas law equation is pV . As mentioned in the example on the number of molecules in an ideal gas, pressure multiplied by volume has units of energy. The energy of a gas can be changed when the gas does work as it increases in volume, something we explored in the preceding chapter, and the amount of work is related to the pressure. This is the process that occurs in gasoline or steam engines and turbines, as we'll see in the next chapter.

Note:

The Ideal Gas Law

Step 1. Examine the situation to determine that an ideal gas is involved. Most gases are nearly ideal unless they are close to the boiling point or at pressures far above atmospheric pressure.

Step 2. Make a list of what quantities are given or can be inferred from the problem as stated (identify the known quantities).

Step 3. Identify exactly what needs to be determined in the problem (identify the unknown quantities). A written list is useful.

Step 4. Determine whether the number of molecules or the number of moles is known or asked for to decide whether to use the ideal gas law as $pV = Nk_{\text{B}}T$, where N is the number of molecules, or $pV = nRT$, where n is the number of moles.

Step 5. Convert known values into proper SI units (K for temperature, Pa for pressure, m^3 for volume, molecules for N , and moles for n). If the units of the knowns are consistent with one of the non-SI values of R , you can leave them in those units. Be sure to use absolute temperature and absolute pressure.

Step 6. Solve the ideal gas law for the quantity to be determined (the unknown quantity). You may need to take a ratio of final states to initial states to eliminate the unknown quantities that are kept fixed.

Step 7. Substitute the known quantities, along with their units, into the appropriate equation and obtain numerical solutions complete with units.

Step 8. Check the answer to see if it is reasonable: Does it make sense?

The Van der Waals Equation of State

We have repeatedly noted that the ideal gas law is an approximation. How can it be improved upon? The **van der Waals equation of state** (named after the Dutch physicist Johannes van der Waals, 1837–1923) improves it by taking into account two factors. First, the attractive forces between molecules, which are stronger at higher density and reduce the pressure, are taken into account by adding to the pressure a term equal to the square of the molar density multiplied by a positive coefficient a . Second, the volume of the molecules is represented by a positive constant b , which can be thought of as the volume of a mole of molecules. This is subtracted from the total volume to give the remaining volume that the molecules can move in. The constants a and b are determined experimentally for each gas. The resulting equation is

Note:
Equation:

$$\left[p + a \left(\frac{n}{V} \right)^2 \right] (V - nb) = nRT.$$

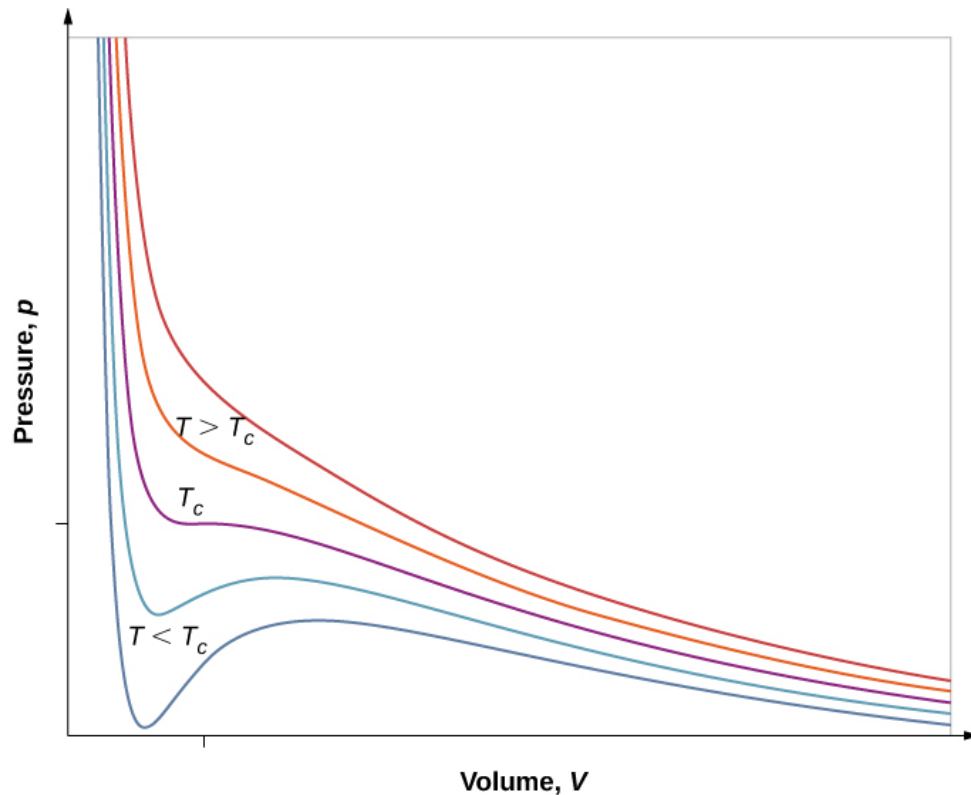
In the limit of low density (small n), the a and b terms are negligible, and we have the ideal gas law, as we should for low density. On the other hand, if $V - nb$ is small, meaning that the molecules are very close together, the pressure must be higher to give the same nRT , as we would expect in the situation of a highly compressed gas. However, the increase in pressure is less than that argument would suggest, because at high density the $(n/V)^2$ term is significant. Since it's positive, it causes a lower pressure to give the same nRT .

The van der Waals equation of state works well for most gases under a wide variety of conditions. As we'll see in the next module, it even predicts the gas-liquid transition.

***pV* Diagrams**

We can examine aspects of the behavior of a substance by plotting a ***pV* diagram**, which is a graph of pressure versus volume. When the substance behaves like an ideal gas, the ideal gas law $pV = nRT$ describes the relationship between its pressure and volume. On a *pV* diagram, it's common to plot an *isotherm*, which is a curve showing p as a function of V with the number of molecules and the temperature fixed. Then, for an ideal gas, $pV = \text{constant}$. For example, the volume of the gas decreases as the pressure increases. The resulting graph is a hyperbola.

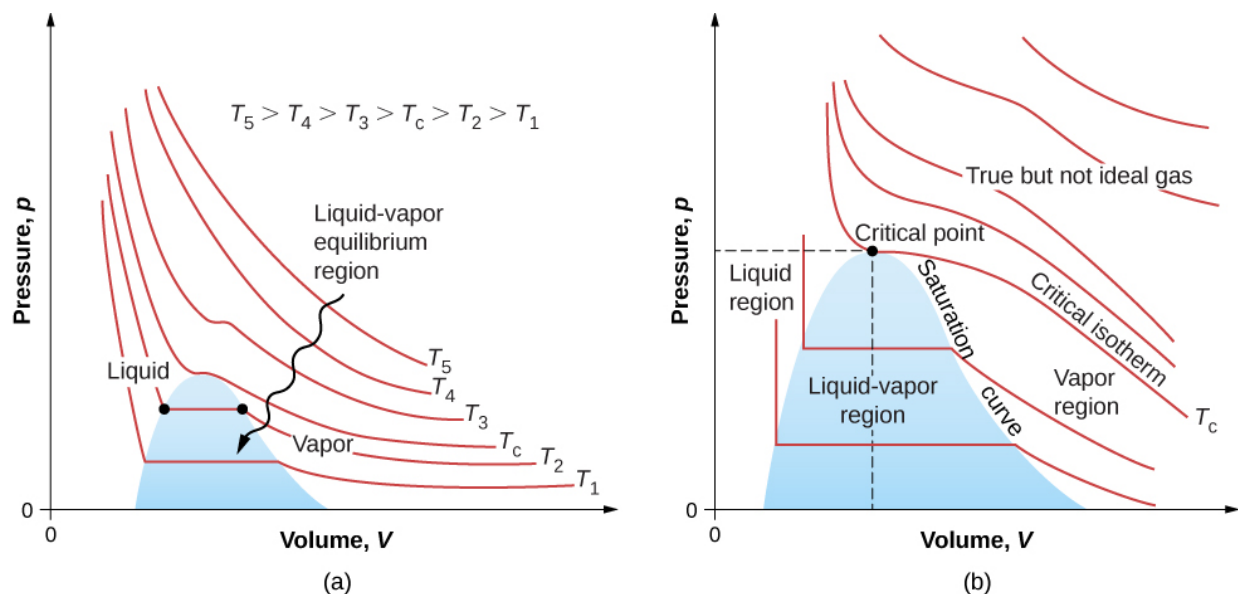
However, if we assume the van der Waals equation of state, the isotherms become more interesting, as shown in [\[link\]](#). At high temperatures, the curves are approximately hyperbolas, representing approximately ideal behavior at various fixed temperatures. At lower temperatures, the curves look less and less like hyperbolas—that is, the gas is not behaving ideally. There is a **critical temperature** T_c at which the curve has a point with zero slope. Below that temperature, the curves do not decrease monotonically; instead, they each have a “hump,” meaning that for a certain range of volume, increasing the volume increases the pressure.



pV diagram for a Van der Waals gas at various temperatures. The red curves are calculated at temperatures above the critical temperature and the blue curves at temperatures below it. The blue curves have an oscillation in which volume (V) increases with increasing pressure (P), an impossible situation, so they must be corrected as in [\[link\]](#). (credit: “Eman”/Wikimedia Commons)

Such behavior would be completely unphysical. Instead, the curves are understood as describing a liquid-gas phase transition. The oscillating part of the curve is replaced by a horizontal line, showing that as the volume increases at constant temperature, the pressure stays constant. That behavior corresponds to boiling and condensation; when a substance is at its boiling temperature for a particular pressure, it can increase in volume as some of the liquid turns to gas, or decrease as some of the gas turns to liquid, without any change in temperature or pressure.

[\[link\]](#) shows similar isotherms that are more realistic than those based on the van der Waals equation. The steep parts of the curves to the left of the transition region show the liquid phase, which is almost incompressible—a slight decrease in volume requires a large increase in pressure. The flat parts show the liquid-gas transition; the blue regions that they define represent combinations of pressure and volume where liquid and gas can coexist.



pV diagrams. (a) Each curve (isotherm) represents the relationship between p and V at a fixed temperature; the upper curves are at higher temperatures. The lower curves are not hyperbolas because the gas is no longer an ideal gas. (b) An expanded portion of the pV diagram for low temperatures, where the phase can change from a gas to a liquid. The term “vapor” refers to the gas phase when it exists at a temperature below the boiling temperature.

The isotherms above T_c do not go through the liquid-gas transition. Therefore, liquid cannot exist above that temperature, which is the critical temperature (described in the chapter on temperature and heat). At sufficiently low pressure above that temperature, the gas has the density of a liquid but will not condense; the gas is said to be **supercritical**. At higher pressure, it is solid. Carbon dioxide, for example, has no liquid phase at a temperature above 31.0 °C. The critical pressure is the maximum pressure at which the liquid can exist. The point on the pV diagram at the critical pressure and temperature is the critical point (which you learned about in the chapter on temperature and heat). [\[link\]](#) lists representative critical temperatures and pressures.

Substance	Critical temperature		Critical pressure	
	K	°C	Pa	atm
Water	647.4	374.3	22.12×10^6	219.0
Sulfur dioxide	430.7	157.6	7.88×10^6	78.0
Ammonia	405.5	132.4	11.28×10^6	111.7
Carbon dioxide	304.2	31.1	7.39×10^6	73.2

Substance	Critical temperature		Critical pressure	
Oxygen	154.8	−118.4	5.08×10^6	50.3
Nitrogen	126.2	−146.9	3.39×10^6	33.6
Hydrogen	33.3	−239.9	1.30×10^6	12.9
Helium	5.3	−267.9	0.229×10^6	2.27

Critical Temperatures and Pressures for Various Substances

Summary

- The ideal gas law relates the pressure and volume of a gas to the number of gas molecules and the temperature of the gas.
- A mole of any substance has a number of molecules equal to the number of atoms in a 12-g sample of carbon-12. The number of molecules in a mole is called Avogadro's number N_A ,

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}.$$

- A mole of any substance has a mass in grams numerically equal to its molecular mass in unified mass units, which can be determined from the periodic table of elements. The ideal gas law can also be written and solved in terms of the number of moles of gas:

Equation:

$$pV = nRT,$$

where n is the number of moles and R is the universal gas constant,

Equation:

$$R = 8.31 \text{ J/mol} \cdot \text{K}.$$

- The ideal gas law is generally valid at temperatures well above the boiling temperature.
- The van der Waals equation of state for gases is valid closer to the boiling point than the ideal gas law.
- Above the critical temperature and pressure for a given substance, the liquid phase does not exist, and the sample is “supercritical.”

Conceptual Questions

Exercise:

Problem:

Two H_2 molecules can react with one O_2 molecule to produce two H_2O molecules. How many moles of hydrogen molecules are needed to react with one mole of oxygen molecules?

Solution:

2 moles, as that will contain twice as many molecules as the 1 mole of oxygen

Exercise:

Problem:

Under what circumstances would you expect a gas to behave significantly differently than predicted by the ideal gas law?

Exercise:**Problem:**

A constant-volume gas thermometer contains a fixed amount of gas. What property of the gas is measured to indicate its temperature?

Solution:

pressure

Exercise:**Problem:**

Inflate a balloon at room temperature. Leave the inflated balloon in the refrigerator overnight. What happens to the balloon, and why?

Exercise:**Problem:**

In the last chapter, free convection was explained as the result of buoyant forces on hot fluids. Explain the upward motion of air in flames based on the ideal gas law.

Solution:

The flame contains hot gas (heated by combustion). The pressure is still atmospheric pressure, in mechanical equilibrium with the air around it (or roughly so). The density of the hot gas is proportional to its number density N/V (neglecting the difference in composition between the gas in the flame and the surrounding air). At higher temperature than the surrounding air, the ideal gas law says that $N/V = p/k_B T$ is less than that of the surrounding air. Therefore the hot air has lower density than the surrounding air and is lifted by the buoyant force.

Problems**Exercise:****Problem:**

The gauge pressure in your car tires is $2.50 \times 10^5 \text{ N/m}^2$ at a temperature of 35.0°C when you drive it onto a ship in Los Angeles to be sent to Alaska. What is their gauge pressure on a night in Alaska when their temperature has dropped to -40.0°C ? Assume the tires have not gained or lost any air.

Exercise:

Problem:

Suppose a gas-filled incandescent light bulb is manufactured so that the gas inside the bulb is at atmospheric pressure when the bulb has a temperature of $20.0\text{ }^{\circ}\text{C}$. (a) Find the gauge pressure inside such a bulb when it is hot, assuming its average temperature is $60.0\text{ }^{\circ}\text{C}$ (an approximation) and neglecting any change in volume due to thermal expansion or gas leaks. (b) The actual final pressure for the light bulb will be less than calculated in part (a) because the glass bulb will expand. Is this effect significant?

Solution:

a. 0.137 atm ; b. $p_g = (1\text{ atm})\frac{T_2V_1}{T_1V_2} - 1\text{ atm}$. Because of the expansion of the glass, $V_2 = 0.99973$. Multiplying by that factor does not make any significant difference.

Exercise:**Problem:**

People buying food in sealed bags at high elevations often notice that the bags are puffed up because the air inside has expanded. A bag of pretzels was packed at a pressure of 1.00 atm and a temperature of $22.0\text{ }^{\circ}\text{C}$. When opened at a summer picnic in Santa Fe, New Mexico, at a temperature of $32.0\text{ }^{\circ}\text{C}$, the volume of the air in the bag is 1.38 times its original volume. What is the pressure of the air?

Exercise:**Problem:**

How many moles are there in (a) 0.0500 g of N_2 gas ($M = 28.0\text{ g/mol}$)? (b) 10.0 g of CO_2 gas ($M = 44.0\text{ g/mol}$)? (c) How many molecules are present in each case?

Solution:

a. $1.79 \times 10^{-3}\text{ mol}$; b. 0.227 mol ; c. 1.08×10^{21} molecules for the nitrogen, 1.37×10^{23} molecules for the carbon dioxide

Exercise:**Problem:**

A cubic container of volume 2.00 L holds 0.500 mol of nitrogen gas at a temperature of $25.0\text{ }^{\circ}\text{C}$. What is the net force due to the nitrogen on one wall of the container? Compare that force to the sample's weight.

Exercise:**Problem:**

Calculate the number of moles in the 2.00-L volume of air in the lungs of the average person. Note that the air is at $37.0\text{ }^{\circ}\text{C}$ (body temperature) and that the total volume in the lungs is several times the amount inhaled in a typical breath as given in [\[link\]](#).

Solution:

$7.84 \times 10^{-2}\text{ mol}$

Exercise:

Problem:

An airplane passenger has 100 cm^3 of air in his stomach just before the plane takes off from a sea-level airport. What volume will the air have at cruising altitude if cabin pressure drops to $7.50 \times 10^4 \text{ N/m}^2$?

Exercise:**Problem:**

A company advertises that it delivers helium at a gauge pressure of $1.72 \times 10^7 \text{ Pa}$ in a cylinder of volume 43.8 L. How many balloons can be inflated to a volume of 4.00 L with that amount of helium? Assume the pressure inside the balloons is $1.01 \times 10^5 \text{ Pa}$ and the temperature in the cylinder and the balloons is 25.0°C .

Solution:

$$1.87 \times 10^3$$

Exercise:**Problem:**

According to <http://hyperphysics.phy-astr.gsu.edu/hbase/solar/venusenv.html>, the atmosphere of Venus is approximately 96.5% CO_2 and 3.5% N_2 by volume. On the surface, where the temperature is about 750 K and the pressure is about 90 atm, what is the density of the atmosphere?

Exercise:**Problem:**

An expensive vacuum system can achieve a pressure as low as $1.00 \times 10^{-7} \text{ N/m}^2$ at 20.0°C . How many molecules are there in a cubic centimeter at this pressure and temperature?

Solution:

$$2.47 \times 10^7 \text{ molecules}$$

Exercise:**Problem:**

The number density N/V of gas molecules at a certain location in the space above our planet is about $1.00 \times 10^{11} \text{ m}^{-3}$, and the pressure is $2.75 \times 10^{-10} \text{ N/m}^2$ in this space. What is the temperature there?

Exercise:**Problem:**

A bicycle tire contains 2.00 L of gas at an absolute pressure of $7.00 \times 10^5 \text{ N/m}^2$ and a temperature of 18.0°C . What will its pressure be if you let out an amount of air that has a volume of 100 cm^3 at atmospheric pressure? Assume tire temperature and volume remain constant.

Solution:

$$6.95 \times 10^5 \text{ Pa}; 6.86 \text{ atm}$$

Exercise:**Problem:**

In a common demonstration, a bottle is heated and stoppered with a hard-boiled egg that's a little bigger than the bottle's neck. When the bottle is cooled, the pressure difference between inside and outside forces the egg into the bottle. Suppose the bottle has a volume of 0.500 L and the temperature inside it is raised to 80.0 °C while the pressure remains constant at 1.00 atm because the bottle is open. (a) How many moles of air are inside? (b) Now the egg is put in place, sealing the bottle. What is the gauge pressure inside after the air cools back to the ambient temperature of 25 °C but before the egg is forced into the bottle?

Exercise:**Problem:**

A high-pressure gas cylinder contains 50.0 L of toxic gas at a pressure of $1.40 \times 10^7 \text{ N/m}^2$ and a temperature of 25.0 °C. The cylinder is cooled to dry ice temperature (−78.5 °C) to reduce the leak rate and pressure so that it can be safely repaired. (a) What is the final pressure in the tank, assuming a negligible amount of gas leaks while being cooled and that there is no phase change? (b) What is the final pressure if one-tenth of the gas escapes? (c) To what temperature must the tank be cooled to reduce the pressure to 1.00 atm (assuming the gas does not change phase and that there is no leakage during cooling)? (d) Does cooling the tank as in part (c) appear to be a practical solution?

Solution:

a. $9.14 \times 10^6 \text{ Pa}$; b. $8.22 \times 10^6 \text{ Pa}$; c. 2.15 K; d. no

Exercise:**Problem:**

Find the number of moles in 2.00 L of gas at 35.0 °C and under $7.41 \times 10^7 \text{ N/m}^2$ of pressure.

Exercise:**Problem:**

Calculate the depth to which Avogadro's number of table tennis balls would cover Earth. Each ball has a diameter of 3.75 cm. Assume the space between balls adds an extra 25.0% to their volume and assume they are not crushed by their own weight.

Solution:

40.7 km

Exercise:**Problem:**

(a) What is the gauge pressure in a 25.0 °C car tire containing 3.60 mol of gas in a 30.0-L volume? (b) What will its gauge pressure be if you add 1.00 L of gas originally at atmospheric pressure and 25.0 °C? Assume the temperature remains at 25.0 °C and the volume remains constant.

Glossary

Avogadro's number

N_A , the number of molecules in one mole of a substance; $N_A = 6.02 \times 10^{23}$ particles/mole

Boltzmann constant

k_B , a physical constant that relates energy to temperature and appears in the ideal gas law;

$$k_B = 1.38 \times 10^{-23} \text{ J/K}$$

critical temperature

T_c at which the isotherm has a point with zero slope

ideal gas

gas at the limit of low density and high temperature

ideal gas law

physical law that relates the pressure and volume of a gas, far from liquefaction, to the number of gas molecules or number of moles of gas and the temperature of the gas

mole

quantity of a substance whose mass (in grams) is equal to its molecular mass

pV diagram

graph of pressure vs. volume

supercritical

condition of a fluid being at such a high temperature and pressure that the liquid phase cannot exist

universal gas constant

R , the constant that appears in the ideal gas law expressed in terms of moles, given by $R = N_A k_B$

van der Waals equation of state

equation, typically approximate, which relates the pressure and volume of a gas to the number of gas molecules or number of moles of gas and the temperature of the gas

Pressure, Temperature, and RMS Speed

By the end of this section, you will be able to:

- Explain the relations between microscopic and macroscopic quantities in a gas
- Solve problems involving mixtures of gases
- Solve problems involving the distance and time between a gas molecule's collisions

We have examined pressure and temperature based on their macroscopic definitions. Pressure is the force divided by the area on which the force is exerted, and temperature is measured with a thermometer. We can gain a better understanding of pressure and temperature from the **kinetic theory of gases**, the theory that relates the macroscopic properties of gases to the motion of the molecules they consist of. First, we make two assumptions about molecules in an ideal gas.

1. There is a very large number N of molecules, all identical and each having mass m .
2. The molecules obey Newton's laws and are in continuous motion, which is random and isotropic, that is, the same in all directions.

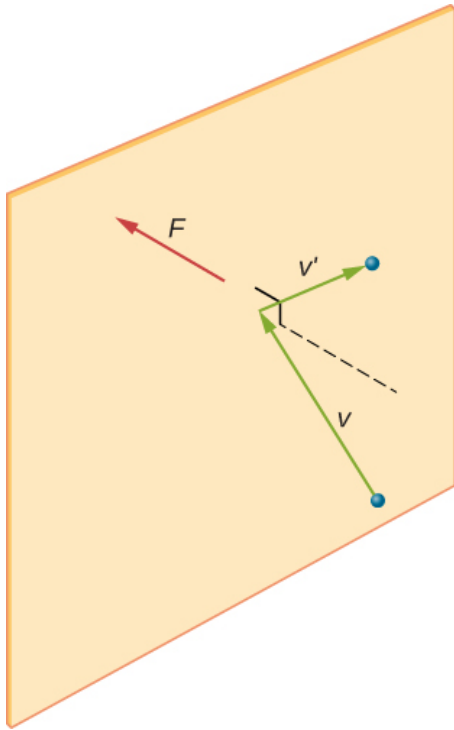
To derive the ideal gas law and the connection between microscopic quantities such as the energy of a typical molecule and macroscopic quantities such as temperature, we analyze a sample of an ideal gas in a rigid container, about which we make two further assumptions:

3. The molecules are much smaller than the average distance between them, so their total volume is much less than that of their container (which has volume V). In other words, we take the Van der Waals constant b , the volume of a mole of gas molecules, to be negligible compared to the volume of a mole of gas in the container.
4. The molecules make perfectly elastic collisions with the walls of the container and with each other. Other forces on them, including gravity and the attractions represented by the Van der Waals constant a , are negligible (as is necessary for the assumption of isotropy).

The collisions between molecules do not appear in the derivation of the ideal gas law. They do not disturb the derivation either, since collisions between molecules moving with random velocities give new random velocities. Furthermore, if the velocities of gas molecules in a container are initially not random and isotropic, molecular collisions are what make them random and isotropic.

We make still further assumptions that simplify the calculations but do not affect the result. First, we let the container be a rectangular box. Second, we begin by considering *monatomic* gases, those whose molecules consist of single atoms, such as helium. Then, we can assume that the atoms have no energy except their translational kinetic energy; for instance, they have neither rotational nor vibrational energy. (Later, we discuss the validity of this assumption for real monatomic gases and dispense with it to consider diatomic and polyatomic gases.)

[\[link\]](#) shows a collision of a gas molecule with the wall of a container, so that it exerts a force on the wall (by Newton's third law). These collisions are the source of pressure in a gas. As the number of molecules increases, the number of collisions, and thus the pressure, increases. Similarly, if the average velocity of the molecules is higher, the gas pressure is higher.



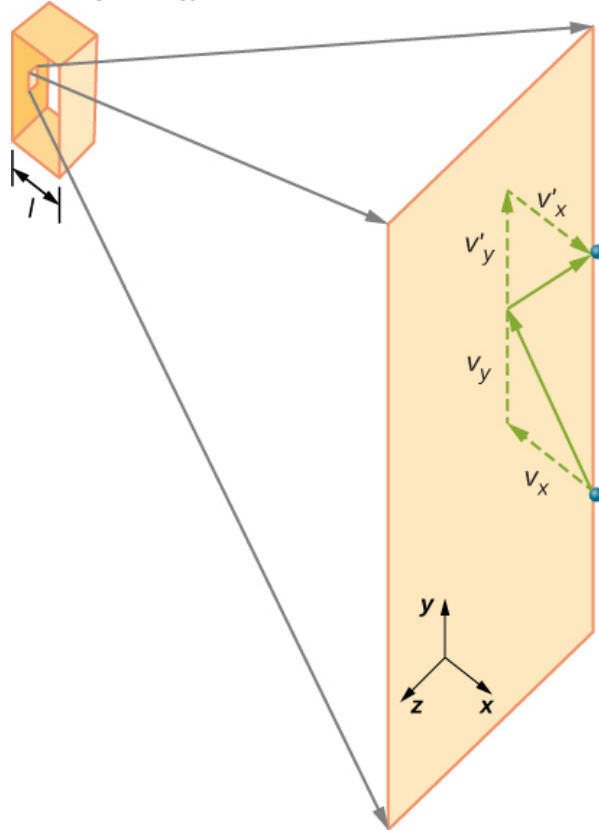
When a molecule collides with a rigid wall, the component of its momentum perpendicular to the wall is reversed. A force is thus exerted on the wall, creating pressure.

In a sample of gas in a container, the randomness of the molecular motion causes the number of collisions of molecules with any part of the wall in a given time to fluctuate. However, because a huge number of molecules collide with the wall in a short time, the number of collisions on the scales of time and space we measure fluctuates by only a tiny, usually unobservable fraction from the average. We can compare this situation to that of a casino, where the outcomes of the bets are random and the casino's takings fluctuate by the minute and the hour. However, over long times such as a year, the casino's takings are very close to the averages expected from the odds. A tank of gas has enormously more molecules than a casino has bettors in a year, and the molecules make enormously more collisions in a second than a casino has bets.

A calculation of the average force exerted by molecules on the walls of the box leads us to the ideal gas law and to the connection between temperature and molecular kinetic energy. (In fact, we will take two averages: one over time to get the average force exerted by one molecule with a given velocity, and then another average over molecules with different velocities.) This approach was developed by Daniel Bernoulli (1700–1782), who is best known in physics for his work on fluid flow (hydrodynamics). Remarkably, Bernoulli did this work before Dalton established the view of matter as consisting of atoms.

[\[link\]](#) shows a container full of gas and an expanded view of an elastic collision of a gas molecule with a wall of the container, broken down into components. We have assumed that a molecule is small compared with the separation of molecules in the gas, and that its interaction with other molecules can be ignored.

Under these conditions, the ideal gas law is experimentally valid. Because we have also assumed the wall is rigid and the particles are points, the collision is elastic (by conservation of energy—there's nowhere for a particle's kinetic energy to go). Therefore, the molecule's kinetic energy remains constant, and hence, its speed and the magnitude of its momentum remain constant as well. This assumption is not always valid, but the results in the rest of this module are also obtained in models that let the molecules exchange energy and momentum with the wall.



Gas in a box exerts an outward pressure on its walls. A molecule colliding with a rigid wall has its velocity and momentum in the x -direction reversed. This direction is perpendicular to the wall. The components of its velocity momentum in the y - and z -directions are not changed, which means there is no force parallel to the wall.

If the molecule's velocity changes in the x -direction, its momentum changes from $-mv_x$ to $+mv_x$. Thus, its change in momentum is $\Delta mv = +mv_x - (-mv_x) = 2mv_x$. According to the impulse-momentum theorem given in the chapter on linear momentum and collisions, the force exerted on the i th molecule, where i labels the molecules from 1 to N , is given by

Equation:

$$F_i = \frac{\Delta p_i}{\Delta t} = \frac{2mv_{ix}}{\Delta t}.$$

(In this equation alone, p represents momentum, not pressure.) There is no force between the wall and the molecule except while the molecule is touching the wall. During the short time of the collision, the force between the molecule and wall is relatively large, but that is not the force we are looking for. We are looking for the average force, so we take Δt to be the average time between collisions of the given molecule with this wall, which is the time in which we expect to find one collision. Let l represent the length of the box in the x -direction. Then Δt is the time the molecule would take to go across the box and back, a distance $2l$, at a speed of v_x . Thus $\Delta t = 2l/v_x$, and the expression for the force becomes

Equation:

$$F_i = \frac{2mv_{ix}}{2l/v_{ix}} = \frac{mv_{ix}^2}{l}.$$

This force is due to *one* molecule. To find the total force on the wall, F , we need to add the contributions of all N molecules:

Equation:

$$F = \sum_{i=1}^N F_i = \sum_{i=1}^N \frac{mv_{ix}^2}{l} = \frac{m}{l} \sum_{i=1}^N v_{ix}^2.$$

We now use the definition of the average, which we denote with a bar, to find the force:

Equation:

$$F = N \frac{m}{l} \left(\frac{1}{N} \sum_{i=1}^N v_{ix}^2 \right) = N \frac{m \bar{v_x^2}}{l}.$$

We want the force in terms of the speed v , rather than the x -component of the velocity. Note that the total velocity squared is the sum of the squares of its components, so that

Equation:

$$\bar{v^2} = \bar{v_x^2} + \bar{v_y^2} + \bar{v_z^2}.$$

With the assumption of isotropy, the three averages on the right side are equal, so

Equation:

$$\bar{v^2} = 3\bar{v_{ix}^2}.$$

Substituting this into the expression for F gives

Equation:

$$F = N \frac{m \bar{v^2}}{3l}.$$

The pressure is F/A , so we obtain

Equation:

$$p = \frac{F}{A} = N \frac{m\bar{v}^2}{3Al} = \frac{Nm\bar{v}^2}{3V},$$

where we used $V = Al$ for the volume. This gives the important result

Note:

Equation:

$$pV = \frac{1}{3} Nm\bar{v}^2.$$

Combining this equation with $pV = Nk_{\text{B}}T$ gives

Equation:

$$\frac{1}{3} Nm\bar{v}^2 = Nk_{\text{B}}T.$$

We can get the average kinetic energy of a molecule, $\frac{1}{2} m\bar{v}^2$, from the left-hand side of the equation by dividing out N and multiplying by $3/2$.

Note:

Average Kinetic Energy per Molecule

The average kinetic energy of a molecule is directly proportional to its absolute temperature:

Equation:

$$\bar{K} = \frac{1}{2} m\bar{v}^2 = \frac{3}{2} k_{\text{B}}T.$$

The equation $\bar{K} = \frac{3}{2} k_{\text{B}}T$ is the average kinetic energy per molecule. Note in particular that nothing in this equation depends on the molecular mass (or any other property) of the gas, the pressure, or anything but the temperature. If samples of helium and xenon gas, with very different molecular masses, are at the same temperature, the molecules have the same average kinetic energy.

The **internal energy** of a thermodynamic system is the sum of the mechanical energies of all of the molecules in it. We can now give an equation for the internal energy of a monatomic ideal gas. In such a gas, the molecules' only energy is their translational kinetic energy. Therefore, denoting the internal energy by E_{int} , we simply have $E_{\text{int}} = N\bar{K}$, or

Note:

Equation:

$$E_{\text{int}} = \frac{3}{2} N k_{\text{B}} T.$$

Often we would like to use this equation in terms of moles:

Equation:

$$E_{\text{int}} = \frac{3}{2} n R T.$$

We can solve $\bar{K} = \frac{1}{2} m \bar{v}^2 = \frac{3}{2} k_{\text{B}} T$ for a typical speed of a molecule in an ideal gas in terms of temperature to determine what is known as the *root-mean-square (rms) speed* of a molecule.

Note:

RMS Speed of a Molecule

The **root-mean-square (rms) speed** of a molecule, or the square root of the average of the square of the speed \bar{v}^2 , is

Equation:

$$v_{\text{rms}} = \sqrt{\bar{v}^2} = \sqrt{\frac{3 k_{\text{B}} T}{m}}.$$

The rms speed is not the average or the most likely speed of molecules, as we will see in [Distribution of Molecular Speeds](#), but it provides an easily calculated estimate of the molecules' speed that is related to their kinetic energy. Again we can write this equation in terms of the gas constant R and the molar mass M in kg/mol:

Note:

Equation:

$$v_{\text{rms}} = \sqrt{\frac{3 R T}{M}}.$$

We digress for a moment to answer a question that may have occurred to you: When we apply the model to atoms instead of theoretical point particles, does rotational kinetic energy change our results? To answer this question, we have to appeal to quantum mechanics. In quantum mechanics, rotational kinetic energy cannot take on just any value; it's limited to a discrete set of values, and the smallest value is inversely proportional to the rotational inertia. The rotational inertia of an atom is tiny because almost all

of its mass is in the nucleus, which typically has a radius less than 10^{-14} m. Thus the minimum rotational energy of an atom is much more than $\frac{1}{2} k_B T$ for any attainable temperature, and the energy available is not enough to make an atom rotate. We will return to this point when discussing diatomic and polyatomic gases in the next section.

Example:

Calculating Kinetic Energy and Speed of a Gas Molecule

(a) What is the average kinetic energy of a gas molecule at 20.0°C (room temperature)? (b) Find the rms speed of a nitrogen molecule (N_2) at this temperature.

Strategy

(a) The known in the equation for the average kinetic energy is the temperature:

Equation:

$$\bar{K} = \frac{1}{2} m \bar{v}^2 = \frac{3}{2} k_B T.$$

Before substituting values into this equation, we must convert the given temperature into kelvin:

$T = (20.0 + 273) \text{ K} = 293 \text{ K}$. We can find the rms speed of a nitrogen molecule by using the equation

Equation:

$$v_{\text{rms}} = \sqrt{\bar{v}^2} = \sqrt{\frac{3k_B T}{m}},$$

but we must first find the mass of a nitrogen molecule. Obtaining the molar mass of nitrogen N_2 from the periodic table, we find

Equation:

$$m = \frac{M}{N_A} = \frac{2 (14.0067) \times 10^{-3} \text{ kg/mol}}{6.02 \times 10^{23} \text{ mol}^{-1}} = 4.65 \times 10^{-26} \text{ kg}.$$

Solution

- a. The temperature alone is sufficient for us to find the average translational kinetic energy. Substituting the temperature into the translational kinetic energy equation gives

Equation:

$$\bar{K} = \frac{3}{2} k_B T = \frac{3}{2} (1.38 \times 10^{-23} \text{ J/K})(293 \text{ K}) = 6.07 \times 10^{-21} \text{ J}.$$

- b. Substituting this mass and the value for k_B into the equation for v_{rms} yields

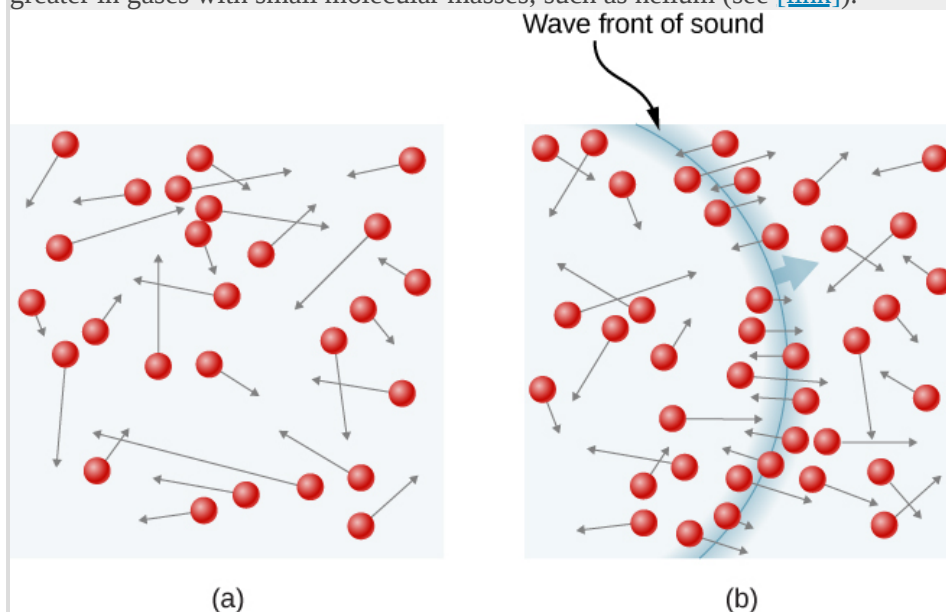
Equation:

$$v_{\text{rms}} = \sqrt{\frac{3k_B T}{m}} = \sqrt{\frac{3(1.38 \times 10^{-23} \text{ J/K})(293 \text{ K})}{4.65 \times 10^{-26} \text{ kg}}} = 511 \text{ m/s}.$$

Significance

Note that the average kinetic energy of the molecule is independent of the type of molecule. The average translational kinetic energy depends only on absolute temperature. The kinetic energy is very small compared to macroscopic energies, so that we do not feel when an air molecule is hitting our skin. On the

other hand, it is much greater than the typical difference in gravitational potential energy when a molecule moves from, say, the top to the bottom of a room, so our neglect of gravitation is justified in typical real-world situations. The rms speed of the nitrogen molecule is surprisingly large. These large molecular velocities do not yield macroscopic movement of air, since the molecules move in all directions with equal likelihood. The *mean free path* (the distance a molecule moves on average between collisions, discussed a bit later in this section) of molecules in air is very small, so the molecules move rapidly but do not get very far in a second. The high value for rms speed is reflected in the speed of sound, which is about 340 m/s at room temperature. The higher the rms speed of air molecules, the faster sound vibrations can be transferred through the air. The speed of sound increases with temperature and is greater in gases with small molecular masses, such as helium (see [\[link\]](#)).



(a) In an ordinary gas, so many molecules move so fast that they collide billions of times every second. (b) Individual molecules do not move very far in a small amount of time, but disturbances like sound waves are transmitted at speeds related to the molecular speeds.

Example:

Calculating Temperature: Escape Velocity of Helium Atoms

To escape Earth's gravity, an object near the top of the atmosphere (at an altitude of 100 km) must travel away from Earth at 11.1 km/s. This speed is called the *escape velocity*. At what temperature would helium atoms have an rms speed equal to the escape velocity?

Strategy

Identify the knowns and unknowns and determine which equations to use to solve the problem.

Solution

1. Identify the knowns: v is the escape velocity, 11.1 km/s.
2. Identify the unknowns: We need to solve for temperature, T . We also need to solve for the mass m of the helium atom.
3. Determine which equations are needed.

- To get the mass m of the helium atom, we can use information from the periodic table:

Equation:

$$m = \frac{M}{N_A}.$$

- To solve for temperature T , we can rearrange

Equation:

$$\frac{1}{2} m \bar{v}^2 = \frac{3}{2} k_B T$$

to yield

Equation:

$$T = \frac{m \bar{v}^2}{3 k_B}.$$

4. Substitute the known values into the equations and solve for the unknowns,

Equation:

$$m = \frac{M}{N_A} = \frac{4.0026 \times 10^{-3} \text{ kg/mol}}{6.02 \times 10^{23} \text{ mol}} = 6.65 \times 10^{-27} \text{ kg}$$

and

Equation:

$$T = \frac{(6.65 \times 10^{-27} \text{ kg}) (11.1 \times 10^3 \text{ m/s})^2}{3 (1.38 \times 10^{-23} \text{ J/K})} = 1.98 \times 10^4 \text{ K}.$$

Significance

This temperature is much higher than atmospheric temperature, which is approximately 250 K (-25°C or -10°F) at high elevation. Very few helium atoms are left in the atmosphere, but many were present when the atmosphere was formed, and more are always being created by radioactive decay (see the chapter on nuclear physics). The reason for the loss of helium atoms is that a small number of helium atoms have speeds higher than Earth's escape velocity even at normal temperatures. The speed of a helium atom changes from one collision to the next, so that at any instant, there is a small but nonzero chance that the atom's speed is greater than the escape velocity. The chance is high enough that over the lifetime of Earth, almost all the helium atoms that have been in the atmosphere have reached escape velocity at high altitudes and escaped from Earth's gravitational pull. Heavier molecules, such as oxygen, nitrogen, and water, have smaller rms speeds, and so it is much less likely that any of them will have speeds greater than the escape velocity. In fact, the likelihood is so small that billions of years are required to lose significant amounts of heavier molecules from the atmosphere. [\[link\]](#) shows the effect of a lack of an atmosphere on the Moon. Because the gravitational pull of the Moon is much weaker, it has lost almost its entire atmosphere. The atmospheres of Earth and other bodies are compared in this chapter's exercises.



This photograph of Apollo 17 Commander Eugene Cernan driving the lunar rover on the Moon in 1972 looks as though it was taken at night with a large spotlight. In fact, the light is coming from the Sun. Because the acceleration due to gravity on the Moon is so low (about $1/6$ that of Earth), the Moon's escape velocity is much smaller. As a result, gas molecules escape very easily from the Moon, leaving it with virtually no atmosphere. Even during the daytime, the sky is black because there is no gas to scatter sunlight.
(credit: Harrison H. Schmitt/NASA)

Note:

Exercise:

Problem:

Check Your Understanding If you consider a very small object, such as a grain of pollen, in a gas, then the number of molecules striking its surface would also be relatively small. Would you expect the grain of pollen to experience any fluctuations in pressure due to statistical fluctuations in the number of gas molecules striking it in a given amount of time?

Solution:

Yes. Such fluctuations actually occur for a body of any size in a gas, but since the numbers of molecules are immense for macroscopic bodies, the fluctuations are a tiny percentage of the number

of collisions, and the averages spoken of in this section vary imperceptibly. Roughly speaking, the fluctuations are inversely proportional to the square root of the number of collisions, so for small bodies, they can become significant. This was actually observed in the nineteenth century for pollen grains in water and is known as Brownian motion.

Vapor Pressure, Partial Pressure, and Dalton's Law

The pressure a gas would create if it occupied the total volume available is called the gas's **partial pressure**. If two or more gases are mixed, they will come to thermal equilibrium as a result of collisions between molecules; the process is analogous to heat conduction as described in the chapter on temperature and heat. As we have seen from kinetic theory, when the gases have the same temperature, their molecules have the same average kinetic energy. Thus, each gas obeys the ideal gas law separately and exerts the same pressure on the walls of a container that it would if it were alone. Therefore, in a mixture of gases, *the total pressure is the sum of partial pressures of the component gases*, assuming ideal gas behavior and no chemical reactions between the components. This law is known as **Dalton's law of partial pressures**, after the English scientist John Dalton (1766–1844) who proposed it. Dalton's law is consistent with the fact that pressures add according to Pascal's principle.

In a mixture of ideal gases in thermal equilibrium, the number of molecules of each gas is proportional to its partial pressure. This result follows from applying the ideal gas law to each in the form $p/n = RT/V$. Because the right-hand side is the same for any gas at a given temperature in a container of a given volume, the left-hand side is the same as well.

- Partial pressure is the pressure a gas would create if it existed alone.
- Dalton's law states that the total pressure is the sum of the partial pressures of all of the gases present.
- For any two gases (labeled 1 and 2) in equilibrium in a container, $\frac{p_1}{n_1} = \frac{p_2}{n_2}$.

An important application of partial pressure is that, in chemistry, it functions as the concentration of a gas in determining the rate of a reaction. Here, we mention only that the partial pressure of oxygen in a person's lungs is crucial to life and health. Breathing air that has a partial pressure of oxygen below 0.16 atm can impair coordination and judgment, particularly in people not acclimated to a high elevation. Lower partial pressures of O_2 have more serious effects; partial pressures below 0.06 atm can be quickly fatal, and permanent damage is likely even if the person is rescued. However, the sensation of needing to breathe, as when holding one's breath, is caused much more by high concentrations of carbon dioxide in the blood than by low concentrations of oxygen. Thus, if a small room or closet is filled with air having a low concentration of oxygen, perhaps because a leaking cylinder of some compressed gas is stored there, a person will not feel any "choking" sensation and may go into convulsions or lose consciousness without noticing anything wrong. Safety engineers give considerable attention to this danger.

Another important application of partial pressure is **vapor pressure**, which is the partial pressure of a vapor at which it is in equilibrium with the liquid (or solid, in the case of sublimation) phase of the same substance. At any temperature, the partial pressure of the water in the air cannot exceed the vapor pressure of the water at that temperature, because whenever the partial pressure reaches the vapor pressure, water condenses out of the air. Dew is an example of this condensation. The temperature at which condensation occurs for a sample of air is called the *dew point*. It is easily measured by slowly cooling a metal ball; the dew point is the temperature at which condensation first appears on the ball.

The vapor pressures of water at some temperatures of interest for meteorology are given in [\[link\]](#).

T (°C)	Vapor Pressure (Pa)
0	610.5
3	757.9
5	872.3
8	1073
10	1228
13	1497
15	1705
18	2063
20	2338
23	2809
25	3167
30	4243
35	5623
40	7376

Vapor Pressure of Water at Various Temperatures

The *relative humidity* (R.H.) at a temperature T is defined by

Equation:

$$\text{R.H.} = \frac{\text{Partial pressure of water vapor at } T}{\text{Vapor pressure of water at } T} \times 100\%.$$

A relative humidity of 100 % means that the partial pressure of water is equal to the vapor pressure; in other words, the air is saturated with water.

Example:

Calculating Relative Humidity

What is the relative humidity when the air temperature is 25 °C and the dew point is 15 °C?

Strategy

We simply look up the vapor pressure at the given temperature and that at the dew point and find the ratio.

Solution

Equation:

$$\text{R.H.} = \frac{\text{Partial pressure of water vapor at } 15^\circ\text{C}}{\text{Partial pressure of water vapor at } 25^\circ\text{C}} \times 100\% = \frac{1705 \text{ Pa}}{3167 \text{ Pa}} \times 100\% = 53.8\%.$$

Significance

R.H. is important to our comfort. The value of 53.8% is within the range of 40% to 60% recommended for comfort indoors.

As noted in the chapter on temperature and heat, the temperature seldom falls below the dew point, because when it reaches the dew point or frost point, water condenses and releases a relatively large amount of latent heat of vaporization.

Mean Free Path and Mean Free Time

We now consider collisions explicitly. The usual first step (which is all we'll take) is to calculate the **mean free path**, λ , the average distance a molecule travels between collisions with other molecules, and the *mean free time* τ , the average time between the collisions of a molecule. If we assume all the molecules are spheres with a radius r , then a molecule will collide with another if their centers are within a distance $2r$ of each other. For a given particle, we say that the area of a circle with that radius, $4\pi r^2$, is the "cross-section" for collisions. As the particle moves, it traces a cylinder with that cross-sectional area. The mean free path is the length λ such that the expected number of other molecules in a cylinder of length λ and cross-section $4\pi r^2$ is 1. If we temporarily ignore the motion of the molecules other than the one we're looking at, the expected number is the number density of molecules, N/V , times the volume, and the volume is $4\pi r^2\lambda$, so we have $(N/V)4\pi r^2\lambda = 1$, or

Equation:

$$\lambda = \frac{V}{4\pi r^2 N}.$$

Taking the motion of all the molecules into account makes the calculation much harder, but the only change is a factor of $\sqrt{2}$. The result is

Note:

Equation:

$$\lambda = \frac{V}{4\sqrt{2}\pi r^2 N}.$$

In an ideal gas, we can substitute $V/N = k_B T/p$ to obtain

Note:

Equation:

$$\lambda = \frac{k_B T}{4\sqrt{2}\pi r^2 p}.$$

The **mean free time** τ is simply the mean free path divided by a typical speed, and the usual choice is the rms speed. Then

Note:

Equation:

$$\tau = \frac{k_{\text{B}}T}{4\sqrt{2}\pi r^2 p v_{\text{rms}}}.$$

Example:

Calculating Mean Free Time

Find the mean free time for argon atoms ($M = 39.9 \text{ g/mol}$) at a temperature of 0°C and a pressure of 1.00 atm . Take the radius of an argon atom to be $1.70 \times 10^{-10} \text{ m}$.

Solution

1. Identify the knowns and convert into SI units. We know the molar mass is 0.0399 kg/mol , the temperature is 273 K , the pressure is $1.01 \times 10^5 \text{ Pa}$, and the radius is $1.70 \times 10^{-10} \text{ m}$.
2. Find the rms speed: $v_{\text{rms}} = \sqrt{\frac{3RT}{M}} = 413 \frac{\text{m}}{\text{s}}$.
3. Substitute into the equation for the mean free time:

Equation:

$$\tau = \frac{k_{\text{B}}T}{4\sqrt{2}\pi r^2 p v_{\text{rms}}} = \frac{(1.38 \times 10^{-23} \text{ J/K})(273 \text{ K})}{4\sqrt{2}\pi(1.70 \times 10^{-10} \text{ m})^2(1.01 \times 10^5 \text{ Pa})(413 \text{ m/s})} = 1.76 \times 10^{-10} \text{ s}.$$

Significance

We can hardly compare this result with our intuition about gas molecules, but it gives us a picture of molecules colliding with extremely high frequency.

Note:

Exercise:

Problem:

Check Your Understanding Which has a longer mean free path, liquid water or water vapor in the air?

Solution:

In a liquid, the molecules are very close together, constantly colliding with one another. For a gas to be nearly ideal, as air is under ordinary conditions, the molecules must be very far apart. Therefore

the mean free path is much longer in the air.

Summary

- Kinetic theory is the atomic description of gases as well as liquids and solids. It models the properties of matter in terms of continuous random motion of molecules.
- The ideal gas law can be expressed in terms of the mass of the gas's molecules and \bar{v}^2 , the average of the molecular speed squared, instead of the temperature.
- The temperature of gases is proportional to the average translational kinetic energy of molecules. Hence, the typical speed of gas molecules v_{rms} is proportional to the square root of the temperature and inversely proportional to the square root of the molecular mass.
- In a mixture of gases, each gas exerts a pressure equal to the total pressure times the fraction of the mixture that the gas makes up.
- The mean free path (the average distance between collisions) and the mean free time of gas molecules are proportional to the temperature and inversely proportional to the molar density and the molecules' cross-sectional area.

Conceptual Questions

Exercise:

Problem:

How is momentum related to the pressure exerted by a gas? Explain on the molecular level, considering the behavior of molecules.

Exercise:

Problem:

If one kind of molecule has double the radius of another and eight times the mass, how do their mean free paths under the same conditions compare? How do their mean free times compare?

Solution:

The mean free path is inversely proportional to the square of the radius, so it decreases by a factor of 4. The mean free time is proportional to the mean free path and inversely proportional to the rms speed, which in turn is inversely proportional to the square root of the mass. That gives a factor of $\sqrt{8}$ in the numerator, so the mean free time decreases by a factor of $\sqrt{2}$.

Exercise:

Problem: What is the average *velocity* of the air molecules in the room where you are right now?

Exercise:

Problem:

Why do the atmospheres of Jupiter, Saturn, Uranus, and Neptune, which are much more massive and farther from the Sun than Earth is, contain large amounts of hydrogen and helium?

Solution:

Since they're more massive, their gravity is stronger, so the escape velocity from them is higher. Since they're farther from the Sun, they're colder, so the speeds of atmospheric molecules including hydrogen and helium are lower. The combination of those facts means that relatively few hydrogen and helium molecules have escaped from the outer planets.

Exercise:

Problem:

Statistical mechanics says that in a gas maintained at a constant temperature through thermal contact with a bigger system (a "reservoir") at that temperature, the fluctuations in internal energy are typically a fraction $1/\sqrt{N}$ of the internal energy. As a fraction of the total internal energy of a mole of gas, how big are the fluctuations in the internal energy? Are we justified in ignoring them?

Exercise:

Problem:

Which is more dangerous, a closet where tanks of nitrogen are stored, or one where tanks of carbon dioxide are stored?

Solution:

One where nitrogen is stored, as excess CO₂ will cause a feeling of suffocating, but excess nitrogen and insufficient oxygen will not.

Problems

In the problems in this section, assume all gases are ideal.

Exercise:

Problem:

A person hits a tennis ball with a mass of 0.058 kg against a wall. The average component of the ball's velocity perpendicular to the wall is 11 m/s, and the ball hits the wall every 2.1 s on average, rebounding with the opposite perpendicular velocity component. (a) What is the average force exerted on the wall? (b) If the part of the wall the person hits has an area of 3.0 m², what is the average pressure on that area?

Solution:

a. 0.61 N; b. 0.20 Pa

Exercise:

Problem:

A person is in a closed room (a racquetball court) with $V = 453 \text{ m}^3$ hitting a ball ($m = 42.0 \text{ g}$) around at random without any pauses. The average kinetic energy of the ball is 2.30 J. (a) What is the average value of v_x^2 ? Does it matter which direction you take to be x ? (b) Applying the methods of this chapter, find the average pressure on the walls? (c) Aside from the presence of only one "molecule" in this problem, what is the main assumption in [Pressure, Temperature, and RMS Speed](#) that does not apply here?

Exercise:

Problem:

Five bicyclists are riding at the following speeds: 5.4 m/s, 5.7 m/s, 5.8 m/s, 6.0 m/s, and 6.5 m/s. (a) What is their average speed? (b) What is their rms speed?

Solution:

a. 5.88 m/s; b. 5.89 m/s

Exercise:**Problem:**

Some incandescent light bulbs are filled with argon gas. What is v_{rms} for argon atoms near the filament, assuming their temperature is 2500 K?

Exercise:**Problem:**

Typical molecular speeds (v_{rms}) are large, even at low temperatures. What is v_{rms} for helium atoms at 5.00 K, less than one degree above helium's liquefaction temperature?

Solution:

177 m/s

Exercise:**Problem:**

What is the average kinetic energy in joules of hydrogen atoms on the 5500 °C surface of the Sun?
(b) What is the average kinetic energy of helium atoms in a region of the solar corona where the temperature is 6.00×10^5 K?

Exercise:**Problem:**

What is the ratio of the average translational kinetic energy of a nitrogen molecule at a temperature of 300 K to the gravitational potential energy of a nitrogen-molecule–Earth system at the ceiling of a 3-m-tall room with respect to the same system with the molecule at the floor?

Solution:

4.54×10^3

Exercise:**Problem:**

What is the total translational kinetic energy of the air molecules in a room of volume 23 m³ if the pressure is 9.5×10^4 Pa (the room is at fairly high elevation) and the temperature is 21 °C? Is any item of data unnecessary for the solution?

Exercise:

Problem:

The product of the pressure and volume of a sample of hydrogen gas at $0.00\text{ }^{\circ}\text{C}$ is 80.0 J . (a) How many moles of hydrogen are present? (b) What is the average translational kinetic energy of the hydrogen molecules? (c) What is the value of the product of pressure and volume at $200\text{ }^{\circ}\text{C}$?

Solution:

a. 0.0352 mol ; b. $5.65 \times 10^{-21}\text{ J}$; c. 139 J

Exercise:**Problem:**

What is the gauge pressure inside a tank of $4.86 \times 10^4\text{ mol}$ of compressed nitrogen with a volume of 6.56 m^3 if the rms speed is 514 m/s ?

Exercise:**Problem:**

If the rms speed of oxygen molecules inside a refrigerator of volume 22.0 ft^3 is 465 m/s , what is the partial pressure of the oxygen? There are 5.71 moles of oxygen in the refrigerator, and the molar mass of oxygen is 32.0 g/mol .

Solution:

21.1 kPa

Exercise:**Problem:**

The escape velocity of any object from Earth is 11.1 km/s . At what temperature would oxygen molecules (molar mass is equal to 32.0 g/mol) have root-mean-square velocity v_{rms} equal to Earth's escape velocity of 11.1 km/s ?

Exercise:**Problem:**

The escape velocity from the Moon is much smaller than that from the Earth, only 2.38 km/s . At what temperature would hydrogen molecules (molar mass is equal to 2.016 g/mol) have a root-mean-square velocity v_{rms} equal to the Moon's escape velocity?

Solution:

458 K

Exercise:**Problem:**

Nuclear fusion, the energy source of the Sun, hydrogen bombs, and fusion reactors, occurs much more readily when the average kinetic energy of the atoms is high—that is, at high temperatures. Suppose you want the atoms in your fusion experiment to have average kinetic energies of $6.40 \times 10^{-14}\text{ J}$. What temperature is needed?

Exercise:**Problem:**

Suppose that the typical speed (v_{rms}) of carbon dioxide molecules (molar mass is 44.0 g/mol) in a flame is found to be 1350 m/s. What temperature does this indicate?

Solution:

$$3.22 \times 10^3 \text{ K}$$

Exercise:**Problem:**

(a) Hydrogen molecules (molar mass is equal to 2.016 g/mol) have v_{rms} equal to 193 m/s. What is the temperature? (b) Much of the gas near the Sun is atomic hydrogen (H rather than H_2). Its temperature would have to be $1.5 \times 10^7 \text{ K}$ for the rms speed v_{rms} to equal the escape velocity from the Sun. What is that velocity?

Exercise:**Problem:**

There are two important isotopes of uranium, ^{235}U and ^{238}U ; these isotopes are nearly identical chemically but have different atomic masses. Only ^{235}U is very useful in nuclear reactors. Separating the isotopes is called uranium enrichment (and is often in the news as of this writing, because of concerns that some countries are enriching uranium with the goal of making nuclear weapons.) One of the techniques for enrichment, gas diffusion, is based on the different molecular speeds of uranium hexafluoride gas, UF_6 . (a) The molar masses of ^{235}U and $^{238}\text{UF}_6$ are 349.0 g/mol and 352.0 g/mol, respectively. What is the ratio of their typical speeds v_{rms} ? (b) At what temperature would their typical speeds differ by 1.00 m/s? (c) Do your answers in this problem imply that this technique may be difficult?

Solution:

a. 1.004; b. 764 K; c. This temperature is equivalent to 915 °F, which is high but not impossible to achieve. Thus, this process is feasible. At this temperature, however, there may be other considerations that make the process difficult. (In general, uranium enrichment by gaseous diffusion is indeed difficult and requires many passes.)

Exercise:**Problem:**

The partial pressure of carbon dioxide in the lungs is about 470 Pa when the total pressure in the lungs is 1.0 atm. What percentage of the air molecules in the lungs is carbon dioxide? Compare your result to the percentage of carbon dioxide in the atmosphere, about 0.033%.

Exercise:**Problem:**

Dry air consists of approximately 78% nitrogen, 21% oxygen, and 1% argon by mole, with trace amounts of other gases. A tank of compressed dry air has a volume of 1.76 cubic feet at a gauge pressure of 2200 pounds per square inch and a temperature of 293 K. How much oxygen does it contain in moles?

Solution:

65 mol

Exercise:**Problem:**

(a) Using data from the previous problem, find the mass of nitrogen, oxygen, and argon in 1 mol of dry air. The molar mass of N_2 is 28.0 g/mol, that of O_2 is 32.0 g/mol, and that of argon is 39.9 g/mol. (b) Dry air is mixed with pentane (C_5H_{12} , molar mass 72.2 g/mol), an important constituent of gasoline, in an air-fuel ratio of 15:1 by mass (roughly typical for car engines). Find the partial pressure of pentane in this mixture at an overall pressure of 1.00 atm.

Exercise:**Problem:**

(a) Given that air is 21 % oxygen, find the minimum atmospheric pressure that gives a relatively safe partial pressure of oxygen of 0.16 atm. (b) What is the minimum pressure that gives a partial pressure of oxygen above the quickly fatal level of 0.06 atm? (c) The air pressure at the summit of Mount Everest (8848 m) is 0.334 atm. Why have a few people climbed it without oxygen, while some who have tried, even though they had trained at high elevation, had to turn back?

Solution:

a. 0.76 atm; b. 0.29 atm; c. The pressure there is barely above the quickly fatal level.

Exercise:**Problem:**

(a) If the partial pressure of water vapor is 8.05 torr, what is the dew point? (760 torr = 1 atm = 101,325 Pa) (b) On a warm day when the air temperature is 35 °C and the dew point is 25 °C, what are the partial pressure of the water in the air and the relative humidity?

Glossary

Dalton's law of partial pressures

physical law that states that the total pressure of a gas is the sum of partial pressures of the component gases

internal energy

sum of the mechanical energies of all of the molecules in it

kinetic theory of gases

theory that derives the macroscopic properties of gases from the motion of the molecules they consist of

mean free path

average distance between collisions of a particle

mean free time

average time between collisions of a particle

partial pressure

pressure a gas would create if it occupied the total volume of space available

root-mean-square (rms) speed

square root of the average of the square (of a quantity)

vapor pressure

partial pressure of a vapor at which it is in equilibrium with the liquid (or solid, in the case of sublimation) phase of the same substance

Heat Capacity and Equipartition of Energy

By the end of this section, you will be able to:

- Solve problems involving heat transfer to and from ideal monatomic gases whose volumes are held constant
- Solve similar problems for non-monatomic ideal gases based on the number of degrees of freedom of a molecule
- Estimate the heat capacities of metals using a model based on degrees of freedom

In the chapter on temperature and heat, we defined the specific heat capacity with the equation $Q = mc\Delta T$, or $c = (1/m)Q/\Delta T$. However, the properties of an ideal gas depend directly on the number of moles in a sample, so here we define specific heat capacity in terms of the number of moles, not the mass. Furthermore, when talking about solids and liquids, we ignored any changes in volume and pressure with changes in temperature—a good approximation for solids and liquids, but for gases, we have to make some condition on volume or pressure changes. Here, we focus on the heat capacity with the volume held constant. We can calculate it for an ideal gas.

Heat Capacity of an Ideal Monatomic Gas at Constant Volume

We define the *molar heat capacity at constant volume* C_V as

Equation:

$$C_V = \frac{1}{n} \frac{Q}{\Delta T}, \text{ with } V \text{ held constant.}$$

This is often expressed in the form

Note:

Equation:

$$Q = nC_V\Delta T.$$

If the volume does not change, there is no overall displacement, so no work is done, and the only change in internal energy is due to the heat flow $\Delta E_{\text{int}} = Q$. (This

statement is discussed further in the next chapter.) We use the equation $E_{\text{int}} = 3nRT/2$ to write $\Delta E_{\text{int}} = 3nR\Delta T/2$ and substitute ΔE for Q to find $Q = 3nR\Delta T/2$, which gives the following simple result for an ideal monatomic gas:

Equation:

$$C_V = \frac{3}{2}R.$$

It is independent of temperature, which justifies our use of finite differences instead of a derivative. This formula agrees well with experimental results.

In the next chapter we discuss the molar specific heat at constant pressure C_p , which is always greater than C_V .

Example:

Calculating Temperature

A sample of 0.125 kg of xenon is contained in a rigid metal cylinder, big enough that the xenon can be modeled as an ideal gas, at a temperature of 20.0 °C. The cylinder is moved outside on a hot summer day. As the xenon comes into equilibrium by reaching the temperature of its surroundings, 180 J of heat are conducted to it through the cylinder walls. What is the equilibrium temperature? Ignore the expansion of the metal cylinder.

Solution

1. Identify the knowns: We know the initial temperature T_1 is 20.0 °C, the heat Q is 180 J, and the mass m of the xenon is 0.125 kg.
2. Identify the unknown. We need the final temperature, so we'll need ΔT .
3. Determine which equations are needed. Because xenon gas is monatomic, we can use $Q = 3nR\Delta T/2$. Then we need the number of moles, $n = m/M$.
4. Substitute the known values into the equations and solve for the unknowns. The molar mass of xenon is 131.3 g, so we obtain

Equation:

$$n = \frac{125 \text{ g}}{131.3 \text{ g/mol}} = 0.952 \text{ mol},$$

Equation:

$$\Delta T = \frac{2Q}{3nR} = \frac{2(180 \text{ J})}{3(0.952 \text{ mol})(8.31 \text{ J/mol} \cdot ^\circ\text{C})} = 15.2 ^\circ\text{C}.$$

Therefore, the final temperature is $35.2 ^\circ\text{C}$. The problem could equally well be solved in kelvin; as a kelvin is the same size as a degree Celsius of temperature change, you would get $\Delta T = 15.2 \text{ K}$.

Significance

The heating of an ideal or almost ideal gas at constant volume is important in car engines and many other practical systems.

Note:

Exercise:

Problem:

Check Your Understanding Suppose 2 moles of helium gas at 200 K are mixed with 2 moles of krypton gas at 400 K in a calorimeter. What is the final temperature?

Solution:

As the number of moles is equal and we know the molar heat capacities of the two gases are equal, the temperature is halfway between the initial temperatures, 300 K.

We would like to generalize our results to ideal gases with more than one atom per molecule. In such systems, the molecules can have other forms of energy beside translational kinetic energy, such as rotational kinetic energy and vibrational kinetic and potential energies. We will see that a simple rule lets us determine the average energies present in these forms and solve problems in much the same way as we have for monatomic gases.

Degrees of Freedom

In the previous section, we found that $\frac{1}{2}m\bar{v}^2 = \frac{3}{2}k_B T$ and $\bar{v}^2 = 3\bar{v}_x^2$, from which it follows that $\frac{1}{2}m\bar{v}_x^2 = \frac{1}{2}k_B T$. The same equation holds for \bar{v}_y^2 and for \bar{v}_z^2 . Thus, we

can look at our energy of $\frac{3}{2}k_{\text{B}}T$ as the sum of contributions of $\frac{1}{2}k_{\text{B}}T$ from each of the three dimensions of translational motion. Shifting to the gas as a whole, we see that the 3 in the formula $C_V = \frac{3}{2}R$ also reflects those three dimensions. We define a **degree of freedom** as an independent possible motion of a molecule, such as each of the three dimensions of translation. Then, letting d represent the number of degrees of freedom, the molar heat capacity at constant volume of a monatomic ideal gas is $C_V = \frac{d}{2}R$, where $d = 3$.

The branch of physics called *statistical mechanics* tells us, and experiment confirms, that C_V of any ideal gas is given by this equation, regardless of the number of degrees of freedom. This fact follows from a more general result, the **equipartition theorem**, which holds in classical (non-quantum) thermodynamics for systems in thermal equilibrium under technical conditions that are beyond our scope. Here, we mention only that in a system, the energy is shared among the degrees of freedom by collisions.

Note:

Equipartition Theorem

The energy of a thermodynamic system in equilibrium is partitioned equally among its degrees of freedom. Accordingly, the molar heat capacity of an ideal gas is proportional to its number of degrees of freedom, d :

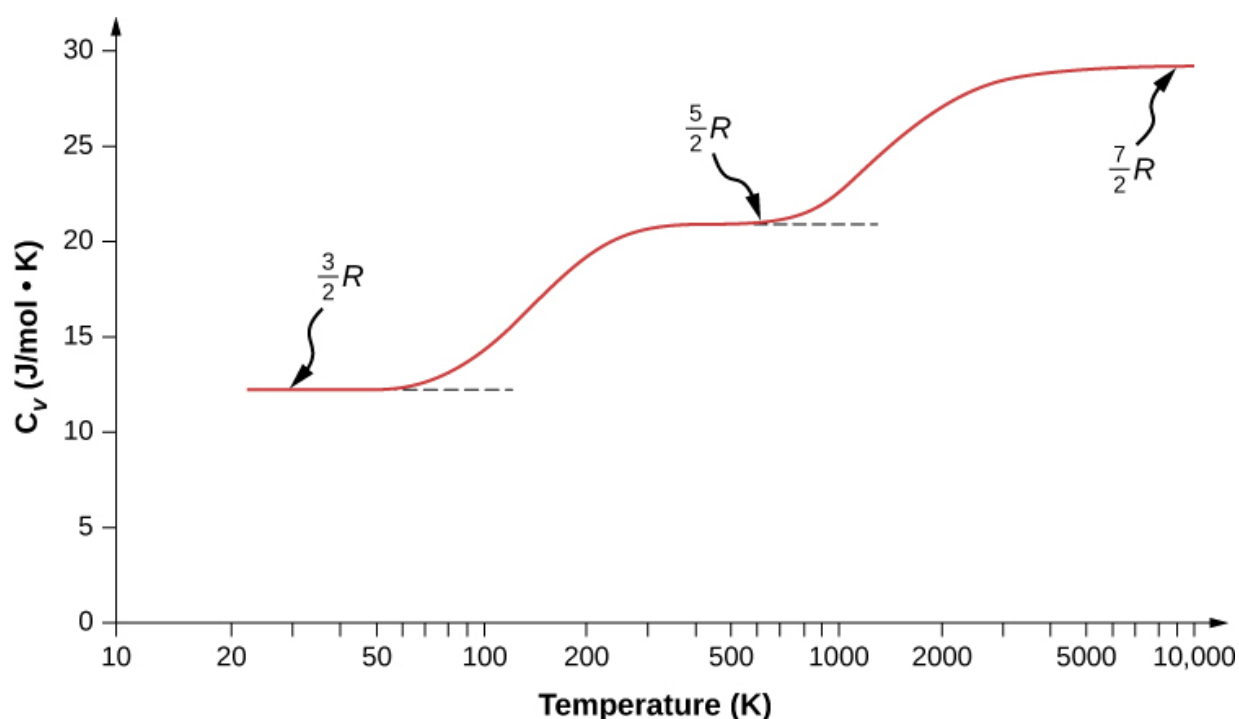
Equation:

$$C_V = \frac{d}{2}R.$$

This result is due to the Scottish physicist James Clerk Maxwell (1831–1871), whose name will appear several more times in this book.

For example, consider a diatomic ideal gas (a good model for nitrogen, N_2 , and oxygen, O_2). Such a gas has more degrees of freedom than a monatomic gas. In addition to the three degrees of freedom for translation, it has two degrees of freedom for rotation perpendicular to its axis. Furthermore, the molecule can vibrate along its axis. This motion is often modeled by imagining a spring connecting the two atoms, and we know from simple harmonic motion that such motion has both kinetic and potential energy. Each of these forms of energy corresponds to a degree of freedom, giving two more.

We might expect that for a diatomic gas, we should use 7 as the number of degrees of freedom; classically, if the molecules of a gas had only translational kinetic energy, collisions between molecules would soon make them rotate and vibrate. However, as explained in the previous module, quantum mechanics controls which degrees of freedom are active. The result is shown in [\[link\]](#). Both rotational and vibrational energies are limited to discrete values. For temperatures below about 60 K, the energies of hydrogen molecules are too low for a collision to bring the rotational state or vibrational state of a molecule from the lowest energy to the second lowest, so the only form of energy is translational kinetic energy, and $d = 3$ or $C_V = 3R/2$ as in a monatomic gas. Above that temperature, the two rotational degrees of freedom begin to contribute, that is, some molecules are excited to the rotational state with the second-lowest energy. (This temperature is much lower than that where rotations of monatomic gases contribute, because diatomic molecules have much higher rotational inertias and hence much lower rotational energies.) From about room temperature (a bit less than 300 K) to about 600 K, the rotational degrees of freedom are fully active, but the vibrational ones are not, and $d = 5$. Then, finally, above about 3000 K, the vibrational degrees of freedom are fully active, and $d = 7$ as the classical theory predicted.



The molar heat capacity of hydrogen as a function of temperature (on a logarithmic scale). The three “steps” or “plateaus” show different numbers of degrees of freedom that the typical energies of molecules must achieve to

activate. Translational kinetic energy corresponds to three degrees of freedom, rotational to another two, and vibrational to yet another two.

Polyatomic molecules typically have one additional rotational degree of freedom at room temperature, since they have comparable moments of inertia around any axis. Thus, at room temperature, they have $d = 6$, and at high temperature, $d = 8$. We usually assume that gases have the theoretical room-temperature values of d .

As shown in [\[link\]](#), the results agree well with experiments for many monatomic and diatomic gases, but the agreement for triatomic gases is only fair. The differences arise from interactions that we have ignored between and within molecules.

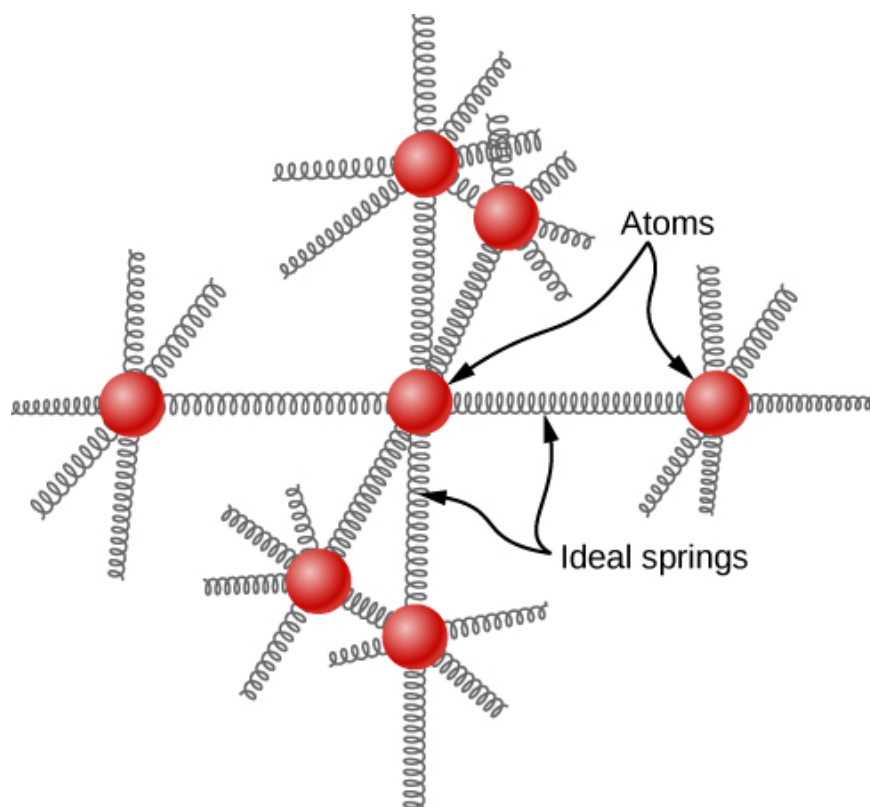
Gas	C_V/R at 25 °C and 1 atm
Ar	1.50
He	1.50
Ne	1.50
CO	2.50
H ₂	2.47
N ₂	2.50
O ₂	2.53
F ₂	2.8
CO ₂	3.48
H ₂ S	3.13
N ₂ O	3.66

C_V/R for Various Monatomic, Diatomic, and Triatomic Gases

What about internal energy for diatomic and polyatomic gases? For such gases, C_V is a function of temperature ([link](#)), so we do not have the kind of simple result we have for monatomic ideal gases.

Molar Heat Capacity of Solid Elements

The idea of equipartition leads to an estimate of the molar heat capacity of solid elements at ordinary temperatures. We can model the atoms of a solid as attached to neighboring atoms by springs ([link](#)).



In a simple model of a solid element, each atom is attached to others by six springs, two for each possible motion: x , y , and z . Each of the three motions corresponds to two degrees of freedom, one for kinetic energy and one for potential energy. Thus $d = 6$.

Analogously to the discussion of vibration in the previous module, each atom has six degrees of freedom: one kinetic and one potential for each of the x -, y -, and z -directions. Accordingly, the molar specific heat of a metal should be $3R$. This result, known as the Law of Dulong and Petit, works fairly well experimentally at room temperature. (For every element, it fails at low temperatures for quantum-mechanical reasons. Since quantum effects are particularly important for low-mass particles, the Law of Dulong and Petit already fails at room temperature for some light elements, such as beryllium and carbon. It also fails for some heavier elements for various reasons beyond what we can cover.)

Note:**Heat Capacity and Equipartition**

The strategy for solving these problems is the same as the one in [Phase Changes](#) for the effects of heat transfer. The only new feature is that you should determine whether the case just presented—ideal gases at constant volume—applies to the problem. (For solid elements, looking up the specific heat capacity is generally better than estimating it from the Law of Dulong and Petit.) In the case of an ideal gas, determine the number d of degrees of freedom from the number of atoms in the gas molecule and use it to calculate C_V (or use C_V to solve for d).

Example:**Calculating Temperature: Calorimetry with an Ideal Gas**

A 300-g piece of solid gallium (a metal used in semiconductor devices) at its melting point of only 30.0°C is in contact with 12.0 moles of air (assumed diatomic) at 95.0°C in an insulated container. When the air reaches equilibrium with the gallium, 202 g of the gallium have melted. Based on those data, what is the heat of fusion of gallium? Assume the volume of the air does not change and there are no other heat transfers.

Strategy

We'll use the equation $Q_{\text{hot}} + Q_{\text{cold}} = 0$. As some of the gallium doesn't melt, we know the final temperature is still the melting point. Then the only Q_{hot} is the heat lost as the air cools, $Q_{\text{hot}} = n_{\text{air}}C_V\Delta T$, where $C_V = 5R/2$. The only Q_{cold} is the latent heat of fusion of the gallium, $Q_{\text{cold}} = m_{\text{Ga}}L_f$. It is positive because heat flows into the gallium.

Solution

1. Set up the equation:

Equation:

$$n_{\text{air}} C_V \Delta T + m_{\text{Ga}} L_f = 0.$$

2. Substitute the known values and solve:

Equation:

$$(12.0 \text{ mol}) \left(\frac{5}{2} \right) \left(8.31 \frac{\text{J}}{\text{mol} \cdot ^\circ\text{C}} \right) (30.0 ^\circ\text{C} - 95.0 ^\circ\text{C}) + (0.202 \text{ kg}) L_f = 0.$$

We solve to find that the heat of fusion of gallium is 80.2 kJ/kg.

Summary

- Every degree of freedom of an ideal gas contributes $\frac{1}{2} k_B T$ per atom or molecule to its changes in internal energy.
- Every degree of freedom contributes $\frac{1}{2} R$ to its molar heat capacity at constant volume C_V .
- Degrees of freedom do not contribute if the temperature is too low to excite the minimum energy of the degree of freedom as given by quantum mechanics. Therefore, at ordinary temperatures, $d = 3$ for monatomic gases, $d = 5$ for diatomic gases, and $d \approx 6$ for polyatomic gases.

Conceptual Questions

Exercise:

Problem:

Experimentally it appears that many polyatomic molecules' vibrational degrees of freedom can contribute to some extent to their energy at room temperature. Would you expect that fact to increase or decrease their heat capacity from the value R ? Explain.

Exercise:

Problem:

One might think that the internal energy of diatomic gases is given by $E_{\text{int}} = 5RT/2$. Do diatomic gases near room temperature have more or less internal energy than that? *Hint:* Their internal energy includes the total energy added in raising the temperature from the boiling point (very low) to room temperature.

Solution:

Less, because at lower temperatures their heat capacity was only $3RT/2$.

Exercise:**Problem:**

You mix 5 moles of H_2 at 300 K with 5 moles of He at 360 K in a perfectly insulated calorimeter. Is the final temperature higher or lower than 330 K?

Problems**Exercise:****Problem:**

To give a helium atom nonzero angular momentum requires about 21.2 eV of energy (that is, 21.2 eV is the difference between the energies of the lowest-energy or ground state and the lowest-energy state with angular momentum). The electron-volt or eV is defined as 1.60×10^{-19} J. Find the temperature T where this amount of energy equals $k_B T/2$. Does this explain why we can ignore the rotational energy of helium for most purposes? (The results for other monatomic gases, and for diatomic gases rotating around the axis connecting the two atoms, have comparable orders of magnitude.)

Solution:

4.92×10^5 K; Yes, that's an impractically high temperature.

Exercise:**Problem:**

(a) How much heat must be added to raise the temperature of 1.5 mol of air from 25.0°C to 33.0°C at constant volume? Assume air is completely diatomic. (b) Repeat the problem for the same number of moles of xenon, Xe.

Exercise:**Problem:**

A sealed, rigid container of 0.560 mol of an unknown ideal gas at a temperature of 30.0°C is cooled to -40.0°C . In the process, 980 J of heat are removed from the gas. Is the gas monatomic, diatomic, or polyatomic?

Solution:

polyatomic

Exercise:**Problem:**

A sample of neon gas (Ne, molar mass $M = 20.2 \text{ g/mol}$) at a temperature of 13.0°C is put into a steel container of mass 47.2 g that's at a temperature of -40.0°C . The final temperature is -28.0°C . (No heat is exchanged with the surroundings, and you can neglect any change in the volume of the container.) What is the mass of the sample of neon?

Exercise:**Problem:**

A steel container of mass 135 g contains 24.0 g of ammonia, NH_3 , which has a molar mass of 17.0 g/mol . The container and gas are in equilibrium at 12.0°C . How much heat has to be removed to reach a temperature of -20.0°C ? Ignore the change in volume of the steel.

Solution:

$$3.08 \times 10^3 \text{ J}$$

Exercise:**Problem:**

A sealed room has a volume of 24 m^3 . It's filled with air, which may be assumed to be diatomic, at a temperature of 24°C and a pressure of $9.83 \times 10^4 \text{ Pa}$. A 1.00-kg block of ice at its melting point is placed in the room. Assume the walls of the room transfer no heat. What is the equilibrium temperature?

Exercise:

Problem:

Heliox, a mixture of helium and oxygen, is sometimes given to hospital patients who have trouble breathing, because the low mass of helium makes it easier to breathe than air. Suppose helium at $25\text{ }^{\circ}\text{C}$ is mixed with oxygen at $35\text{ }^{\circ}\text{C}$ to make a mixture that is 70% helium by mole. What is the final temperature? Ignore any heat flow to or from the surroundings, and assume the final volume is the sum of the initial volumes.

Solution:

$29.2\text{ }^{\circ}\text{C}$

Exercise:**Problem:**

Professional divers sometimes use heliox, consisting of 79% helium and 21% oxygen by mole. Suppose a perfectly rigid scuba tank with a volume of 11 L contains heliox at an absolute pressure of $2.1 \times 10^7\text{ Pa}$ at a temperature of $31\text{ }^{\circ}\text{C}$. (a) How many moles of helium and how many moles of oxygen are in the tank? (b) The diver goes down to a point where the sea temperature is $27\text{ }^{\circ}\text{C}$ while using a negligible amount of the mixture. As the gas in the tank reaches this new temperature, how much heat is removed from it?

Exercise:**Problem:**

In car racing, one advantage of mixing liquid nitrous oxide (N_2O) with air is that the boiling of the “nitrous” absorbs latent heat of vaporization and thus cools the air and ultimately the fuel-air mixture, allowing more fuel-air mixture to go into each cylinder. As a very rough look at this process, suppose 1.0 mol of nitrous oxide gas at its boiling point, $-88\text{ }^{\circ}\text{C}$, is mixed with 4.0 mol of air (assumed diatomic) at $30\text{ }^{\circ}\text{C}$. What is the final temperature of the mixture? Use the measured heat capacity of N_2O at $25\text{ }^{\circ}\text{C}$, which is $30.4\text{ J/mol }^{\circ}\text{C}$. (The primary advantage of nitrous oxide is that it consists of 1/3 oxygen, which is more than air contains, so it supplies more oxygen to burn the fuel. Another advantage is that its decomposition into nitrogen and oxygen releases energy in the cylinder.)

Solution:

$-1.6\text{ }^{\circ}\text{C}$

Glossary

degree of freedom

independent kind of motion possessing energy, such as the kinetic energy of motion in one of the three orthogonal spatial directions

equipartition theorem

theorem that the energy of a classical thermodynamic system is shared equally among its degrees of freedom

Distribution of Molecular Speeds

By the end of this section, you will be able to:

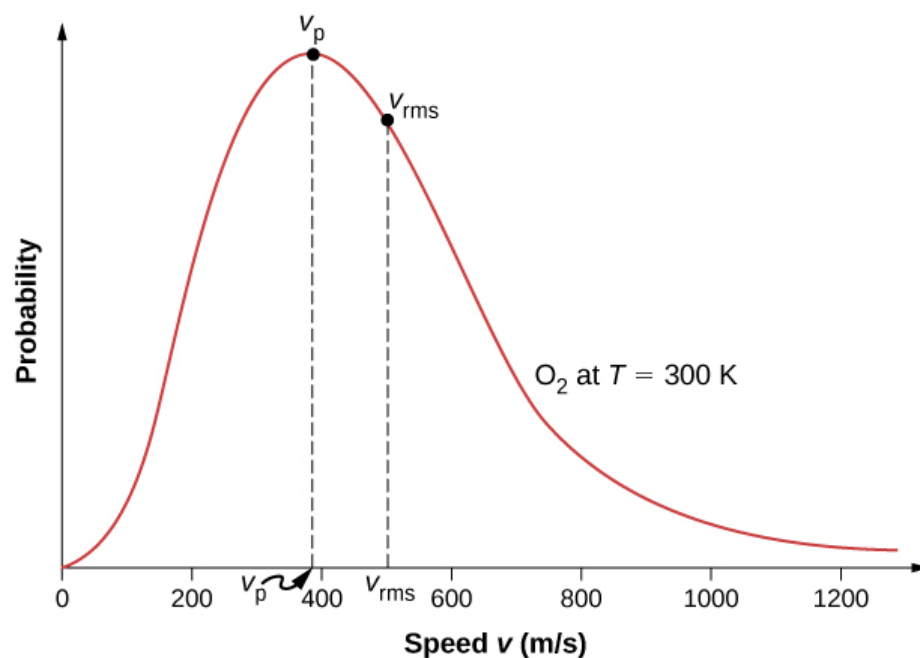
- Describe the distribution of molecular speeds in an ideal gas
- Find the average and most probable molecular speeds in an ideal gas

Particles in an ideal gas all travel at relatively high speeds, but they do not travel at the same speed. The rms speed is one kind of average, but many particles move faster and many move slower. The actual distribution of speeds has several interesting implications for other areas of physics, as we will see in later chapters.

The Maxwell-Boltzmann Distribution

The motion of molecules in a gas is random in magnitude and direction for individual molecules, but a gas of many molecules has a predictable distribution of molecular speeds. This predictable distribution of molecular speeds is known as the **Maxwell-Boltzmann distribution**, after its originators, who calculated it based on kinetic theory, and it has since been confirmed experimentally ([link](#)).

To understand this figure, we must define a distribution function of molecular speeds, since with a finite number of molecules, the probability that a molecule will have exactly a given speed is 0.



The Maxwell-Boltzmann distribution of molecular speeds in an ideal gas. The most likely speed v_p is less than the rms speed v_{rms} .

Although very high speeds are possible, only a tiny fraction of the molecules have speeds that are an order of magnitude greater than

v_{rms} .

We define the distribution function $f(v)$ by saying that the expected number $N(v_1, v_2)$ of particles with speeds between v_1 and v_2 is given by

Equation:

$$N(v_1, v_2) = N \int_{v_1}^{v_2} f(v) dv.$$

[Since N is dimensionless, the unit of $f(v)$ is seconds per meter.] We can write this equation conveniently in differential form:

Equation:

$$dN = N f(v) dv.$$

In this form, we can understand the equation as saying that the number of molecules with speeds between v and $v + dv$ is the total number of molecules in the sample times $f(v)$ times dv . That is, the probability that a molecule's speed is between v and $v + dv$ is $f(v)dv$.

We can now quote Maxwell's result, although the proof is beyond our scope.

Note:

Maxwell-Boltzmann Distribution of Speeds

The distribution function for speeds of particles in an ideal gas at temperature T is

Equation:

$$f(v) = \frac{4}{\sqrt{\pi}} \left(\frac{m}{2k_B T} \right)^{3/2} v^2 e^{(-mv^2/(2k_B T))}.$$

The factors before the v^2 are a normalization constant; they make sure that $N(0, \infty) = N$ by making sure that $\int_0^\infty f(v) dv = 1$. Let's focus on the dependence on v . The factor of v^2 means that $f(0) = 0$ and for small v , the curve looks like a parabola. The factor of $e^{-m_0 v^2 / 2k_B T}$ means that $\lim_{v \rightarrow \infty} f(v) = 0$ and the graph has an exponential tail, which indicates that a few molecules may move at several times the rms speed. The interaction of these factors gives the function the single-peaked shape shown in the figure.

Example:

Calculating the Ratio of Numbers of Molecules Near Given Speeds

In a sample of nitrogen (N_2 , with a molar mass of 28.0 g/mol) at a temperature of 27 °C, find the ratio of the number of molecules with a speed very close to 300 m/s to the number with a speed very close to 100 m/s.

Strategy

Since we're looking at a small range, we can approximate the number of molecules near 100 m/s as $dN_{100} = f(100 \text{ m/s})dv$. Then the ratio we want is

Equation:

$$\frac{dN_{300}}{dN_{100}} = \frac{f(300 \text{ m/s})dv}{f(100 \text{ m/s})dv} = \frac{f(300 \text{ m/s})}{f(100 \text{ m/s})}.$$

All we have to do is take the ratio of the two f values.

Solution

1. Identify the knowns and convert to SI units if necessary.

Equation:

$$T = 300 \text{ K}, k_B = 1.38 \times 10^{-23} \text{ J/K}$$

Equation:

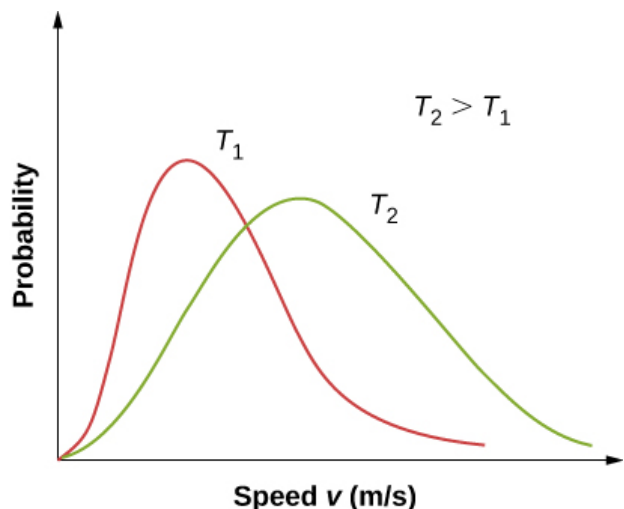
$$M = 0.0280 \text{ kg/mol so } m = 4.65 \times 10^{-26} \text{ kg}$$

2. Substitute the values and solve.

Equation:

$$\begin{aligned} \frac{f(300 \text{ m/s})}{f(100 \text{ m/s})} &= \frac{\frac{4}{\sqrt{\pi}} \left(\frac{m}{2k_B T} \right)^{3/2} (300 \text{ m/s})^2 \exp[-m(300 \text{ m/s})^2/2k_B T]}{\frac{4}{\sqrt{\pi}} \left(\frac{m}{2k_B T} \right)^{3/2} (100 \text{ m/s})^2 \exp[-m(100 \text{ m/s})^2/2k_B T]} \\ &= \frac{(300 \text{ m/s})^2 \exp[-(4.65 \times 10^{-26} \text{ kg})(300 \text{ m/s})^2/2(1.38 \times 10^{-23} \text{ J/K})(300 \text{ K})]}{(100 \text{ m/s})^2 \exp[-(4.65 \times 10^{-26} \text{ kg})(100 \text{ m/s})^2/2(1.38 \times 10^{-23} \text{ J/K})(300 \text{ K})]} \\ &= 3^2 \exp \left[-\frac{(4.65 \times 10^{-26} \text{ kg})[(300 \text{ m/s})^2 - (100 \text{ m/s})^2]}{2(1.38 \times 10^{-23} \text{ J/K})(300 \text{ K})} \right] \\ &= 5.74 \end{aligned}$$

[\[link\]](#) shows that the curve is shifted to higher speeds at higher temperatures, with a broader range of speeds.



The Maxwell-Boltzmann distribution is shifted to higher speeds and broadened at higher temperatures.

Note:

With only a relatively small number of molecules, the distribution of speeds fluctuates around the Maxwell-Boltzmann distribution. However, you can view this [simulation](#) to see the essential features that more massive molecules move slower and have a narrower distribution. Use the set-up “2 Gases, Random Speeds”. Note the display at the bottom comparing histograms of the speed distributions with the theoretical curves.

We can use a probability distribution to calculate average values by multiplying the distribution function by the quantity to be averaged and integrating the product over all possible speeds. (This is analogous to calculating averages of discrete distributions, where you multiply each value by the number of times it occurs, add the results, and divide by the number of values. The integral is analogous to the first two steps, and the normalization is analogous to dividing by the number of values.) Thus the average velocity is

Note:

Equation:

$$\bar{v} = \int_0^{\infty} v f(v) dv = \sqrt{\frac{8}{\pi} \frac{k_B T}{m}} = \sqrt{\frac{8}{\pi} \frac{RT}{M}}.$$

Similarly,

Equation:

$$v_{\text{rms}} = \sqrt{\overline{v^2}} = \sqrt{\int_0^\infty v^2 f(v) dv} = \sqrt{\frac{3k_B T}{m}} = \sqrt{\frac{3RT}{M}}$$

as in [Pressure, Temperature, and RMS Speed](#). The **most probable speed**, also called the **peak speed** v_p , is the speed at the peak of the velocity distribution. (In statistics it would be called the mode.) It is less than the rms speed v_{rms} . The most probable speed can be calculated by the more familiar method of setting the derivative of the distribution function, with respect to v , equal to 0. The result is

Note:

Equation:

$$v_p = \sqrt{\frac{2k_B T}{m}} = \sqrt{\frac{2RT}{M}},$$

which is less than v_{rms} . In fact, the rms speed is greater than both the most probable speed and the average speed.

The peak speed provides a sometimes more convenient way to write the Maxwell-Boltzmann distribution function:

Equation:

$$f(v) = \frac{4v^2}{\sqrt{\pi}v_p^3} e^{-v^2/v_p^2}$$

In the factor $e^{-mv^2/2k_B T}$, it is easy to recognize the translational kinetic energy. Thus, that expression is equal to $e^{-K/k_B T}$. The distribution $f(v)$ can be transformed into a kinetic energy distribution by requiring that $f(K)dK = f(v)dv$. Boltzmann showed that the resulting formula is much more generally applicable if we replace the kinetic energy of translation with the total mechanical energy E . Boltzmann's result is

Equation:

$$f(E) = \frac{2}{\sqrt{\pi}} (k_B T)^{-3/2} \sqrt{E} e^{-E/k_B T} = \frac{2}{\sqrt{\pi} (k_B T)^{3/2}} \frac{\sqrt{E}}{e^{E/k_B T}}.$$

The first part of this equation, with the negative exponential, is the usual way to write it. We give the second part only to remark that $e^{E/k_B T}$ in the denominator is ubiquitous in quantum as well as classical statistical mechanics.

Note:**Speed Distribution**

Step 1. Examine the situation to determine that it relates to the distribution of molecular speeds.

Step 2. Make a list of what quantities are given or can be inferred from the problem as stated (identify the known quantities).

Step 3. Identify exactly what needs to be determined in the problem (identify the unknown quantities). A written list is useful.

Step 4. Convert known values into proper SI units (K for temperature, Pa for pressure, m^3 for volume, molecules for N , and moles for n). In many cases, though, using R and the molar mass will be more convenient than using k_B and the molecular mass.

Step 5. Determine whether you need the distribution function for velocity or the one for energy, and whether you are using a formula for one of the characteristic speeds (average, most probably, or rms), finding a ratio of values of the distribution function, or approximating an integral.

Step 6. Solve the appropriate equation for the ideal gas law for the quantity to be determined (the unknown quantity). Note that if you are taking a ratio of values of the distribution function, the normalization factors divide out. Or if approximating an integral, use the method asked for in the problem.

Step 7. Substitute the known quantities, along with their units, into the appropriate equation and obtain numerical solutions complete with units.

We can now gain a qualitative understanding of a puzzle about the composition of Earth's atmosphere. Hydrogen is by far the most common element in the universe, and helium is by far the second-most common. Moreover, helium is constantly produced on Earth by radioactive decay. Why are those elements so rare in our atmosphere? The answer is that gas molecules that reach speeds above Earth's escape velocity, about 11 km/s, can escape from the atmosphere into space. Because of the lower mass of hydrogen and helium molecules, they move at higher speeds than other gas molecules, such as nitrogen and oxygen. Only a few exceed escape velocity, but far fewer heavier molecules do. Thus, over the billions of years that Earth has existed, far more hydrogen and helium molecules have escaped from the atmosphere than other molecules, and hardly any of either is now present.

We can also now take another look at evaporative cooling, which we discussed in the chapter on temperature and heat. Liquids, like gases, have a distribution of molecular energies. The highest-energy molecules are those that can escape from the intermolecular attractions of the liquid. Thus, when some liquid evaporates, the molecules left behind have a lower average energy, and the liquid has a lower temperature.

Summary

- The motion of individual molecules in a gas is random in magnitude and direction. However, a gas of many molecules has a predictable distribution of molecular speeds, known as the Maxwell-Boltzmann distribution.
- The average and most probable velocities of molecules having the Maxwell-Boltzmann speed distribution, as well as the rms velocity, can be calculated from the temperature and molecular mass.

Key Equations

Ideal gas law in terms of molecules	$pV = Nk_{\text{B}}T$
Ideal gas law ratios if the amount of gas is constant	$\frac{p_1V_1}{T_1} = \frac{p_2V_2}{T_2}$
Ideal gas law in terms of moles	$pV = nRT$
Van der Waals equation	$\left[p + a\left(\frac{n}{V}\right)^2 \right] (V - nb) = nRT$
Pressure, volume, and molecular speed	$pV = \frac{1}{3}Nm\bar{v}^2$
Root-mean-square speed	$v_{\text{rms}} = \sqrt{\frac{3RT}{M}} = \sqrt{\frac{3k_{\text{B}}T}{m}}$
Mean free path	$\lambda = \frac{V}{4\sqrt{2}\pi r^2 N} = \frac{k_{\text{B}}T}{4\sqrt{2}\pi r^2 p}$
Mean free time	$\tau = \frac{k_{\text{B}}T}{4\sqrt{2}\pi r^2 p v_{\text{rms}}}$
The following two equations apply only to a monatomic ideal gas:	
Average kinetic energy of a molecule	$\bar{K} = \frac{3}{2}k_{\text{B}}T$
Internal energy	$E_{\text{int}} = \frac{3}{2}Nk_{\text{B}}T.$
Heat in terms of molar heat capacity at constant volume	$Q = nC_V\Delta T$
Molar heat capacity at constant volume for an ideal gas with d degrees of freedom	$C_V = \frac{d}{2}R$
Maxwell–Boltzmann speed distribution	$f(v) = \frac{4}{\sqrt{\pi}}\left(\frac{m}{2k_{\text{B}}T}\right)^{3/2}v^2e^{-mv^2/2k_{\text{B}}T}$
Average velocity of a molecule	$\bar{v} = \sqrt{\frac{8}{\pi}\frac{k_{\text{B}}T}{m}} = \sqrt{\frac{8}{\pi}\frac{RT}{M}}$
Peak velocity of a molecule	$v_p = \sqrt{\frac{2k_{\text{B}}T}{m}} = \sqrt{\frac{2RT}{M}}$

Conceptual Questions

Exercise:

Problem:

One cylinder contains helium gas and another contains krypton gas at the same temperature. Mark each of these statements true, false, or impossible to determine from the given information. (a) The rms speeds of atoms in the two gases are the same. (b) The average kinetic energies of atoms in the two gases are the same. (c) The internal energies of 1 mole of gas in each cylinder are the same. (d) The pressures in the two cylinders are the same.

Solution:

a. false; b. true; c. true; d. true

Exercise:**Problem:**

Repeat the previous question if one gas is still helium but the other is changed to fluorine, F_2 .

Exercise:**Problem:**

An ideal gas is at a temperature of 300 K. To double the average speed of its molecules, what does the temperature need to be changed to?

Solution:

1200 K

Problems**Exercise:****Problem:**

In a sample of hydrogen sulfide ($M = 34.1$ g/mol) at a temperature of 3.00×10^2 K, estimate the ratio of the number of molecules that have speeds very close to v_{rms} to the number that have speeds very close to $2v_{\text{rms}}$.

Exercise:**Problem:**

Using the approximation $\int_{v_1}^{v_1+\Delta v} f(v)dv \approx f(v_1)\Delta v$ for small Δv , estimate the fraction of nitrogen molecules at a temperature of 3.00×10^2 K that have speeds between 290 m/s and 291 m/s.

Solution:

0.00157

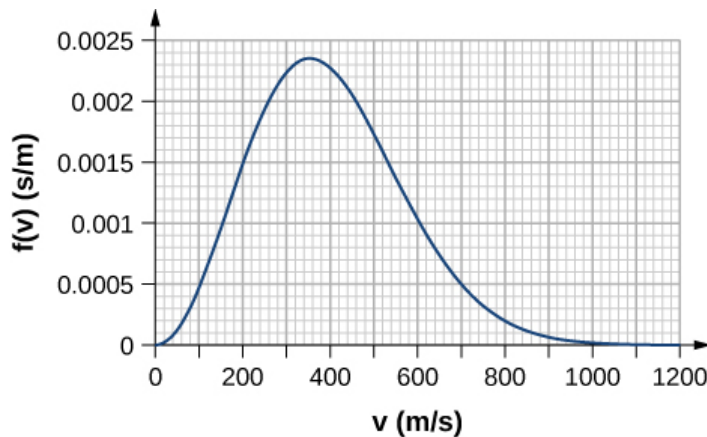
Exercise:

Problem:

Using the method of the preceding problem, estimate the fraction of nitric oxide (NO) molecules at a temperature of 250 K that have energies between 3.45×10^{-21} J and 3.50×10^{-21} J.

Exercise:**Problem:**

By counting squares in the following figure, estimate the fraction of argon atoms at $T = 300$ K that have speeds between 600 m/s and 800 m/s. The curve is correctly normalized. The value of a square is its length as measured on the x -axis times its height as measured on the y -axis, with the units given on those axes.



Solution:

About 0.072. Answers may vary slightly. A more accurate answer is 0.074.

Exercise:**Problem:**

Using a numerical integration method such as Simpson's rule, find the fraction of molecules in a sample of oxygen gas at a temperature of 250 K that have speeds between 100 m/s and 150 m/s. The molar mass of oxygen (O_2) is 32.0 g/mol. A precision to two significant digits is enough.

Exercise:**Problem:**

Find (a) the most probable speed, (b) the average speed, and (c) the rms speed for nitrogen molecules at 295 K.

Solution:

a. 419 m/s; b. 472 m/s; c. 513 m/s

Exercise:

Problem: Repeat the preceding problem for nitrogen molecules at 2950 K.

Exercise:

Problem:

At what temperature is the average speed of carbon dioxide molecules ($M = 44.0 \text{ g/mol}$) 510 m/s?

Solution:

541 K

Exercise:

Problem:

The most probable speed for molecules of a gas at 296 K is 263 m/s. What is the molar mass of the gas? (You might like to figure out what the gas is likely to be.)

Exercise:

Problem:

a) At what temperature do oxygen molecules have the same average speed as helium atoms ($M = 4.00 \text{ g/mol}$) have at 300 K? b) What is the answer to the same question about most probable speeds? c) What is the answer to the same question about rms speeds?

Solution:

2400 K for all three parts

Additional Problems

Exercise:

Problem:

In the deep space between galaxies, the density of molecules (which are mostly single atoms) can be as low as 10^6 atoms/m^3 , and the temperature is a frigid 2.7 K. What is the pressure? (b) What volume (in m^3) is occupied by 1 mol of gas? (c) If this volume is a cube, what is the length of its sides in kilometers?

Exercise:

Problem:

(a) Find the density in SI units of air at a pressure of 1.00 atm and a temperature of 20°C , assuming that air is 78% N_2 , 21% O_2 , and 1% Ar, (b) Find the density of the atmosphere on Venus, assuming that it's 96% CO_2 and 4% N_2 , with a temperature of 737 K and a pressure of 92.0 atm.

Solution:

a. 1.20 kg/m^3 ; b. 65.9 kg/m^3

Exercise:

Problem:

The air inside a hot-air balloon has a temperature of 370 K and a pressure of 101.3 kPa, the same as that of the air outside. Using the composition of air as 78% N_2 , 21% O_2 , and 1% Ar, find the density of the air inside the balloon.

Exercise:

Problem:

When an air bubble rises from the bottom to the top of a freshwater lake, its volume increases by 80 %. If the temperatures at the bottom and the top of the lake are 4.0 and 10 °C, respectively, how deep is the lake?

Solution:

7.9 m

Exercise:

Problem:

(a) Use the ideal gas equation to estimate the temperature at which 1.00 kg of steam (molar mass $M = 18.0 \text{ g/mol}$) at a pressure of $1.50 \times 10^6 \text{ Pa}$ occupies a volume of 0.220 m^3 . (b) The van der Waals constants for water are $a = 0.5537 \text{ Pa} \cdot \text{m}^6/\text{mol}^2$ and $b = 3.049 \times 10^{-5} \text{ m}^3/\text{mol}$. Use the Van der Waals equation of state to estimate the temperature under the same conditions. (c) The actual temperature is 779 K. Which estimate is better?

Exercise:

Problem:

One process for decaffeinating coffee uses carbon dioxide ($M = 44.0 \text{ g/mol}$) at a molar density of about $14,600 \text{ mol/m}^3$ and a temperature of about 60 °C. (a) Is CO_2 a solid, liquid, gas, or supercritical fluid under those conditions? (b) The van der Waals constants for carbon dioxide are $a = 0.3658 \text{ Pa} \cdot \text{m}^6/\text{mol}^2$ and $b = 4.286 \times 10^{-5} \text{ m}^3/\text{mol}$. Using the van der Waals equation, estimate the pressure of CO_2 at that temperature and density.

Solution:

a. supercritical fluid; b. $3.00 \times 10^7 \text{ Pa}$

Exercise:

Problem:

On a winter day when the air temperature is 0 °C, the relative humidity is 50 %. Outside air comes inside and is heated to a room temperature of 20 °C. What is the relative humidity of the air inside the room. (Does this problem show why inside air is so dry in winter?)

Exercise:

Problem:

On a warm day when the air temperature is $30\text{ }^{\circ}\text{C}$, a metal can is slowly cooled by adding bits of ice to liquid water in it. Condensation first appears when the can reaches $15\text{ }^{\circ}\text{C}$. What is the relative humidity of the air?

Solution:

40.18 %

Exercise:**Problem:**

(a) People often think of humid air as “heavy.” Compare the densities of air with 0% relative humidity and 100% relative humidity when both are at 1 atm and $30\text{ }^{\circ}\text{C}$. Assume that the dry air is an ideal gas composed of molecules with a molar mass of 29.0 g/mol and the moist air is the same gas mixed with water vapor. (b) As discussed in the chapter on the applications of Newton’s laws, the air resistance felt by projectiles such as baseballs and golf balls is approximately $F_D = C\rho Av^2/2$, where ρ is the mass density of the air, A is the cross-sectional area of the projectile, and C is the projectile’s drag coefficient. For a fixed air pressure, describe qualitatively how the range of a projectile changes with the relative humidity. (c) When a thunderstorm is coming, usually the humidity is high and the air pressure is low. Do those conditions give an advantage or disadvantage to home-run hitters?

Exercise:**Problem:**

The mean free path for helium at a certain temperature and pressure is $2.10 \times 10^{-7}\text{ m}$. The radius of a helium atom can be taken as $1.10 \times 10^{-11}\text{ m}$. What is the measure of the density of helium under those conditions (a) in molecules per cubic meter and (b) in moles per cubic meter?

Solution:

a. $2.21 \times 10^{27}\text{ molecules/m}^3$; b. $3.67 \times 10^3\text{ mol/m}^3$

Exercise:**Problem:**

The mean free path for methane at a temperature of 269 K and a pressure of $1.11 \times 10^5\text{ Pa}$ is $4.81 \times 10^{-8}\text{ m}$. Find the effective radius r of the methane molecule.

Exercise:

Problem:

In the chapter on fluid mechanics, Bernoulli's equation for the flow of incompressible fluids was explained in terms of changes affecting a small volume dV of fluid. Such volumes are a fundamental idea in the study of the flow of compressible fluids such as gases as well. For the equations of hydrodynamics to apply, the mean free path must be much less than the linear size of such a volume, $a \approx dV^{1/3}$. For air in the stratosphere at a temperature of 220 K and a pressure of 5.8 kPa, how big should a be for it to be 100 times the mean free path? Take the effective radius of air molecules to be 1.88×10^{-11} m, which is roughly correct for N_2 .

Solution:

8.2 mm

Exercise:**Problem:**

Find the total number of collisions between molecules in 1.00 s in 1.00 L of nitrogen gas at standard temperature and pressure (0 °C, 1.00 atm). Use 1.88×10^{-10} m as the effective radius of a nitrogen molecule. (The number of collisions per second is the reciprocal of the collision time.) Keep in mind that each collision involves two molecules, so if one molecule collides once in a certain period of time, the collision of the molecule it hit cannot be counted.

Exercise:**Problem:**

(a) Estimate the specific heat capacity of sodium from the Law of Dulong and Petit. The molar mass of sodium is 23.0 g/mol. (b) What is the percent error of your estimate from the known value, 1230 J/kg · °C?

Solution:

a. 1080 J/kg · °C; b. 12 %

Exercise:**Problem:**

A sealed, perfectly insulated container contains 0.630 mol of air at 20.0 °C and an iron stirring bar of mass 40.0 g. The stirring bar is magnetically driven to a kinetic energy of 50.0 J and allowed to slow down by air resistance. What is the equilibrium temperature?

Exercise:**Problem:**

Find the ratio $f(v_p)/f(v_{rms})$ for hydrogen gas ($M = 2.02$ g/mol) at a temperature of 77.0 K.

Solution:

$2\sqrt{e}/3$ or about 1.10

Exercise:

Problem:

Unreasonable results. (a) Find the temperature of 0.360 kg of water, modeled as an ideal gas, at a pressure of 1.01×10^5 Pa if it has a volume of 0.615 m^3 . (b) What is unreasonable about this answer? How could you get a better answer?

Exercise:**Problem:**

Unreasonable results. (a) Find the average speed of hydrogen sulfide, H_2S , molecules at a temperature of 250 K. Its molar mass is 31.4 g/mol (b) The result isn't very unreasonable, but why is it less reliable than those for, say, neon or nitrogen?

Solution:

a. 411 m/s; b. According to [\[link\]](#), the C_V of H_2S is significantly different from the theoretical value, so the ideal gas model does not describe it very well at room temperature and pressure, and the Maxwell-Boltzmann speed distribution for ideal gases may not hold very well, even less well at a lower temperature.

Challenge Problems**Exercise:****Problem:**

An airtight dispenser for drinking water is $25 \text{ cm} \times 10 \text{ cm}$ in horizontal dimensions and 20 cm tall. It has a tap of negligible volume that opens at the level of the bottom of the dispenser. Initially, it contains water to a level 3.0 cm from the top and air at the ambient pressure, 1.00 atm, from there to the top. When the tap is opened, water will flow out until the gauge pressure at the bottom of the dispenser, and thus at the opening of the tap, is 0. What volume of water flows out? Assume the temperature is constant, the dispenser is perfectly rigid, and the water has a constant density of 1000 kg/m^3 .

Exercise:**Problem:**

Eight bumper cars, each with a mass of 322 kg, are running in a room 21.0 m long and 13.0 m wide. They have no drivers, so they just bounce around on their own. The rms speed of the cars is 2.50 m/s. Repeating the arguments of [Pressure, Temperature, and RMS Speed](#), find the average force per unit length (analogous to pressure) that the cars exert on the walls.

Solution:

29.5 N/m

Exercise:

Problem: Verify that $v_p = \sqrt{\frac{2k_B T}{m}}$.

Exercise:**Problem:**

Verify the normalization equation $\int_0^\infty f(v)dv = 1$. In doing the integral, first make the substitution $u = \sqrt{\frac{m}{2k_B T}} v = \frac{v}{v_p}$. This “scaling” transformation gives you all features of the answer except for the integral, which is a dimensionless numerical factor. You’ll need the formula

$$\int_0^\infty x^2 e^{-x^2} dx = \frac{\sqrt{\pi}}{4}$$

to find the numerical factor and verify the normalization.

Solution:

Substituting $v = \sqrt{\frac{2k_B T}{m}} u$ and $dv = \sqrt{\frac{2k_B T}{m}} du$ gives

$$\begin{aligned} \int_0^\infty \frac{4}{\sqrt{\pi}} \left(\frac{m}{2k_B T} \right)^{3/2} v^2 e^{-mv^2/2k_B T} dv &= \int_0^\infty \frac{4}{\sqrt{\pi}} \left(\frac{m}{2k_B T} \right)^{3/2} \left(\frac{2k_B T}{m} \right) u^2 e^{-u^2} \sqrt{\frac{2k_B T}{m}} du \\ &= \int_0^\infty \frac{4}{\sqrt{\pi}} u^2 e^{-u^2} du = \frac{4}{\sqrt{\pi}} \frac{\sqrt{\pi}}{4} = 1 \end{aligned}$$

Exercise:**Problem:**

Verify that $\bar{v} = \sqrt{\frac{8}{\pi} \frac{k_B T}{m}}$. Make the same scaling transformation as in the preceding problem.

Exercise:

Problem: Verify that $v_{\text{rms}} = \sqrt{\bar{v}^2} = \sqrt{\frac{3k_B T}{m}}$.

Solution:

Making the scaling transformation as in the previous problems, we find that

$$\bar{v}^2 = \int_0^\infty \frac{4}{\sqrt{\pi}} \left(\frac{m}{2k_B T} \right)^{3/2} v^2 v^2 e^{-mv^2/2k_B T} dv = \int_0^\infty \frac{4}{\sqrt{\pi}} \frac{2k_B T}{m} u^4 e^{-u^2} du.$$

As in the previous problem, we integrate by parts:

$$\int_0^\infty u^4 e^{-u^2} du = \left[-\frac{1}{2} u^3 e^{-u^2} \right]_0^\infty + \frac{3}{2} \int_0^\infty u^2 e^{-u^2} du.$$

Again, the first term is 0, and we were given in an earlier problem that the integral in the second term equals $\frac{\sqrt{\pi}}{4}$. We now have

$$\bar{v}^2 = \frac{4}{\sqrt{\pi}} \frac{2k_B T}{m} \frac{3}{2} \frac{\sqrt{\pi}}{4} = \frac{3k_B T}{m}.$$

Taking the square root of both sides gives the desired result: $v_{\text{rms}} = \sqrt{\frac{3k_B T}{m}}$.

Glossary

Maxwell-Boltzmann distribution

function that can be integrated to give the probability of finding ideal gas molecules with speeds in the range between the limits of integration

most probable speed

speed near which the speeds of most molecules are found, the peak of the speed distribution function

peak speed

same as “most probable speed”

Introduction

class="introduction"

A weak cold front of air pushes all the smog in northeastern China into a giant smog blanket over the Yellow Sea, as captured by NASA's Terra satellite in 2012. To understand changes in weather and climate, such as the event shown here, you need a thorough knowledge of thermodynamics . (credit: modification of work by NASA)



Heat is the transfer of energy due to a temperature difference between two systems. Heat describes the process of converting from one form of energy into another. A car engine, for example, burns gasoline. Heat is produced when the burned fuel is chemically transformed into mostly CO_2 and H_2O , which are gases at the combustion temperature. These gases exert a force on a piston through a displacement, doing work and converting the piston's kinetic energy into a variety of other forms—into the car's kinetic energy; into electrical energy to run the spark plugs, radio, and lights; and back into stored energy in the car's battery.

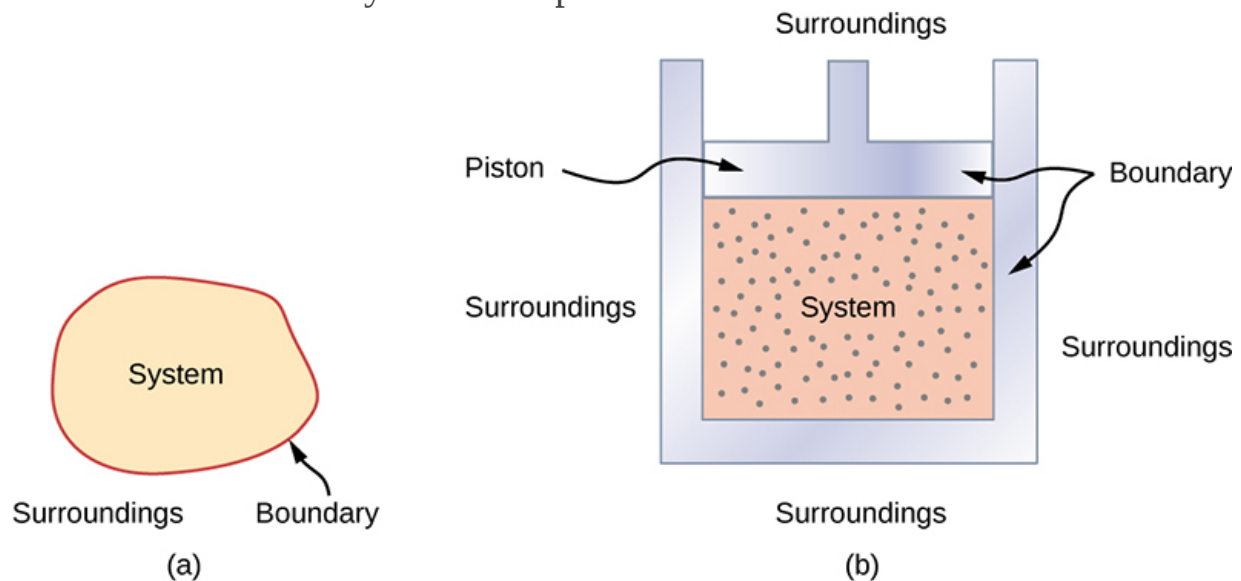
Energy is conserved in all processes, including those associated with thermodynamic systems. The roles of heat transfer and internal energy change vary from process to process and affect how work is done by the system in that process. We will see that the first law of thermodynamics explains that a change in the internal energy of a system comes from changes in heat or work. Understanding the laws that govern thermodynamic processes and the relationship between the system and its surroundings is therefore paramount in gaining scientific knowledge of energy and energy consumption.

Thermodynamic Systems

By the end of this section, you will be able to:

- Define a thermodynamic system, its boundary, and its surroundings
- Explain the roles of all the components involved in thermodynamics
- Define thermal equilibrium and thermodynamic temperature
- Link an equation of state to a system

A **thermodynamic system** includes anything whose thermodynamic properties are of interest. It is embedded in its **surroundings** or **environment**; it can exchange heat with, and do work on, its environment through a **boundary**, which is the imagined wall that separates the system and the environment ([\[link\]](#)). In reality, the immediate surroundings of the system are interacting with it directly and therefore have a much stronger influence on its behavior and properties. For example, if we are studying a car engine, the burning gasoline inside the cylinder of the engine is the thermodynamic system; the piston, exhaust system, radiator, and air outside form the surroundings of the system. The boundary then consists of the inner surfaces of the cylinder and piston.



(a) A system, which can include any relevant process or value, is self-contained in an area. The surroundings may also have relevant information; however, the surroundings are important to study only if

the situation is an open system. (b) The burning gasoline in the cylinder of a car engine is an example of a thermodynamic system.

Normally, a system must have some interactions with its surroundings. A system is called an isolated and **closed system** if it is completely separated from its environment—for example, a gas that is surrounded by immovable and thermally insulating walls. In reality, a closed system does not exist unless the entire universe is treated as the system, or it is used as a model for an actual system that has minimal interactions with its environment. Most systems are known as an **open system**, which can exchange energy and/or matter with its surroundings ([\[link\]](#)).



(a)



(b)

(a) This boiling tea kettle is an open thermodynamic system. It transfers heat and matter (steam) to its surroundings. (b) A pressure cooker is a good approximation to a closed system. A little steam escapes through the top valve to prevent explosion. (credit a: modification of work by Gina Hamilton; credit b: modification of work by Jane Whitney)

When we examine a thermodynamic system, we ignore the difference in behavior from place to place inside the system for a given moment. In other words, we concentrate on the macroscopic properties of the system, which are the averages of the microscopic properties of all the molecules or entities in the system. Any thermodynamic system is therefore treated as a continuum that has the same behavior everywhere inside. We assume the system is in **equilibrium**. You could have, for example, a temperature gradient across the system. However, when we discuss a thermodynamic system in this chapter, we study those that have uniform properties throughout the system.

Before we can carry out any study on a thermodynamic system, we need a fundamental characterization of the system. When we studied a mechanical system, we focused on the forces and torques on the system, and their balances dictated the mechanical equilibrium of the system. In a similar way, we should examine the heat transfer between a thermodynamic system and its environment or between the different parts of the system, and its balance should dictate the thermal equilibrium of the system. Intuitively, such a balance is reached if the temperature becomes the same for different objects or parts of the system in thermal contact, and the net heat transfer over time becomes zero.

Thus, when we say two objects (a thermodynamic system and its environment, for example) are in thermal equilibrium, we mean that they are at the same temperature, as we discussed in [Temperature and Heat](#). Let us consider three objects at temperatures T_1 , T_2 , and T_3 , respectively. How do we know whether they are in thermal equilibrium? The governing principle here is the zeroth law of thermodynamics, as described in [Temperature and Heat](#) on temperature and heat:

If object 1 is in thermal equilibrium with objects 2 and 3, respectively, then objects 2 and 3 must also be in thermal equilibrium.

Mathematically, we can simply write the zeroth law of thermodynamics as **Equation:**

$$\text{If } T_1 = T_2 \text{ and } T_1 = T_3, \text{ then } T_2 = T_3.$$

This is the most fundamental way of defining temperature: Two objects must be at the same temperature thermodynamically if the net heat transfer between them is zero when they are put in thermal contact and have reached a thermal equilibrium.

The zeroth law of thermodynamics is equally applicable to the different parts of a closed system and requires that the temperature everywhere inside the system be the same if the system has reached a thermal equilibrium. To simplify our discussion, we assume the system is uniform with only one type of material—for example, water in a tank. The measurable properties of the system at least include its volume, pressure, and temperature. The range of specific relevant variables depends upon the system. For example, for a stretched rubber band, the relevant variables would be length, tension, and temperature. The relationship between these three basic properties of the system is called the **equation of state** of the system and is written symbolically *for a closed system* as

Note:

Equation:

$$f(p, V, T) = 0,$$

where V , p , and T are the volume, pressure, and temperature of the system at a given condition.

In principle, this equation of state exists for any thermodynamic system but is not always readily available. The forms of $f(p, V, T) = 0$ for many materials have been determined either experimentally or theoretically. In the preceding chapter, we saw an example of an equation of state for an ideal gas, $f(p, V, T) = pV - nRT = 0$.

We have so far introduced several physical properties that are relevant to the thermodynamics of a thermodynamic system, such as its volume,

pressure, and temperature. We can separate these quantities into two generic categories. The quantity associated with an amount of matter is an **extensive variable**, such as the volume and the number of moles. The other properties of a system are **intensive variables**, such as the pressure and temperature. An extensive variable doubles its value if the amount of matter in the system doubles, provided all the intensive variables remain the same. For example, the volume or total energy of the system doubles if we double the amount of matter in the system while holding the temperature and pressure of the system unchanged.

Summary

- A thermodynamic system, its boundary, and its surroundings must be defined with all the roles of the components fully explained before we can analyze a situation.
- Thermal equilibrium is reached with two objects if a third object is in thermal equilibrium with the other two separately.
- A general equation of state for a closed system has the form $f(p, V, T) = 0$, with an ideal gas as an illustrative example.

Conceptual Questions

Exercise:

Problem:

Consider these scenarios and state whether work is done by the system on the environment (SE) or by the environment on the system (ES): (a) opening a carbonated beverage; (b) filling a flat tire; (c) a sealed empty gas can expands on a hot day, bowing out the walls.

Solution:

a. SE; b. ES; c. ES

Problems

Exercise:**Problem:**

A gas follows $pV = bp + c_T$ on an isothermal curve, where p is the pressure, V is the volume, b is a constant, and c is a function of temperature. Show that a temperature scale under an isochoric process can be established with this gas and is identical to that of an ideal gas.

Solution:

$p(V - b) = -c_T$ is the temperature scale desired and mirrors the ideal gas if under constant volume.

Exercise:**Problem:**

A mole of gas has isobaric expansion coefficient $dV/dT = R/p$ and isochoric pressure-temperature coefficient $dp/dT = p/T$. Find the equation of state of the gas.

Exercise:**Problem:**

Find the equation of state of a solid that has an isobaric expansion coefficient $dV/dT = 2cT - bp$ and an isothermal pressure-volume coefficient $dV/dp = -bT$.

Solution:

$$V - bpT + cT^2 = 0$$

Glossary

boundary

imagined walls that separate the system and its surroundings

closed system

system that is mechanically and thermally isolated from its environment

environment
outside of the system being studied

equation of state
describes properties of matter under given physical conditions

equilibrium
thermal balance established between two objects or parts within a system

extensive variable
variable that is proportional to the amount of matter in the system

intensive variable
variable that is independent of the amount of matter in the system

open system
system that can exchange energy and/or matter with its surroundings

surroundings
environment that interacts with an open system

thermodynamic system
object and focus of thermodynamic study

Work, Heat, and Internal Energy

By the end of this section, you will be able to:

- Describe the work done by a system, heat transfer between objects, and internal energy change of a system
- Calculate the work, heat transfer, and internal energy change in a simple process

We discussed the concepts of work and energy earlier in mechanics. Examples and related issues of heat transfer between different objects have also been discussed in the preceding chapters. Here, we want to expand these concepts to a thermodynamic system and its environment. Specifically, we elaborated on the concepts of heat and heat transfer in the previous two chapters. Here, we want to understand how work is done by or to a thermodynamic system; how heat is transferred between a system and its environment; and how the total energy of the system changes under the influence of the work done and heat transfer.

Work Done by a System

A force created from any source can do work by moving an object through a displacement. Then how does a thermodynamic system do work? [\[link\]](#) shows a gas confined to a cylinder that has a movable piston at one end. If the gas expands against the piston, it exerts a force through a distance and does work on the piston. If the piston compresses the gas as it is moved inward, work is also done—in this case, on the gas. The work associated with such volume changes can be determined as follows: Let the gas pressure on the piston face be p . Then the force on the piston due to the gas is pA , where A is the area of the face. When the piston is pushed outward an infinitesimal distance dx , the magnitude of the work done by the gas is

Equation:

$$dW = F dx = pA dx.$$

Since the change in volume of the gas is $dV = A dx$, this becomes

Equation:

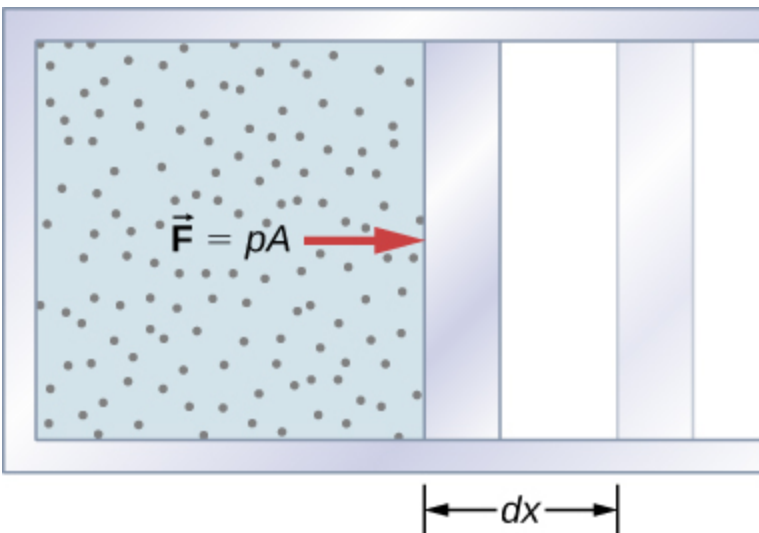
$$dW = pdV.$$

For a finite change in volume from V_1 to V_2 , we can integrate this equation from V_1 to V_2 to find the net work:

Note:

Equation:

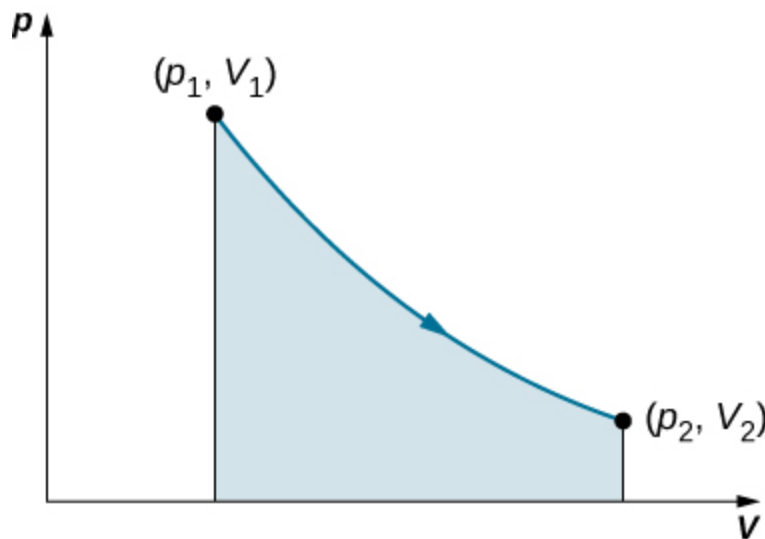
$$W = \int_{V_1}^{V_2} pdV.$$



The work done by a confined gas in moving a piston a distance dx is given by
 $dW = Fdx = pdV.$

This integral is only meaningful for a **quasi-static process**, which means a process that takes place in infinitesimally small steps, keeping the system at

thermal equilibrium. (We examine this idea in more detail later in this chapter.) Only then does a well-defined mathematical relationship (the equation of state) exist between the pressure and volume. This relationship can be plotted on a pV diagram of pressure versus volume, where the curve is the change of state. We can approximate such a process as one that occurs slowly, through a series of equilibrium states. The integral is interpreted graphically as the area under the pV curve (the shaded area of [link](#)). Work done by the gas is positive for expansion and negative for compression.

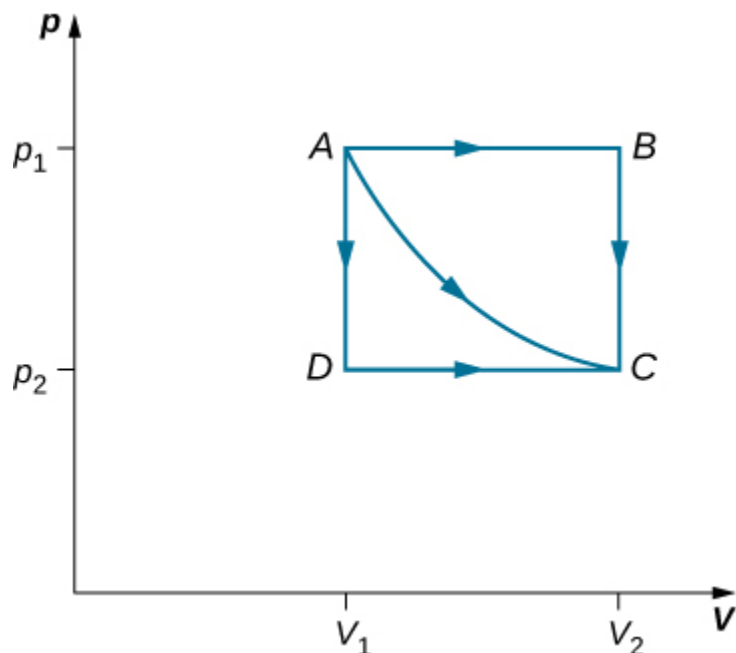


When a gas expands slowly from V_1 to V_2 , the work done by the system is represented by the shaded area under the pV curve.

Consider the two processes involving an ideal gas that are represented by paths AC and ABC in [link](#). The first process is an isothermal expansion, with the volume of the gas changing its volume from V_1 to V_2 . This isothermal process is represented by the curve between points A and C. The gas is kept at a constant temperature T by keeping it in thermal equilibrium with a heat reservoir at that temperature. From [link](#) and the ideal gas law,

Equation:

$$W = \int_{V_1}^{V_2} p dV = \int_{V_1}^{V_2} \left(\frac{nRT}{V} \right) dV.$$



The paths ABC , AC , and ADC represent three different quasi-static transitions between the equilibrium states A and C .

The expansion is isothermal, so T remains constant over the entire process. Since n and R are also constant, the only variable in the integrand is V , so the work done by an ideal gas in an isothermal process is

Equation:

$$W = nRT \int_{V_1}^{V_2} \frac{dV}{V} = nRT \ln \frac{V_2}{V_1}.$$

Notice that if $V_2 > V_1$ (expansion), W is positive, as expected.

The straight lines from A to B and then from B to C represent a different process. Here, a gas at a pressure p_1 first expands isobarically (constant pressure) and quasi-statically from V_1 to V_2 , after which it cools quasi-statically at the constant volume V_2 until its pressure drops to p_2 . From A to B , the pressure is constant at p , so the work over this part of the path is

Equation:

$$W = \int_{V_1}^{V_2} p dV = p_1 \int_{V_1}^{V_2} dV = p_1(V_2 - V_1).$$

From B to C , there is no change in volume and therefore no work is done. The net work over the path ABC is then

Equation:

$$W = p_1(V_2 - V_1) + 0 = p_1(V_2 - V_1).$$

A comparison of the expressions for the work done by the gas in the two processes of [\[link\]](#) shows that they are quite different. This illustrates a very important property of thermodynamic work: It is *path dependent*. We cannot determine the work done by a system as it goes from one equilibrium state to another unless we know its thermodynamic path. Different values of the work are associated with different paths.

Example:

Isothermal Expansion of a van der Waals Gas

Studies of a van der Waals gas require an adjustment to the ideal gas law that takes into consideration that gas molecules have a definite volume (see [The Kinetic Theory of Gases](#)). One mole of a van der Waals gas has an equation of state

Equation:

$$\left(p + \frac{a}{V^2}\right)(V - b) = RT,$$

where a and b are two parameters for a specific gas. Suppose the gas expands isothermally and quasi-statically from volume V_1 to volume V_2 . How much work is done by the gas during the expansion?

Strategy

Because the equation of state is given, we can use [\[link\]](#) to express the pressure in terms of V and T . Furthermore, temperature T is a constant under the isothermal condition, so V becomes the only changing variable under the integral.

Solution

To evaluate this integral, we must express p as a function of V . From the given equation of state, the gas pressure is

Equation:

$$p = \frac{RT}{V - b} - \frac{a}{V^2}.$$

Because T is constant under the isothermal condition, the work done by 1 mol of a van der Waals gas in expanding from a volume V_1 to a volume V_2 is thus

Equation:

$$\begin{aligned} W &= \int_{V_1}^{V_2} \left(\frac{RT}{V - b} - \frac{a}{V^2} \right) dV = RT \ln(V - b) + \frac{a}{V} \Big|_{V_1}^{V_2} \\ &= RT \ln \left(\frac{V_2 - b}{V_1 - b} \right) + a \left(\frac{1}{V_2} - \frac{1}{V_1} \right). \end{aligned}$$

Significance

By taking into account the volume of molecules, the expression for work is much more complex. If, however, we set $a = 0$ and $b = 0$, we see that the expression for work matches exactly the work done by an isothermal process for one mole of an ideal gas.

Note:

Exercise:

Problem:

Check Your Understanding How much work is done by the gas, as given in [\[link\]](#), when it expands quasi-statically along the path *ADC*?

Solution:

$$p_2(V_2 - V_1)$$

Internal Energy

The **internal energy** E_{int} of a thermodynamic system is, by definition, the sum of the mechanical energies of all the molecules or entities in the system. If the kinetic and potential energies of molecule i are K_i and U_i , respectively, then the internal energy of the system is the average of the total mechanical energy of all the entities:

Note:

Equation:

$$E_{\text{int}} = \sum_i (\bar{K}_i + \bar{U}_i),$$

where the summation is over all the molecules of the system, and the bars over K and U indicate average values. The kinetic energy K_i of an individual molecule includes contributions due to its rotation and vibration, as well as its translational energy $m_i v_i^2/2$, where v_i is the molecule's speed measured relative to the center of mass of the system. The potential energy

U_i is associated only with the interactions between molecule i and the other molecules of the system. In fact, neither the system's location nor its motion is of any consequence as far as the internal energy is concerned. The internal energy of the system is not affected by moving it from the basement to the roof of a 100-story building or by placing it on a moving train.

In an ideal monatomic gas, each molecule is a single atom. Consequently, there is no rotational or vibrational kinetic energy and $K_i = m_i v_i^2 / 2$. Furthermore, there are no interatomic interactions (collisions notwithstanding), so $U_i = \text{constant}$, which we set to zero. The internal energy is therefore due to translational kinetic energy only and

Equation:

$$E_{\text{int}} = \sum_i \bar{K}_i = \sum_i \frac{1}{2} m_i v_i^2.$$

From the discussion in the preceding chapter, we know that the average kinetic energy of a molecule in an ideal monatomic gas is

Equation:

$$\frac{1}{2} m_i \bar{v}_i^2 = \frac{3}{2} k_B T,$$

where T is the Kelvin temperature of the gas. Consequently, the average mechanical energy per molecule of an ideal monatomic gas is also $3k_B T / 2$, that is,

Equation:

$$K_i + U_i = \bar{K}_i = \frac{3}{2} k_B T.$$

The internal energy is just the number of molecules multiplied by the average mechanical energy per molecule. Thus for n moles of an ideal monatomic gas,

Note:

Equation:

$$E_{\text{int}} = nN_A \left(\frac{3}{2} k_B T \right) = \frac{3}{2} nRT.$$

Notice that the internal energy of a given quantity of an ideal monatomic gas depends on just the temperature and is completely independent of the pressure and volume of the gas. For other systems, the internal energy cannot be expressed so simply. However, an increase in internal energy can often be associated with an increase in temperature.

We know from the zeroth law of thermodynamics that when two systems are placed in thermal contact, they eventually reach thermal equilibrium, at which point they are at the same temperature. As an example, suppose we mix two monatomic ideal gases. Now, the energy per molecule of an ideal monatomic gas is proportional to its temperature. Thus, when the two gases are mixed, the molecules of the hotter gas must lose energy and the molecules of the colder gas must gain energy. This continues until thermal equilibrium is reached, at which point, the temperature, and therefore the average translational kinetic energy per molecule, is the same for both gases. The approach to equilibrium for real systems is somewhat more complicated than for an ideal monatomic gas. Nevertheless, we can still say that energy is exchanged between the systems until their temperatures are the same.

Summary

- Positive (negative) work is done by a thermodynamic system when it expands (contracts) under an external pressure.
- Heat is the energy transferred between two objects (or two parts of a system) because of a temperature difference.
- Internal energy of a thermodynamic system is its total mechanical energy.

Conceptual Questions

Exercise:

Problem:

Is it possible to determine whether a change in internal energy is caused by heat transferred, by work performed, or by a combination of the two?

Exercise:

Problem:

When a liquid is vaporized, its change in internal energy is not equal to the heat added. Why?

Solution:

Some of the energy goes into changing the phase of the liquid to gas.

Exercise:

Problem:

Why does a bicycle pump feel warm as you inflate your tire?

Exercise:

Problem:

Is it possible for the temperature of a system to remain constant when heat flows into or out of it? If so, give examples.

Solution:

Yes, as long as the work done equals the heat added there will be no change in internal energy and thereby no change in temperature. When water freezes or when ice melts while removing or adding heat, respectively, the temperature remains constant.

Problems

Exercise:

Problem:

A gas at a pressure of 2.00 atm undergoes a quasi-static isobaric expansion from 3.00 to 5.00 L. How much work is done by the gas?

Exercise:

Problem:

It takes 500 J of work to compress quasi-statically 0.50 mol of an ideal gas to one-fifth its original volume. Calculate the temperature of the gas, assuming it remains constant during the compression.

Solution:

74 K

Exercise:

Problem:

It is found that, when a dilute gas expands quasi-statically from 0.50 to 4.0 L, it does 250 J of work. Assuming that the gas temperature remains constant at 300 K, how many moles of gas are present?

Exercise:

Problem:

In a quasi-static isobaric expansion, 500 J of work are done by the gas. If the gas pressure is 0.80 atm, what is the fractional increase in the volume of the gas, assuming it was originally at 20.0 L?

Solution:

0.31

Exercise:

Problem:

When a gas undergoes a quasi-static isobaric change in volume from 10.0 to 2.0 L, 15 J of work from an external source are required. What is the pressure of the gas?

Exercise:**Problem:**

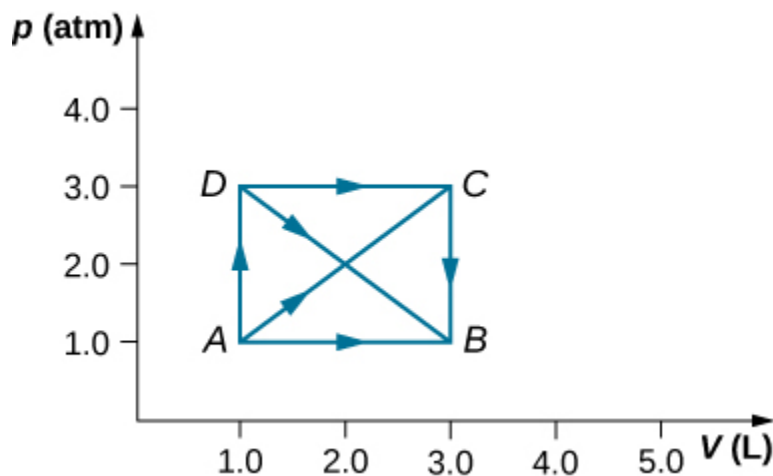
An ideal gas expands quasi-statically and isothermally from a state with pressure p and volume V to a state with volume $4V$. Show that the work done by the gas in the expansion is $pV(\ln 4)$.

Solution:

$$pV\ln(4)$$

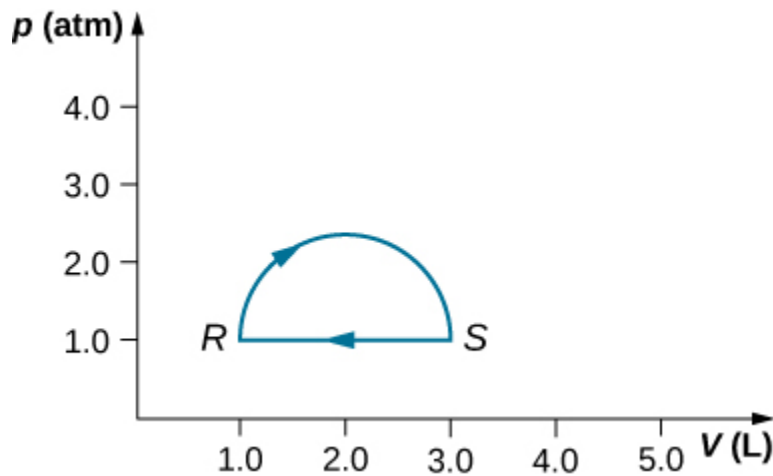
Exercise:**Problem:**

As shown below, calculate the work done by the gas in the quasi-static processes represented by the paths (a) AB; (b) ADB; (c) ACB; and (d) ADCB.

**Exercise:**

Problem:

(a) Calculate the work done on the gas along the closed path shown below. The curved section between R and S is semicircular. (b) If the process is carried out in the opposite direction, what is the work done on the gas?



Solution:

a. 160 J; b. -160 J

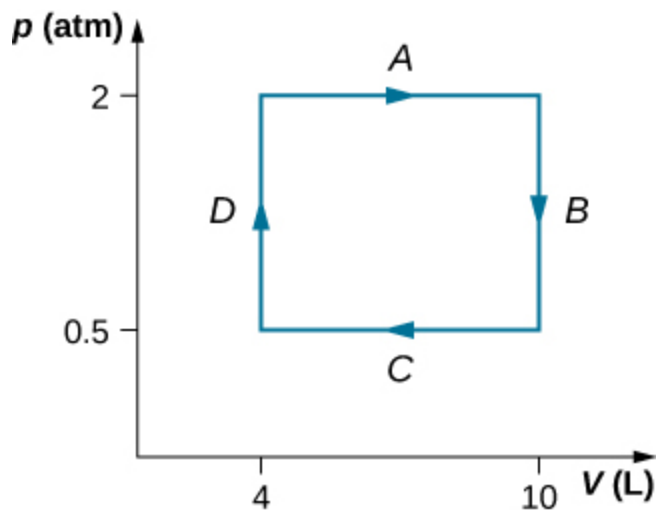
Exercise:**Problem:**

An ideal gas expands quasi-statically to three times its original volume. Which process requires more work from the gas, an isothermal process or an isobaric one? Determine the ratio of the work done in these processes.

Exercise:

Problem:

A dilute gas at a pressure of 2.0 atm and a volume of 4.0 L is taken through the following quasi-static steps: (a) an isobaric expansion to a volume of 10.0 L, (b) an isochoric change to a pressure of 0.50 atm, (c) an isobaric compression to a volume of 4.0 L, and (d) an isochoric change to a pressure of 2.0 atm. Show these steps on a pV diagram and determine from your graph the net work done by the gas.

Solution:

$$W = 900 \text{ J}$$

Exercise:**Problem:**

What is the average mechanical energy of the atoms of an ideal monatomic gas at 300 K?

Exercise:

Problem:

What is the internal energy of 6.00 mol of an ideal monatomic gas at 200 °C ?

Solution:

$$3.53 \times 10^4 \text{ J}$$

Exercise:**Problem:**

Calculate the internal energy of 15 mg of helium at a temperature of 0 °C.

Exercise:**Problem:**

Two monatomic ideal gases A and B are at the same temperature. If 1.0 g of gas A has the same internal energy as 0.10 g of gas B, what are (a) the ratio of the number of moles of each gas and (b) the ratio of the atomic masses of the two gases?

Solution:

a. 1:1; b. 10:1

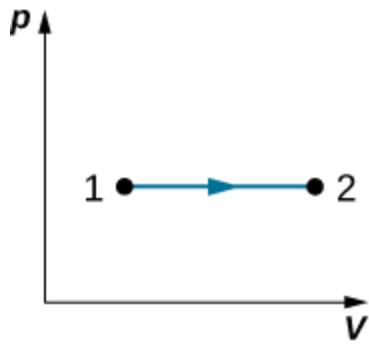
Exercise:**Problem:**

The van der Waals coefficients for oxygen are $a = 0.138 \text{ J} \cdot \text{m}^3/\text{mol}^2$ and $b = 3.18 \times 10^{-5} \text{ m}^3/\text{mol}$. Use these values to draw a van der Waals isotherm of oxygen at 100 K. On the same graph, draw isotherms of one mole of an ideal gas.

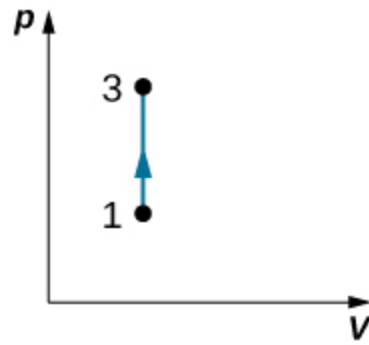
Exercise:

Problem:

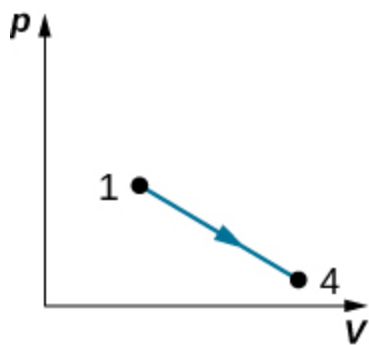
Find the work done in the quasi-static processes shown below. The states are given as (p, V) values for the points in the pV plane: 1 (3 atm, 4 L), 2 (3 atm, 6 L), 3 (5 atm, 4 L), 4 (2 atm, 6 L), 5 (4 atm, 2 L), 6 (5 atm, 5 L), and 7 (2 atm, 5 L).



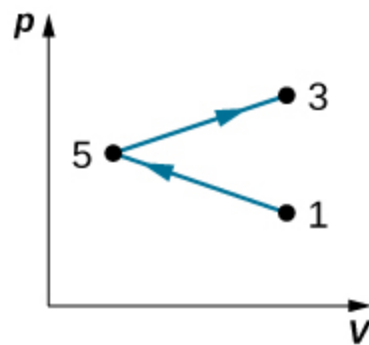
(a)



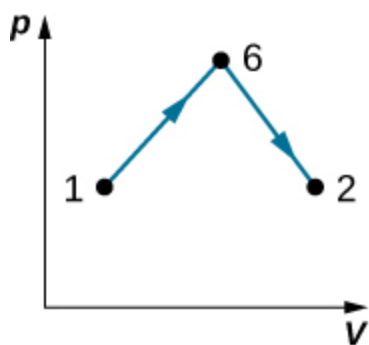
(b)



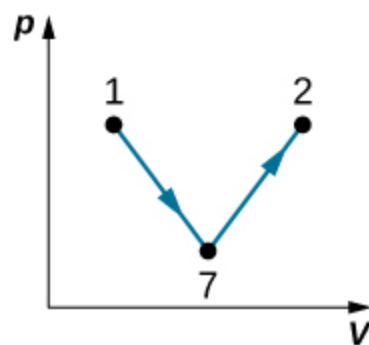
(c)



(d)



(e)



(f)

Solution:

a. 600 J; b. 0; c. 500 J; d. 200 J; e. 800 J; f. 500 J

Glossary

internal energy

average of the total mechanical energy of all the molecules or entities in the system

quasi-static process

evolution of a system that goes so slowly that the system involved is always in thermodynamic equilibrium

First Law of Thermodynamics

By the end of this section, you will be able to:

- State the first law of thermodynamics and explain how it is applied
- Explain how heat transfer, work done, and internal energy change are related in any thermodynamic process

Now that we have seen how to calculate internal energy, heat, and work done for a thermodynamic system undergoing change during some process, we can see how these quantities interact to affect the amount of change that can occur. This interaction is given by the first law of thermodynamics. British scientist and novelist C. P. Snow (1905–1980) is credited with a joke about the four laws of thermodynamics. His humorous statement of the first law of thermodynamics is stated “you can’t win,” or in other words, you cannot get more energy out of a system than you put into it. We will see in this chapter how internal energy, heat, and work all play a role in the first law of thermodynamics.

Suppose Q represents the heat exchanged between a system and the environment, and W is the work done by or on the system. The first law states that the change in internal energy of that system is given by $Q - W$. Since added heat increases the internal energy of a system, Q is positive when it is added to the system and negative when it is removed from the system.

When a gas expands, it does work and its internal energy decreases. Thus, W is positive when work is done by the system and negative when work is done on the system. This sign convention is summarized in [\[link\]](#). The **first law of thermodynamics** is stated as follows:

Note:

First Law of Thermodynamics

Associated with every equilibrium state of a system is its internal energy E_{int} . The change in E_{int} for any transition between two equilibrium states is

Equation:

$$\Delta E_{\text{int}} = Q - W$$

where Q and W represent, respectively, the heat exchanged by the system and the work done by or on the system.

Thermodynamic Sign Conventions for Heat and Work	
Process	Convention
Heat added to system	$Q > 0$
Heat removed from system	$Q < 0$
Work done by system	$W > 0$
Work done on system	$W < 0$

The first law is a statement of energy conservation. It tells us that a system can exchange energy with its surroundings by the transmission of heat and by the performance of work. The net energy exchanged is then equal to the change in the total mechanical energy of the molecules of the system (i.e., the system's internal energy). Thus, if a system is isolated, its internal energy must remain constant.

Although Q and W both depend on the thermodynamic path taken between two equilibrium states, their difference $Q - W$ does not. [\[link\]](#) shows the pV diagram of a system that is making the transition from A to B repeatedly along different thermodynamic paths. Along path 1, the system absorbs heat Q_1 and does work W_1 ; along path 2, it absorbs heat Q_2 and does work W_2 , and so on. The values of Q_i and W_i may vary from path to path, but we have **Equation:**

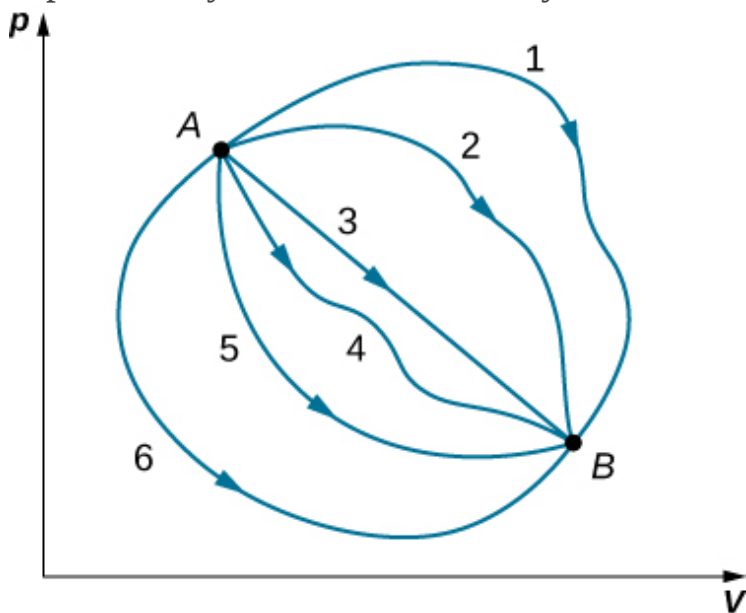
$$Q_1 - W_1 = Q_2 - W_2 = \cdots = Q_i - W_i = \cdots,$$

or

Equation:

$$\Delta E_{\text{int}1} = \Delta E_{\text{int}2} = \cdots = \Delta E_{\text{int}i} = \cdots.$$

That is, the change in the internal energy of the system between A and B is path independent. In the chapter on potential energy and the conservation of energy, we encountered another path-independent quantity: the change in potential energy between two arbitrary points in space. This change represents the negative of the work done by a conservative force between the two points. The potential energy is a function of spatial coordinates, whereas the internal energy is a function of thermodynamic variables. For example, we might write $E_{\text{int}}(T, p)$ for the internal energy. Functions such as internal energy and potential energy are known as *state functions* because their values depend solely on the state of the system.



Different thermodynamic paths taken by a system in going from state A to state B .

For all transitions, the change in the internal energy of the system

$$\Delta E_{\text{int}} = Q - W \text{ is the same.}$$

Often the first law is used in its differential form, which is

Equation:

$$dE_{\text{int}} = dQ - dW.$$

Here dE_{int} is an infinitesimal change in internal energy when an infinitesimal amount of heat dQ is exchanged with the system and an infinitesimal amount of work dW is done by (positive in sign) or on (negative in sign) the system.

Example:

Changes of State and the First Law

During a thermodynamic process, a system moves from state A to state B , it is supplied with 400 J of heat and does 100 J of work. (a) For this transition, what is the system's change in internal energy? (b) If the system then moves from state B back to state A , what is its change in internal energy? (c) If in moving from A to B along a different path, $W'_{AB} = 400$ J of work is done on the system, how much heat does it absorb?

Strategy

The first law of thermodynamics relates the internal energy change, work done by the system, and the heat transferred to the system in a simple equation. The internal energy is a function of state and is therefore fixed at any given point regardless of how the system reaches the state.

Solution

- a. From the first law, the change in the system's internal energy is

Equation:

$$\Delta E_{\text{int}AB} = Q_{AB} - W_{AB} = 400 \text{ J} - 100 \text{ J} = 300 \text{ J}.$$

- b. Consider a closed path that passes through the states A and B . Internal energy is a state function, so ΔE_{int} is zero for a closed path. Thus

Equation:

$$\Delta E_{\text{int}} = \Delta E_{\text{int}AB} + \Delta E_{\text{int}BA} = 0,$$

and

Equation:

$$\Delta E_{\text{int}AB} = -\Delta E_{\text{int}BA}.$$

This yields

Equation:

$$\Delta E_{\text{int}BA} = -300 \text{ J}.$$

c. The change in internal energy is the same for any path, so

Equation:

$$\begin{aligned}\Delta E_{\text{int}AB} &= \Delta E'_{\text{int}AB} = Q'_{AB} - W'_{AB}; \\ 300 \text{ J} &= Q'_{AB} - (-400 \text{ J}),\end{aligned}$$

and the heat exchanged is

Equation:

$$Q'_{AB} = -100 \text{ J}.$$

The negative sign indicates that the system loses heat in this transition.

Significance

When a closed cycle is considered for the first law of thermodynamics, the change in internal energy around the whole path is equal to zero. If friction were to play a role in this example, less work would result from this heat added. [\[link\]](#) takes into consideration what happens if friction plays a role.

Notice that in [\[link\]](#), we did not assume that the transitions were quasi-static. This is because the first law is not subject to such a restriction. It describes transitions between equilibrium states but is not concerned with the intermediate states. The system does not have to pass through only equilibrium states. For example, if a gas in a steel container at a well-defined temperature and pressure is made to explode by means of a spark, some of the gas may condense, different gas molecules may combine to form new

compounds, and there may be all sorts of turbulence in the container—but eventually, the system will settle down to a new equilibrium state. This system is clearly not in equilibrium during its transition; however, its behavior is still governed by the first law because the process starts and ends with the system in equilibrium states.

Example:**Polishing a Fitting**

A machinist polishes a 0.50-kg copper fitting with a piece of emery cloth for 2.0 min. He moves the cloth across the fitting at a constant speed of 1.0 m/s by applying a force of 20 N, tangent to the surface of the fitting. (a) What is the total work done on the fitting by the machinist? (b) What is the increase in the internal energy of the fitting? Assume that the change in the internal energy of the cloth is negligible and that no heat is exchanged between the fitting and its environment. (c) What is the increase in the temperature of the fitting?

Strategy

The machinist's force over a distance that can be calculated from the speed and time given is the work done on the system. The work, in turn, increases the internal energy of the system. This energy can be interpreted as the heat that raises the temperature of the system via its heat capacity. Be careful with the sign of each quantity.

Solution

- a. The power created by a force on an object or the rate at which the machinist does frictional work on the fitting is $\vec{\mathbf{F}} \cdot \vec{\mathbf{v}} = -Fv$. Thus, in an elapsed time Δt (2.0 min), the work done on the fitting is

Equation:

$$\begin{aligned} W &= -Fv\Delta t = -(20 \text{ N})(1.0 \text{ m/s})(1.2 \times 10^2 \text{ s}) \\ &= -2.4 \times 10^3 \text{ J.} \end{aligned}$$

- b. By assumption, no heat is exchanged between the fitting and its environment, so the first law gives for the change in the internal energy of the fitting:

Equation:

$$\Delta E_{\text{int}} = -W = 2.4 \times 10^3 \text{ J}.$$

- c. Since ΔE_{int} is path independent, the effect of the $2.4 \times 10^3 \text{ J}$ of work is the same as if it were supplied at atmospheric pressure by a transfer of heat. Thus,

Equation:

$$2.4 \times 10^3 \text{ J} = mc\Delta T = (0.50 \text{ kg})(3.9 \times 10^2 \text{ J/kg} \cdot ^\circ\text{C})\Delta T,$$

and the increase in the temperature of the fitting is

Equation:

$$\Delta T = 12 ^\circ\text{C},$$

where we have used the value for the specific heat of copper,
 $c = 3.9 \times 10^2 \text{ J/kg} \cdot ^\circ\text{C}$.

Significance

If heat were released, the change in internal energy would be less and cause less of a temperature change than what was calculated in the problem.

Note:**Exercise:****Problem:**

Check Your Understanding The quantities below represent four different transitions between the same initial and final state. Fill in the blanks.

$Q \text{ (J)}$	$W \text{ (J)}$	$\Delta E_{\text{int}} \text{ (J)}$
-80	-120	
90		
	40	
	-40	

Solution:

Line 1, $\Delta E_{\text{int}} = 40 \text{ J}$; line 2, $W = 50 \text{ J}$ and $\Delta E_{\text{int}} = 40 \text{ J}$; line 3, $Q = 80 \text{ J}$ and $\Delta E_{\text{int}} = 40 \text{ J}$; and line 4, $Q = 0$ and $\Delta E_{\text{int}} = 40 \text{ J}$

Example:

An Ideal Gas Making Transitions between Two States

Consider the quasi-static expansions of an ideal gas between the equilibrium states A and C of [\[link\]](#). If 515 J of heat are added to the gas as it traverses the path ABC , how much heat is required for the transition along ADC ?

Assume that

$p_1 = 2.10 \times 10^5 \text{ N/m}^2$, $p_2 = 1.05 \times 10^5 \text{ N/m}^2$, $V_1 = 2.25 \times 10^{-3} \text{ m}^3$, and $V_2 = 4.50 \times 10^{-3} \text{ m}^3$.

Strategy

The difference in work done between process ABC and process ADC is the area enclosed by $ABCD$. Because the change of the internal energy (a function of state) is the same for both processes, the difference in work is thus the same as the difference in heat transferred to the system.

Solution

For path ABC , the heat added is $Q_{ABC} = 515 \text{ J}$ and the work done by the gas is the area under the path on the pV diagram, which is

Equation:

$$W_{ABC} = p_1(V_2 - V_1) = 473 \text{ J.}$$

Along ADC , the work done by the gas is again the area under the path:

Equation:

$$W_{ADC} = p_2(V_2 - V_1) = 236 \text{ J.}$$

Then using the strategy we just described, we have

Equation:

$$Q_{ADC} - Q_{ABC} = W_{ADC} - W_{ABC},$$

which leads to

Equation:

$$Q_{ADC} = Q_{ABC} + W_{ADC} - W_{ABC} = (515 + 236 - 473) \text{ J} = 278 \text{ J.}$$

Significance

The work calculations in this problem are made simple since no work is done along AD and BC and along AB and DC ; the pressure is constant over the volume change, so the work done is simply $p\Delta V$. An isothermal line could also have been used, as we have derived the work for an isothermal process as $W = nRT \ln \frac{V_2}{V_1}$.

Example:

Isothermal Expansion of an Ideal Gas

Heat is added to 1 mol of an ideal monatomic gas confined to a cylinder with a movable piston at one end. The gas expands quasi-statically at a constant temperature of 300 K until its volume increases from V to $3V$. (a) What is the change in internal energy of the gas? (b) How much work does the gas do? (c) How much heat is added to the gas?

Strategy

(a) Because the system is an ideal gas, the internal energy only changes when the temperature changes. (b) The heat added to the system is therefore purely used to do work that has been calculated in [Work, Heat, and Internal Energy](#). (c) Lastly, the first law of thermodynamics can be used to calculate the heat added to the gas.

Solution

- a. We saw in the preceding section that the internal energy of an ideal monatomic gas is a function only of temperature. Since $\Delta T = 0$, for this process, $\Delta E_{\text{int}} = 0$.
- b. The quasi-static isothermal expansion of an ideal gas was considered in the preceding section and was found to be

Equation:

$$\begin{aligned} W &= nRT \ln \frac{V_2}{V_1} = nRT \ln \frac{3V}{V} \\ &= (1.00 \text{ mol})(8.314 \text{ J/K} \cdot \text{mol})(300 \text{ K})(\ln 3) = 2.74 \times 10^3 \text{ J}. \end{aligned}$$

- c. With the results of parts (a) and (b), we can use the first law to determine the heat added:

Equation:

$$\Delta E_{\text{int}} = Q - W = 0,$$

which leads to

Equation:

$$Q = W = 2.74 \times 10^3 \text{ J}.$$

Significance

An isothermal process has no change in the internal energy. Based on that, the first law of thermodynamics reduces to $Q = W$.

Note:

Exercise:

Problem:

Check Your Understanding Why was it necessary to state that the process of [\[link\]](#) is quasi-static?

Solution:

So that the process is represented by the curve $p = nRT/V$ on the pV plot for the evaluation of work.

Example:**Vaporizing Water**

When 1.00 g of water at 100 °C changes from the liquid to the gas phase at atmospheric pressure, its change in volume is $1.67 \times 10^{-3} \text{ m}^3$. (a) How much heat must be added to vaporize the water? (b) How much work is done by the water against the atmosphere in its expansion? (c) What is the change in the internal energy of the water?

Strategy

We can first figure out how much heat is needed from the latent heat of vaporization of the water. From the volume change, we can calculate the work done from $W = p\Delta V$ because the pressure is constant. Then, the first law of thermodynamics provides us with the change in the internal energy.

Solution

- a. With L_v representing the latent heat of vaporization, the heat required to vaporize the water is

Equation:

$$Q = mL_v = (1.00 \text{ g})(2.26 \times 10^3 \text{ J/g}) = 2.26 \times 10^3 \text{ J}.$$

- b. Since the pressure on the system is constant at $1.00 \text{ atm} = 1.01 \times 10^5 \text{ N/m}^2$, the work done by the water as it is vaporized is

Equation:

$$W = p\Delta V = (1.01 \times 10^5 \text{ N/m}^2)(1.67 \times 10^{-3} \text{ m}^3) = 169 \text{ J}.$$

- c. From the first law, the thermal energy of the water during its vaporization changes by

Equation:

$$\Delta E_{\text{int}} = Q - W = 2.26 \times 10^3 \text{ J} - 169 \text{ J} = 2.09 \times 10^3 \text{ J}.$$

Significance

We note that in part (c), we see a change in internal energy, yet there is no change in temperature. Ideal gases that are not undergoing phase changes have the internal energy proportional to temperature. Internal energy in general is the sum of all energy in the system.

Note:

Exercise:

Problem:

Check Your Understanding When 1.00 g of ammonia boils at atmospheric pressure and $-33.0\text{ }^{\circ}\text{C}$, its volume changes from 1.47 to 1130 cm^3 . Its heat of vaporization at this pressure is $1.37 \times 10^6\text{ J/kg}$. What is the change in the internal energy of the ammonia when it vaporizes?

Solution:

$1.26 \times 10^3\text{ J}$.

Note:

View this [site](#) to learn about how the first law of thermodynamics. First, pump some heavy species molecules into the chamber. Then, play around by doing work (pushing the wall to the right where the person is located) to see how the internal energy changes (as seen by temperature). Then, look at how heat added changes the internal energy. Finally, you can set a parameter constant such as temperature and see what happens when you do work to keep the temperature constant (*Note:* You might see a change in these variables initially if you are moving around quickly in the simulation, but ultimately, this value will return to its equilibrium value).

Summary

- The internal energy of a thermodynamic system is a function of state and thus is unique for every equilibrium state of the system.
- The increase in the internal energy of the thermodynamic system is given by the heat added to the system less the work done by the system in any thermodynamics process.

Conceptual Questions

Exercise:

Problem:

What does the first law of thermodynamics tell us about the energy of the universe?

Exercise:

Problem:

Does adding heat to a system always increase its internal energy?

Solution:

If more work is done on the system than heat added, the internal energy of the system will actually decrease.

Exercise:

Problem:

A great deal of effort, time, and money has been spent in the quest for a so-called perpetual-motion machine, which is defined as a hypothetical machine that operates or produces useful work indefinitely and/or a hypothetical machine that produces more work or energy than it consumes. Explain, in terms of the first law of thermodynamics, why or why not such a machine is likely to be constructed.

Problems

Exercise:**Problem:**

When a dilute gas expands quasi-statically from 0.50 to 4.0 L, it does 250 J of work. Assuming that the gas temperature remains constant at 300 K, (a) what is the change in the internal energy of the gas? (b) How much heat is absorbed by the gas in this process?

Exercise:**Problem:**

In an expansion of gas, 500 J of work are done by the gas. If the internal energy of the gas increased by 80 J in the expansion, how much heat does the gas absorb?

Solution:

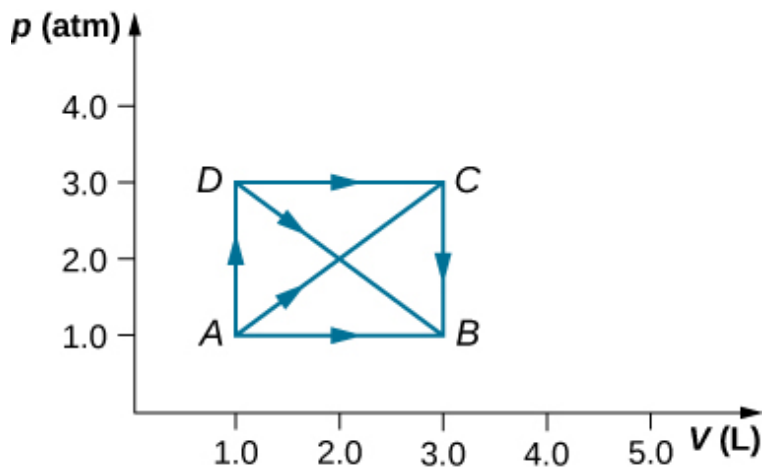
580 J

Exercise:**Problem:**

An ideal gas expands quasi-statically and isothermally from a state with pressure p and volume V to a state with volume $4V$. How much heat is added to the expanding gas?

Exercise:**Problem:**

As shown below, if the heat absorbed by the gas along AB is 400 J, determine the quantities of heat absorbed along (a) ADB; (b) ACB; and (c) ADCB.



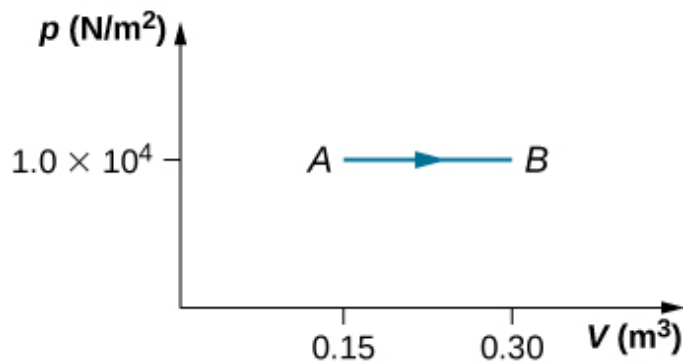
Solution:

a. 600 J; b. 600 J; c. 800 J

Exercise:

Problem:

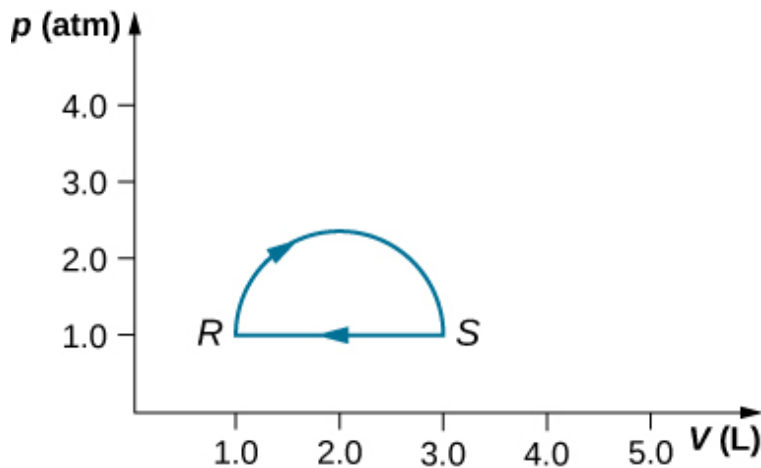
During the isobaric expansion from A to B represented below, 3,100 J of heat are added to the gas. What is the change in its internal energy?



Exercise:

Problem:

(a) What is the change in internal energy for the process represented by the closed path shown below? (b) How much heat is exchanged? (c) If the path is traversed in the opposite direction, how much heat is exchanged?

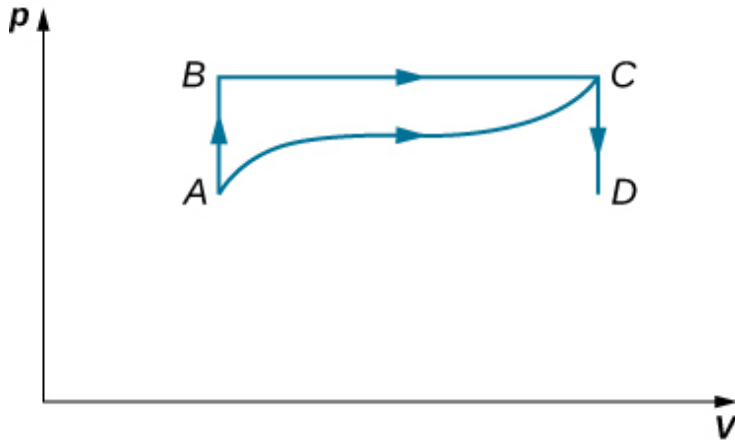


Solution:

a. 0; b. 160 J; c. -160 J

Exercise:**Problem:**

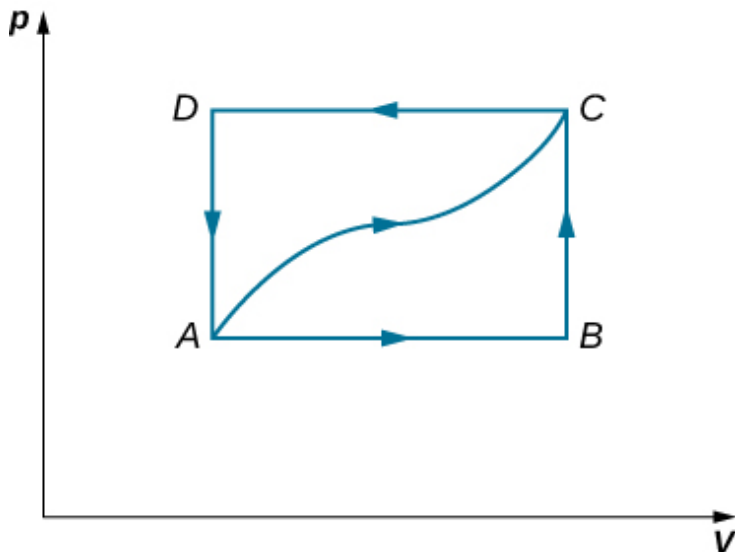
When a gas expands along path AC shown below, it does 400 J of work and absorbs either 200 or 400 J of heat. (a) Suppose you are told that along path ABC, the gas absorbs either 200 or 400 J of heat. Which of these values is correct? (b) Give the correct answer from part (a), how much work is done by the gas along ABC? (c) Along CD, the internal energy of the gas decreases by 50 J. How much heat is exchanged by the gas along this path?



Exercise:

Problem:

When a gas expands along AB (see below), it does 20 J of work and absorbs 30 J of heat. When the gas expands along AC , it does 40 J of work and absorbs 70 J of heat. (a) How much heat does the gas exchange along BC ? (b) When the gas makes the transition from C to A along CDA , 60 J of work are done on it from C to D . How much heat does it exchange along CDA ?



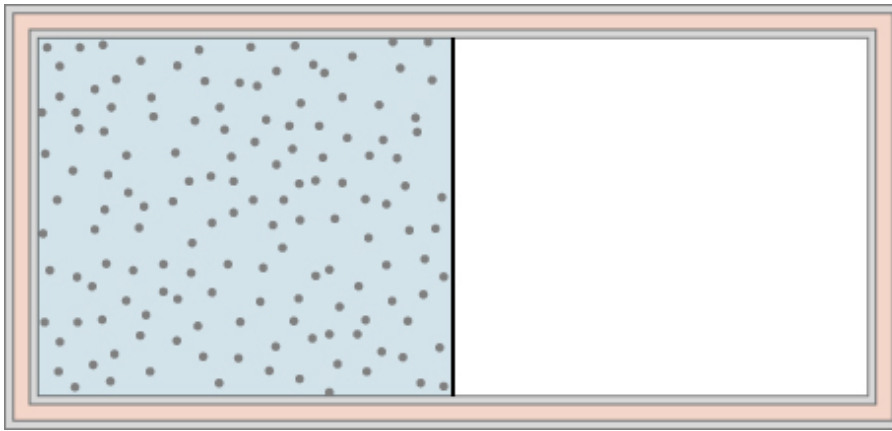
Solution:

a. 20 J; b. 90 J

Exercise:

Problem:

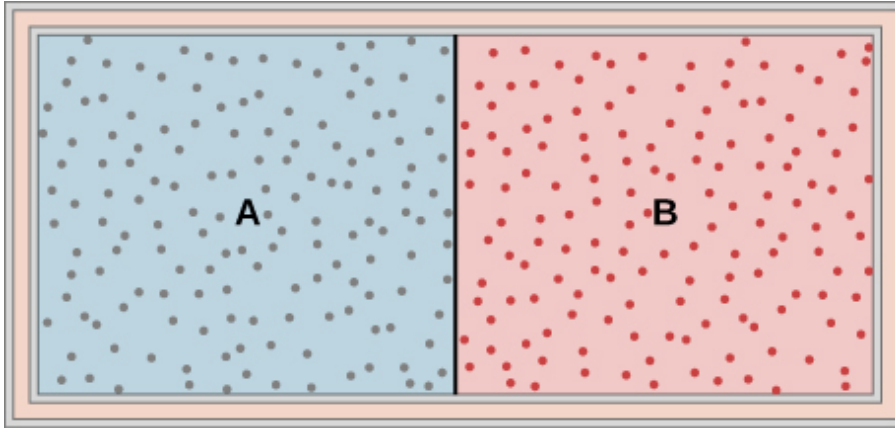
A dilute gas is stored in the left chamber of a container whose walls are perfectly insulating (see below), and the right chamber is evacuated. When the partition is removed, the gas expands and fills the entire container. Calculate the work done by the gas. Does the internal energy of the gas change in this process?



Exercise:

Problem:

Ideal gases A and B are stored in the left and right chambers of an insulated container, as shown below. The partition is removed and the gases mix. Is any work done in this process? If the temperatures of A and B are initially equal, what happens to their common temperature after they are mixed?



Solution:

No work is done and they reach the same common temperature.

Exercise:

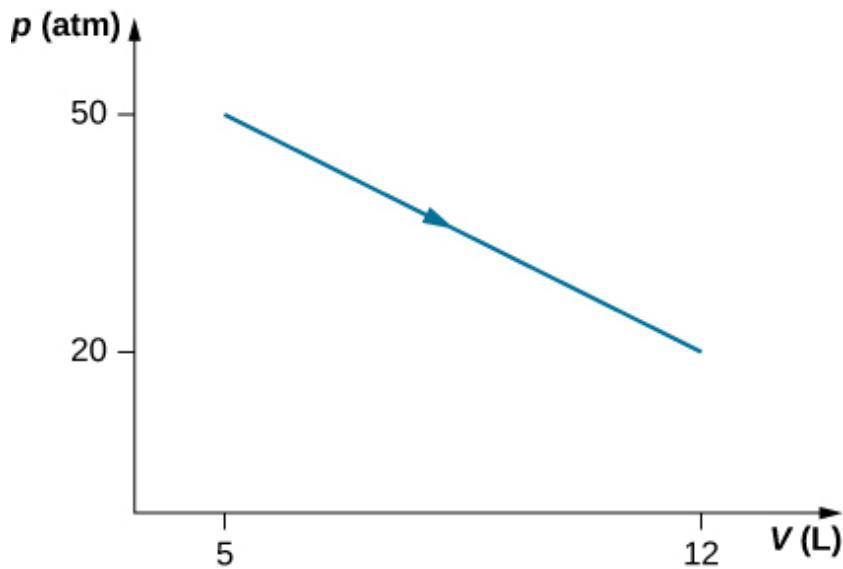
Problem:

An ideal monatomic gas at a pressure of $2.0 \times 10^5 \text{ N/m}^2$ and a temperature of 300 K undergoes a quasi-static isobaric expansion from 2.0×10^3 to $4.0 \times 10^3 \text{ cm}^3$. (a) What is the work done by the gas? (b) What is the temperature of the gas after the expansion? (c) How many moles of gas are there? (d) What is the change in internal energy of the gas? (e) How much heat is added to the gas?

Exercise:

Problem:

Consider the process for steam in a cylinder shown below. Suppose the change in the internal energy in this process is 30 kJ. Find the heat entering the system.



Solution:

54,500 J

Exercise:

Problem:

The state of 30 moles of steam in a cylinder is changed in a cyclic manner from a-b-c-a, where the pressure and volume of the states are: a (30 atm, 20 L), b (50 atm, 20 L), and c (50 atm, 45 L). Assume each change takes place along the line connecting the initial and final states in the pV plane. (a) Display the cycle in the pV plane. (b) Find the net work done by the steam in one cycle. (c) Find the net amount of heat flow in the steam over the course of one cycle.

Exercise:

Problem:

A monatomic ideal gas undergoes a quasi-static process that is described by the function $p(V) = p_1 + 3(V - V_1)$, where the starting state is (p_1, V_1) and the final state (p_2, V_2) . Assume the system consists of n moles of the gas in a container that can exchange heat with the environment and whose volume can change freely. (a) Evaluate the work done by the gas during the change in the state. (b) Find the change in internal energy of the gas. (c) Find the heat input to the gas during the change. (d) What are initial and final temperatures?

Solution:

a. $(p_1 - 3V_1)(V_2 - V_1) + \frac{3}{2}(V_2^2 - V_1^2)$; b. $\frac{3}{2}(p_2 V_2 - p_1 V_1)$; c. the sum of parts (a) and (b); d. $T_1 = \frac{p_1 V_1}{nR}$ and $T_2 = \frac{p_2 V_2}{nR}$

Exercise:**Problem:**

A metallic container of fixed volume of $2.5 \times 10^{-3} \text{ m}^3$ immersed in a large tank of temperature 27°C contains two compartments separated by a freely movable wall. Initially, the wall is kept in place by a stopper so that there are 0.02 mol of the nitrogen gas on one side and 0.03 mol of the oxygen gas on the other side, each occupying half the volume. When the stopper is removed, the wall moves and comes to a final position. The movement of the wall is controlled so that the wall moves in infinitesimal quasi-static steps. (a) Find the final volumes of the two sides assuming the ideal gas behavior for the two gases. (b) How much work does each gas do on the other? (c) What is the change in the internal energy of each gas? (d) Find the amount of heat that enters or leaves each gas.

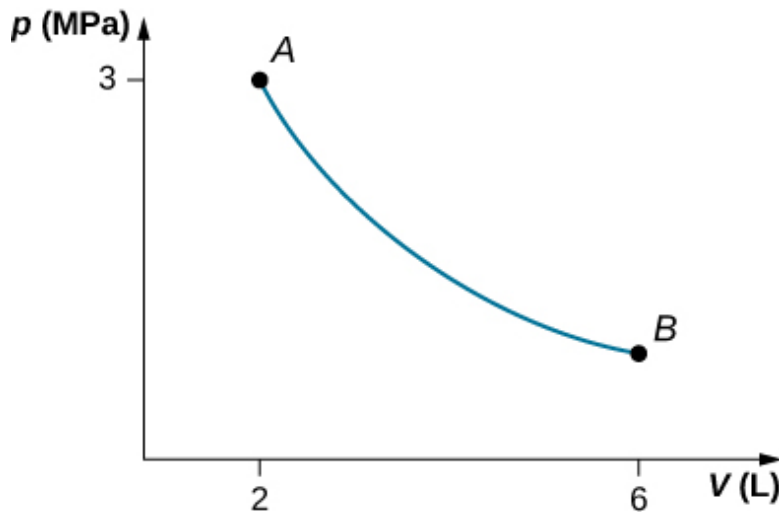
Exercise:

Problem:

A gas in a cylindrical closed container is adiabatically and quasi-statically expanded from a state A (3 MPa, 2 L) to a state B with volume of 6 L along the path $1.8 pV = \text{constant}$. (a) Plot the path in the pV plane. (b) Find the amount of work done by the gas and the change in the internal energy of the gas during the process.

Solution:

a.



;

b. $W = 4.39 \text{ kJ}$, $\Delta E_{\text{int}} = -4.39 \text{ kJ}$

Glossary

first law of thermodynamics

the change in internal energy for any transition between two equilibrium states is $\Delta E_{\text{int}} = Q - W$

Thermodynamic Processes

By the end of this section, you will be able to:

- Define a thermodynamic process
- Distinguish between quasi-static and non-quasi-static processes
- Calculate physical quantities, such as the heat transferred, work done, and internal energy change for isothermal, adiabatic, and cyclical thermodynamic processes

In solving mechanics problems, we isolate the body under consideration, analyze the external forces acting on it, and then use Newton's laws to predict its behavior. In thermodynamics, we take a similar approach. We start by identifying the part of the universe we wish to study; it is also known as our system. (We defined a system at the beginning of this chapter as anything whose properties are of interest to us; it can be a single atom or the entire Earth.) Once our system is selected, we determine how the environment, or surroundings, interact with the system. Finally, with the interaction understood, we study the thermal behavior of the system with the help of the laws of thermodynamics.

The thermal behavior of a system is described in terms of *thermodynamic variables*. For an ideal gas, these variables are pressure, volume, temperature, and the number of molecules or moles of the gas. Different types of systems are generally characterized by different sets of variables. For example, the thermodynamic variables for a stretched rubber band are tension, length, temperature, and mass.

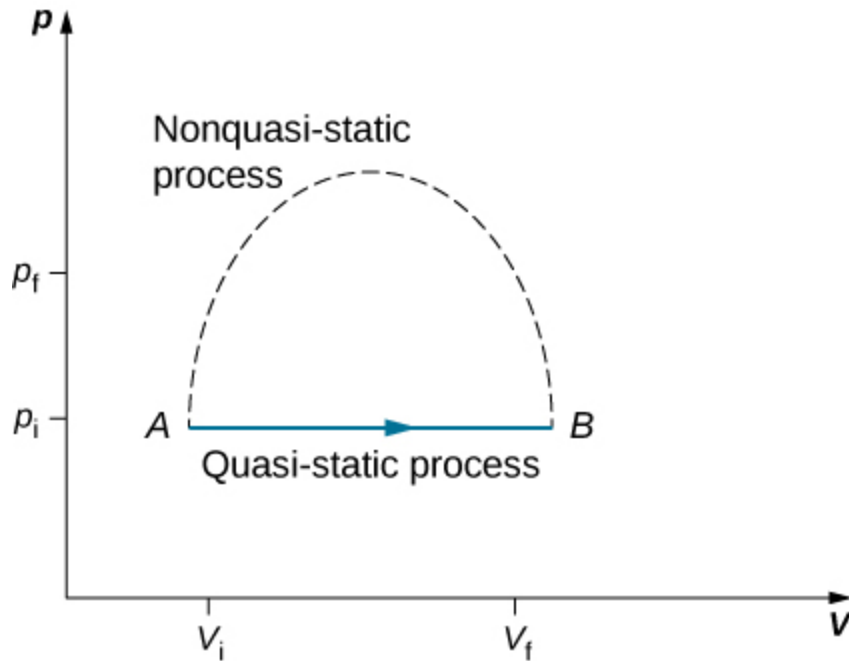
The state of a system can change as a result of its interaction with the environment. The change in a system can be fast or slow and large or small. The manner in which a state of a system can change from an initial state to a final state is called a **thermodynamic process**. For analytical purposes in thermodynamics, it is helpful to divide up processes as either *quasi-static* or *non-quasi-static*, as we now explain.

Quasi-static and Non-quasi-static Processes

A quasi-static process refers to an idealized or imagined process where the change in state is made infinitesimally slowly so that at each instant, the system can be assumed to be at a thermodynamic equilibrium with itself and with the environment. For instance, imagine heating 1 kg of water from a temperature 20°C to 21°C at a constant pressure of 1 atmosphere. To heat the water very slowly, we may imagine placing the container with water in a large bath that can be slowly heated such that the temperature of the bath can rise infinitesimally slowly from 20°C to 21°C . If we put 1 kg of water at 20°C directly into a bath at 21°C , the temperature of the water will rise rapidly to 21°C in a non-quasi-static way.

Quasi-static processes are done slowly enough that the system remains at thermodynamic equilibrium at each instant, despite the fact that the system changes over time. The thermodynamic equilibrium of the system is necessary for the system to have well-defined values of macroscopic properties such as the temperature and the pressure of the system at each instant of the process. Therefore, quasi-static processes can be shown as well-defined paths in state space of the system.

Since quasi-static processes cannot be completely realized for any finite change of the system, all processes in nature are non-quasi-static. Examples of quasi-static and non-quasi-static processes are shown in [\[link\]](#). Despite the fact that all finite changes must occur essentially non-quasi-statically at some stage of the change, we can imagine performing infinitely many quasi-static process corresponding to every quasi-static process. Since quasi-static processes can be analyzed analytically, we mostly study quasi-static processes in this book. We have already seen that in a quasi-static process the work by a gas is given by $\int p dV$.

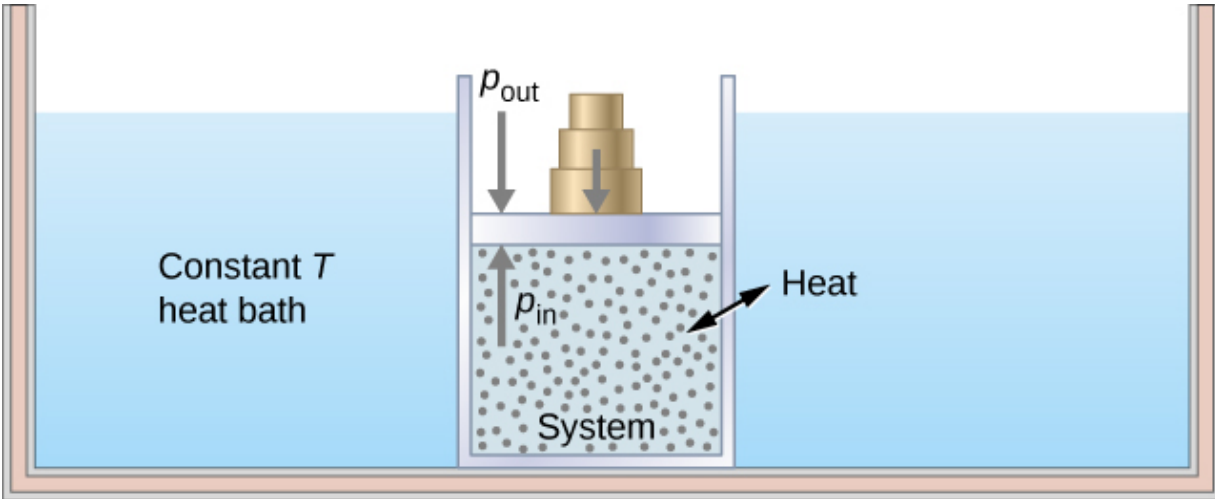


Quasi-static and non-quasi-static processes between states A and B of a gas. In a quasi-static process, the path of the process between A and B can be drawn in a state diagram since all the states that the system goes through are known. In a non-quasi-static process, the states between A and B are not known, and hence no path can be drawn. It may follow the dashed line as shown in the figure or take a very different path.

Isothermal Processes

An **isothermal process** is a change in the state of the system at a constant temperature. This process is accomplished by keeping the system in thermal equilibrium with a large heat bath during the process. Recall that a heat bath is an idealized “infinitely” large system whose temperature does not change. In practice, the temperature of a finite bath is controlled by either adding or removing a finite amount of energy as the case may be.

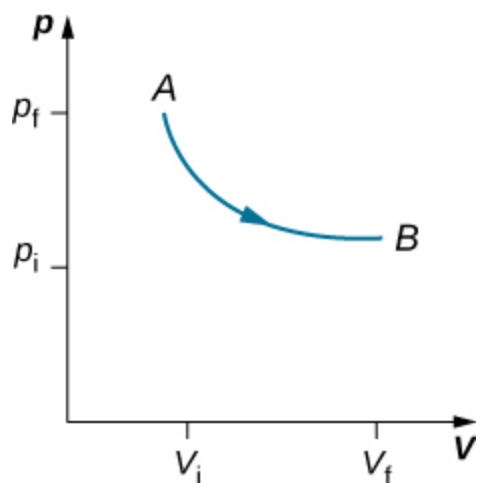
As an illustration of an isothermal process, consider a cylinder of gas with a movable piston immersed in a large water tank whose temperature is maintained constant. Since the piston is freely movable, the pressure inside P_{in} is balanced by the pressure outside P_{out} by some weights on the piston, as in [\[link\]](#).



Expanding a system at a constant temperature. Removing weights on the piston leads to an imbalance of forces on the piston, which causes the piston to move up. As the piston moves up, the temperature is lowered momentarily, which causes heat to flow from the heat bath to the system. The energy to move the piston eventually comes from the heat bath.

As weights on the piston are removed, an imbalance of forces on the piston develops. The net nonzero force on the piston would cause the piston to accelerate, resulting in an increase in volume. The expansion of the gas cools the gas to a lower temperature, which makes it possible for the heat to enter from the heat bath into the system until the temperature of the gas is reset to the temperature of the heat bath. If weights are removed in infinitesimal steps, the pressure in the system decreases infinitesimally slowly. This way, an isothermal process can be conducted quasi-statically. An isothermal line on a (p, V) diagram is represented by a curved line from

starting point A to finishing point B , as seen in [\[link\]](#). For an ideal gas, an isothermal process is hyperbolic, since for an ideal gas at constant temperature, $p \propto \frac{1}{V}$.



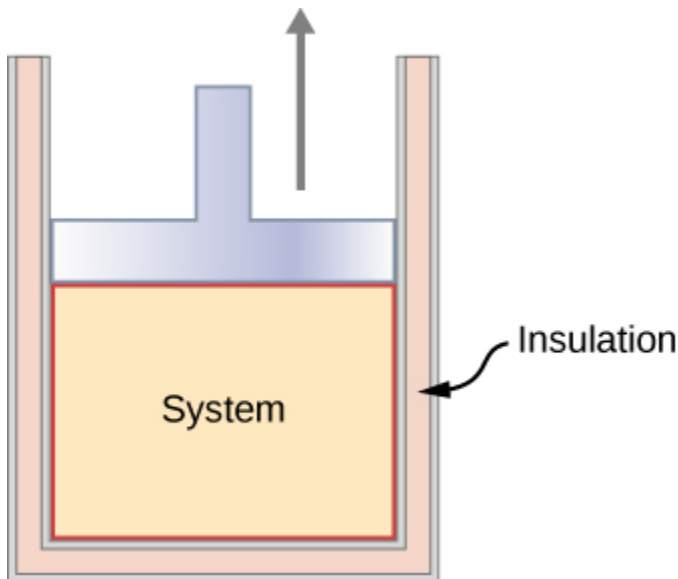
An isothermal expansion from a state labeled A to another state labeled B on a pV diagram. The curve represents the relation between pressure and volume in an ideal gas at constant temperature.

An isothermal process studied in this chapter is quasi-statically performed, since to be isothermal throughout the change of volume, you must be able to state the temperature of the system at each step, which is possible only if the system is in thermal equilibrium continuously. The system must go out of equilibrium for the state to change, but for quasi-static processes, we imagine that the process is conducted in infinitesimal steps such that these departures from equilibrium can be made as brief and as small as we like.

Other quasi-static processes of interest for gases are isobaric and isochoric processes. An **isobaric process** is a process where the pressure of the system does not change, whereas an **isochoric process** is a process where the volume of the system does not change.

Adiabatic Processes

In an **adiabatic process**, the system is insulated from its environment so that although the state of the system changes, no heat is allowed to enter or leave the system, as seen in [\[link\]](#). An adiabatic process can be conducted either quasi-statically or non-quasi-statically. When a system expands adiabatically, it must do work against the outside world, and therefore its energy goes down, which is reflected in the lowering of the temperature of the system. An adiabatic expansion leads to a lowering of temperature, and an adiabatic compression leads to an increase of temperature. We discuss adiabatic expansion again in [Adiabatic Processes for an ideal Gas](#).



An insulated piston with a hot, compressed gas is released. The piston moves up, the volume expands, and the pressure and temperature decrease. The internal

energy goes into work. If the expansion occurs within a time frame in which negligible heat can enter the system, then the process is called adiabatic. Ideally, during an adiabatic process no heat enters or exits the system.

Cyclic Processes

We say that a system goes through a **cyclic process** if the state of the system at the end is same as the state at the beginning. Therefore, state properties such as temperature, pressure, volume, and internal energy of the system do not change over a complete cycle:

Equation:

$$\Delta E_{\text{int}} = 0.$$

When the first law of thermodynamics is applied to a cyclic process, we obtain a simple relation between heat into the system and the work done by the system over the cycle:

Equation:

$$Q = W \text{ (cyclic process).}$$

Thermodynamic processes are also distinguished by whether or not they are reversible. A **reversible process** is one that can be made to retrace its path by differential changes in the environment. Such a process must therefore also be quasi-static. Note, however, that a quasi-static process is not necessarily reversible, since there may be dissipative forces involved. For example, if friction occurred between the piston and the walls of the cylinder containing the gas, the energy lost to friction would prevent us from reproducing the original states of the system.

We considered several thermodynamic processes:

1. An isothermal process, during which the system's temperature remains constant
2. An adiabatic process, during which no heat is transferred to or from the system
3. An isobaric process, during which the system's pressure does not change
4. An isochoric process, during which the system's volume does not change

Many other processes also occur that do not fit into any of these four categories.

Note:

View this [site](#) to set up your own process in a pV diagram. See if you can calculate the values predicted by the simulation for heat, work, and change in internal energy.

Summary

- The thermal behavior of a system is described in terms of thermodynamic variables. For an ideal gas, these variables are pressure, volume, temperature, and number of molecules or moles of the gas.
- For systems in thermodynamic equilibrium, the thermodynamic variables are related by an equation of state.
- A heat reservoir is so large that when it exchanges heat with other systems, its temperature does not change.
- A quasi-static process takes place so slowly that the system involved is always in thermodynamic equilibrium.
- A reversible process is one that can be made to retrace its path and both the temperature and pressure are uniform throughout the system.

- There are several types of thermodynamic processes, including (a) isothermal, where the system's temperature is constant; (b) adiabatic, where no heat is exchanged by the system; (c) isobaric, where the system's pressure is constant; and (d) isochoric, where the system's volume is constant.
- As a consequence of the first law of thermodynamics, here is a summary of the thermodynamic processes: (a) isothermal: $\Delta E_{\text{int}} = 0, Q = W$; (b) adiabatic: $Q = 0, \Delta E_{\text{int}} = -W$; (c) isobaric: $\Delta E_{\text{int}} = Q - W$; and (d) isochoric: $W = 0, \Delta E_{\text{int}} = Q$.

Conceptual Questions

Exercise:

Problem:

When a gas expands isothermally, it does work. What is the source of energy needed to do this work?

Solution:

The system must be in contact with a heat source that allows heat to flow into the system.

Exercise:

Problem:

If the pressure and volume of a system are given, is the temperature always uniquely determined?

Exercise:

Problem:

It is unlikely that a process can be isothermal unless it is a very slow process. Explain why. Is the same true for isobaric and isochoric processes? Explain your answer.

Solution:

Isothermal processes must be slow to make sure that as heat is transferred, the temperature does not change. Even for isobaric and isochoric processes, the system must be in thermal equilibrium with slow changes of thermodynamic variables.

Problems

Exercise:

Problem:

Two moles of a monatomic ideal gas at (5 MPa, 5 L) is expanded isothermally until the volume is doubled (step 1). Then it is cooled isochorically until the pressure is 1 MPa (step 2). The temperature drops in this process. The gas is now compressed isothermally until its volume is back to 5 L, but its pressure is now 2 MPa (step 3). Finally, the gas is heated isochorically to return to the initial state (step 4). (a) Draw the four processes in the pV plane. (b) Find the total work done by the gas.

Exercise:

Problem:

Consider a transformation from point A to B in a two-step process. First, the pressure is lowered from 3 MPa at point A to a pressure of 1 MPa, while keeping the volume at 2 L by cooling the system. The state reached is labeled C . Then the system is heated at a constant pressure to reach a volume of 6 L in the state B . (a) Find the amount of work done on the ACB path. (b) Find the amount of heat exchanged by the system when it goes from A to B on the ACB path. (c) Compare the change in the internal energy when the AB process occurs isothermally with the AB change through the two-step process on the ACB path.

Solution:

a. 4000 J; b. -4000 J; c. It does not depend on the process.

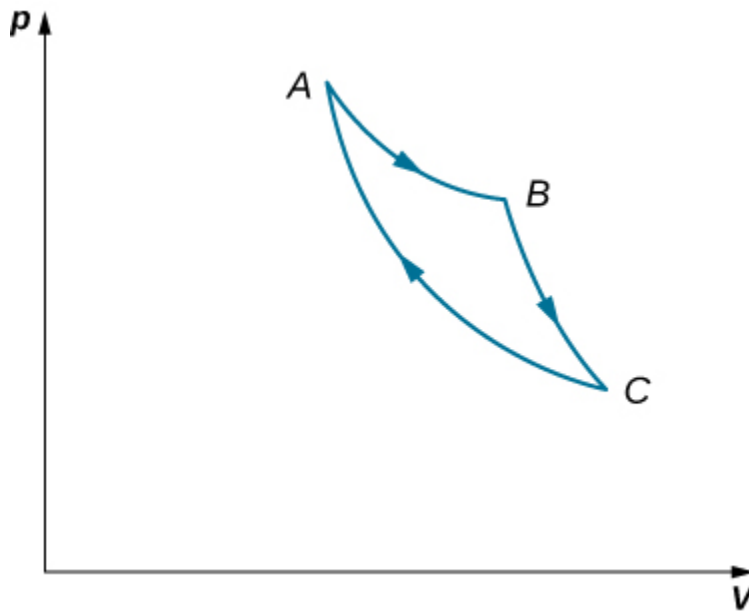
Exercise:

Problem:

Consider a cylinder with a movable piston containing n moles of an ideal gas. The entire apparatus is immersed in a constant temperature bath of temperature T kelvin. The piston is then pushed slowly so that the pressure of the gas changes quasi-statically from p_1 to p_2 at constant temperature T . Find the work done by the gas in terms of n , R , T , p_1 , and p_2 .

Exercise:**Problem:**

An ideal gas expands isothermally along AB and does 700 J of work (see below). (a) How much heat does the gas exchange along AB? (b) The gas then expands adiabatically along BC and does 400 J of work. When the gas returns to A along CA, it exhausts 100 J of heat to its surroundings. How much work is done on the gas along this path?



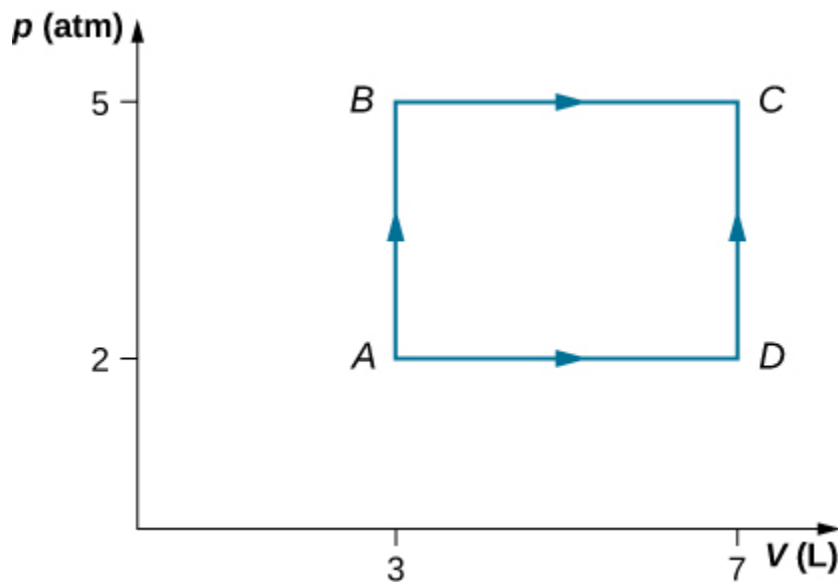
Solution:

a. 700 J; b. 500 J

Exercise:

Problem:

Consider the processes shown below for a monatomic gas. (a) Find the work done in each of the processes AB, BC, AD, and DC. (b) Find the internal energy change in processes AB and BC. (c) Find the internal energy difference between states C and A. (d) Find the total heat added in the ADC process. (e) From the information given, can you find the heat added in process AD? Why or why not?

**Exercise:****Problem:**

Two moles of helium gas are placed in a cylindrical container with a piston. The gas is at room temperature 25°C and under a pressure of $3.0 \times 10^5 \text{ Pa}$. When the pressure from the outside is decreased while keeping the temperature the same as the room temperature, the volume of the gas doubles. (a) Find the work the external agent does on the gas in the process. (b) Find the heat exchanged by the gas and indicate whether the gas takes in or gives up heat. Assume ideal gas behavior.

Solution:

a. $-3\,400 \text{ J}$; b. 3400 J enters the gas

Exercise:

Problem:

An amount of n moles of a monatomic ideal gas in a conducting container with a movable piston is placed in a large thermal heat bath at temperature T_1 and the gas is allowed to come to equilibrium. After the equilibrium is reached, the pressure on the piston is lowered so that the gas expands at constant temperature. The process is continued quasi-statically until the final pressure is $4/3$ of the initial pressure p_1 . (a) Find the change in the internal energy of the gas. (b) Find the work done by the gas. (c) Find the heat exchanged by the gas, and indicate, whether the gas takes in or gives up heat.

Glossary

adiabatic process

process during which no heat is transferred to or from the system

cyclic process

process in which the state of the system at the end is same as the state at the beginning

isobaric process

process during which the system's pressure does not change

isochoric process

process during which the system's volume does not change

isothermal process

process during which the system's temperature remains constant

reversible process

process that can be reverted to restore both the system and its environment back to their original states together

thermodynamic process

manner in which a state of a system can change from initial state to final state

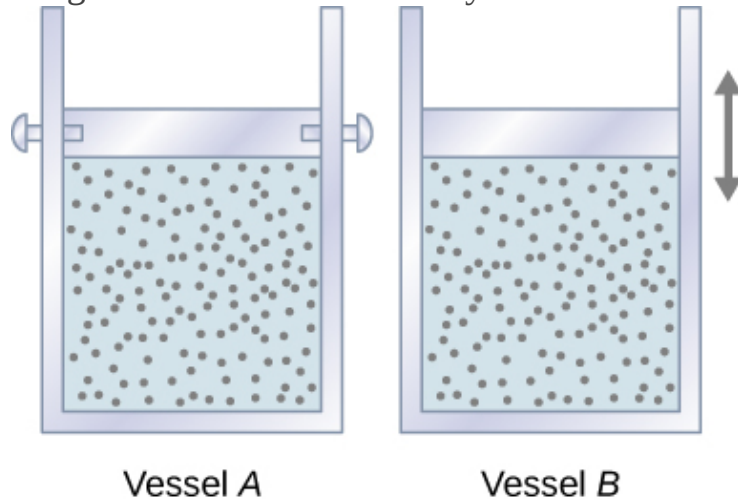
Heat Capacities of an Ideal Gas

By the end of this section, you will be able to:

- Define heat capacity of an ideal gas for a specific process
- Calculate the specific heat of an ideal gas for either an isobaric or isochoric process
- Explain the difference between the heat capacities of an ideal gas and a real gas
- Estimate the change in specific heat of a gas over temperature ranges

We learned about specific heat and molar heat capacity in [Temperature and Heat](#); however, we have not considered a process in which heat is added. We do that in this section. First, we examine a process where the system has a constant volume, then contrast it with a system at constant pressure and show how their specific heats are related.

Let's start with looking at [\[link\]](#), which shows two vessels *A* and *B*, each containing 1 mol of the same type of ideal gas at a temperature T and a volume V . The only difference between the two vessels is that the piston at the top of *A* is fixed, whereas the one at the top of *B* is free to move against a constant external pressure p . We now consider what happens when the temperature of the gas in each vessel is slowly increased to $T + dT$ with the addition of heat.



Two vessels are identical except that the piston at the top of *A* is fixed, whereas that atop *B* is free to move against a constant external pressure p .

Since the piston of vessel *A* is fixed, the volume of the enclosed gas does not change. Consequently, the gas does no work, and we have from the first law

Equation:

$$dE_{\text{int}} = dQ - dW = dQ.$$

We represent the fact that the heat is exchanged at constant volume by writing

Equation:

$$dQ = C_V n dT,$$

where C_V is the **molar heat capacity at constant volume** of the gas. In addition, since $dE_{\text{int}} = dQ$ for this particular process,

Equation:

$$dE_{\text{int}} = C_V n dT.$$

We obtained this equation assuming the volume of the gas was fixed. However, internal energy is a state function that depends on only the temperature of an ideal gas. Therefore, $dE_{\text{int}} = C_V n dT$ gives the change in internal energy of an ideal gas for any process involving a temperature change dT .

When the gas in vessel *B* is heated, it expands against the movable piston and does work $dW = p dV$. In this case, the heat is added at constant pressure, and we write

Equation:

$$dQ = C_p n dT,$$

where C_p is the **molar heat capacity at constant pressure** of the gas.

Furthermore, since the ideal gas expands against a constant pressure,

Equation:

$$d(pV) = d(RnT)$$

becomes

Equation:

$$pdV = RndT.$$

Finally, inserting the expressions for dQ and pdV into the first law, we obtain

Equation:

$$dE_{\text{int}} = dQ - pdV = (C_p n - Rn)dT.$$

We have found dE_{int} for both an isochoric and an isobaric process. Because the internal energy of an ideal gas depends only on the temperature, dE_{int} must be the same for both processes. Thus,

Equation:

$$C_V ndT = (C_p n - Rn)dT,$$

and

Note:

Equation:

$$C_p = C_V + R.$$

The derivation of [\[link\]](#) was based only on the ideal gas law. Consequently, this relationship is approximately valid for all dilute gases, whether monatomic like He, diatomic like O₂, or polyatomic like CO₂ or NH₃.

In the preceding chapter, we found the molar heat capacity of an ideal gas under constant volume to be

Equation:

$$C_V = \frac{d}{2}R,$$

where d is the number of degrees of freedom of a molecule in the system. [\[link\]](#) shows the molar heat capacities of some dilute ideal gases at room temperature. The heat capacities of real gases are somewhat higher than those predicted by the expressions of C_V and C_p given in [\[link\]](#). This indicates that vibrational motion in polyatomic molecules is significant, even at room temperature. Nevertheless, the difference in the molar heat capacities, $C_p - C_V$, is very close to R , even for the polyatomic gases.

Molar Heat Capacities of Dilute Ideal Gases at Room Temperature				
Type of Molecule	Gas	C_p (J/mol K)	C_V (J/mol K)	$C_p - C_V$ (J/mol K)
Monatomic	Ideal	$\frac{5}{2}R = 20.79$	$\frac{3}{2}R = 12.47$	$R = 8.31$
Diatomic	Ideal	$\frac{7}{2}R = 29.10$	$\frac{5}{2}R = 20.79$	$R = 8.31$
Polyatomic	Ideal	$4R = 33.26$	$3R = 24.94$	$R = 8.31$

Summary

- For an ideal gas, the molar capacity at constant pressure C_p is given by $C_p = C_V + R = dR/2 + R$, where d is the number of degrees of freedom of each molecule/entity in the system.
- A real gas has a specific heat close to but a little bit higher than that of the corresponding ideal gas with $C_p \simeq C_V + R$.

Conceptual Questions

Exercise:

Problem:

How can an object transfer heat if the object does not possess a discrete quantity of heat?

Exercise:

Problem:

Most materials expand when heated. One notable exception is water between $0\text{ }^{\circ}\text{C}$ and $4\text{ }^{\circ}\text{C}$, which actually decreases in volume with the increase in temperature. Which is greater for water in this temperature region, C_p or C_V ?

Solution:

Typically C_p is greater than C_V because when expansion occurs under constant pressure, it does work on the surroundings. Therefore, heat can go into internal energy and work. Under constant volume, all heat goes into internal energy. In this example, water contracts upon heating, so if we add heat at constant pressure, work is done on the water by surroundings and therefore, C_p is less than C_V .

Exercise:

Problem:

Why are there two specific heats for gases C_p and C_V , yet only one given for solid?

Problems

Exercise:

Problem:

The temperature of an ideal monatomic gas rises by 8.0 K . What is the change in the internal energy of 1 mol of the gas at constant volume?

Solution:

100 J

Exercise:**Problem:**

For a temperature increase of $10\text{ }^{\circ}\text{C}$ at constant volume, what is the heat absorbed by (a) 3.0 mol of a dilute monatomic gas; (b) 0.50 mol of a dilute diatomic gas; and (c) 15 mol of a dilute polyatomic gas?

Exercise:**Problem:**

If the gases of the preceding problem are initially at 300 K, what are their internal energies after they absorb the heat?

Solution:

a. 370 J; b. 100 J; c. 500 J

Exercise:**Problem:**

Consider 0.40 mol of dilute carbon dioxide at a pressure of 0.50 atm and a volume of 50 L. What is the internal energy of the gas?

Exercise:**Problem:**

When 400 J of heat are slowly added to 10 mol of an ideal monatomic gas, its temperature rises by $10\text{ }^{\circ}\text{C}$. What is the work done on the gas?

Solution:

850 J

Exercise:

Problem:

One mole of a dilute diatomic gas occupying a volume of 10.00 L expands against a constant pressure of 2.000 atm when it is slowly heated. If 400.0 J of heat are added in the process, what is its final volume?

Glossary

molar heat capacity at constant pressure

quantifies the ratio of the amount of heat added removed to the temperature while measuring at constant pressure

molar heat capacity at constant volume

quantifies the ratio of the amount of heat added removed to the temperature while measuring at constant volume

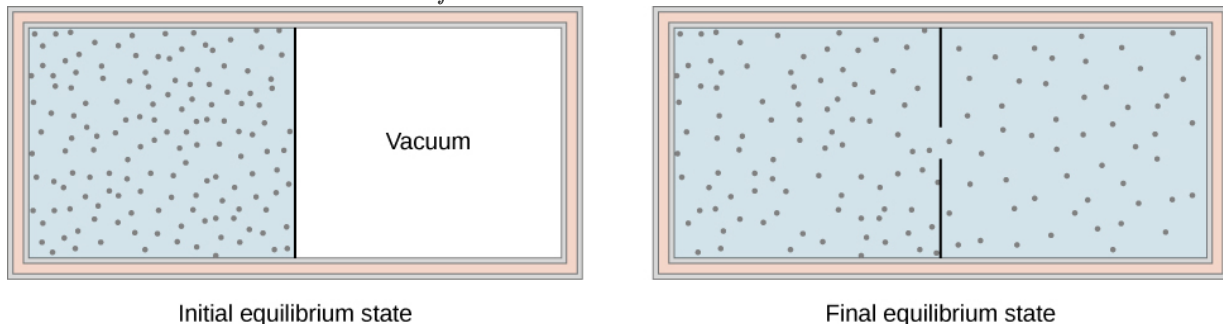
Adiabatic Processes for an Ideal Gas

By the end of this section, you will be able to:

- Define adiabatic expansion of an ideal gas
- Demonstrate the qualitative difference between adiabatic and isothermal expansions

When an ideal gas is compressed adiabatically ($Q = 0$), work is done on it and its temperature increases; in an adiabatic expansion, the gas does work and its temperature drops. Adiabatic compressions actually occur in the cylinders of a car, where the compressions of the gas-air mixture take place so quickly that there is no time for the mixture to exchange heat with its environment. Nevertheless, because work is done on the mixture during the compression, its temperature does rise significantly. In fact, the temperature increases can be so large that the mixture can explode without the addition of a spark. Such explosions, since they are not timed, make a car run poorly—it usually “knocks.” Because ignition temperature rises with the octane of gasoline, one way to overcome this problem is to use a higher-octane gasoline.

Another interesting adiabatic process is the free expansion of a gas. [\[link\]](#) shows a gas confined by a membrane to one side of a two-compartment, thermally insulated container. When the membrane is punctured, gas rushes into the empty side of the container, thereby expanding freely. Because the gas expands “against a vacuum” ($p = 0$), it does no work, and because the vessel is thermally insulated, the expansion is adiabatic. With $Q = 0$ and $W = 0$ in the first law, $\Delta E_{\text{int}} = 0$, so $E_{\text{int}i} = E_{\text{int}f}$ for the free expansion.



The gas in the left chamber expands freely into the right chamber when the membrane is punctured.

If the gas is ideal, the internal energy depends only on the temperature. Therefore, when an ideal gas expands freely, its temperature does not change.

A quasi-static, adiabatic expansion of an ideal gas is represented in [\[link\]](#), which shows an insulated cylinder that contains 1 mol of an ideal gas. The gas is made to expand quasi-statically by removing one grain of sand at a time from the top of the piston. When the gas expands by dV , the change in its temperature is dT . The work done by the gas in the expansion is $dW = pdV$; $dQ = 0$ because the cylinder is insulated; and the change in the internal energy of the gas is, from [\[link\]](#), $dE_{\text{int}} = C_V n dT$. Therefore, from the first law,

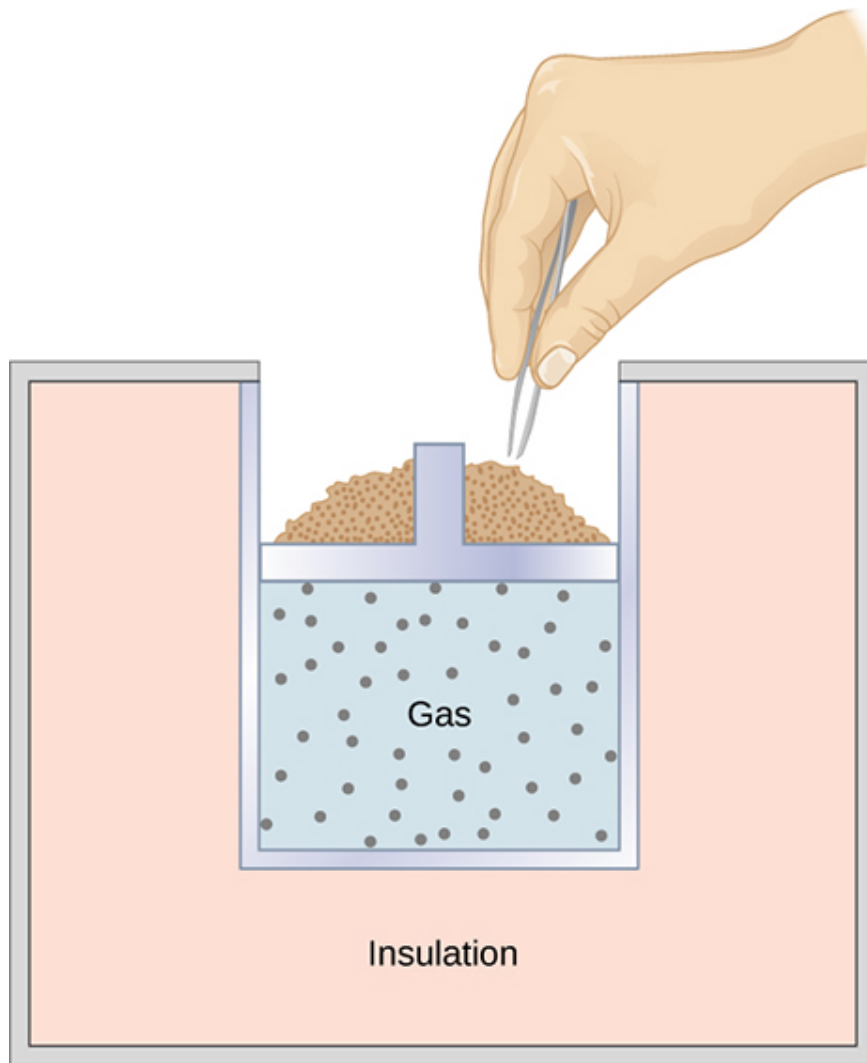
Equation:

$$C_V n dT = 0 - pdV = -pdV,$$

so

Equation:

$$dT = -\frac{pdV}{C_V n}.$$



When sand is removed from the piston one grain at a time, the gas expands adiabatically and quasi-statically in the insulated vessel.

Also, for 1 mol of an ideal gas,

Equation:

$$d(pV) = d(RnT),$$

so

Equation:

$$pdV + Vdp = RndT$$

and

Equation:

$$dT = \frac{pdV + Vdp}{Rn}.$$

We now have two equations for dT . Upon equating them, we find that

Equation:

$$C_V n V dp + (C_V n + Rn) p dV = 0.$$

Now, we divide this equation by npV and use $C_p = C_V + R$. We are then left with

Equation:

$$C_V \frac{dp}{p} + C_p \frac{dV}{V} = 0,$$

which becomes

Equation:

$$\frac{dp}{p} + \gamma \frac{dV}{V} = 0,$$

where we define γ as the ratio of the molar heat capacities:

Note:

Equation:

$$\gamma = \frac{C_p}{C_V}.$$

Thus,

Equation:

$$\int \frac{dp}{p} + \gamma \int \frac{dV}{V} = 0$$

and

Equation:

$$\ln p + \gamma \ln V = \text{constant}.$$

Finally, using $\ln(A^x) = x \ln A$ and $\ln AB = \ln A + \ln B$, we can write this in the form

Note:

Equation:

$$pV^\gamma = \text{constant}.$$

This equation is the condition that must be obeyed by an ideal gas in a quasi-static adiabatic process. For example, if an ideal gas makes a quasi-static adiabatic transition from a state with pressure and volume p_1 and V_1 to a state with p_2 and V_2 , then it must be true that $p_1 V_1^\gamma = p_2 V_2^\gamma$.

The adiabatic condition of [\[link\]](#) can be written in terms of other pairs of thermodynamic variables by combining it with the ideal gas law. In doing this, we find that

Equation:

$$p^{1-\gamma} T^\gamma = \text{constant}$$

and

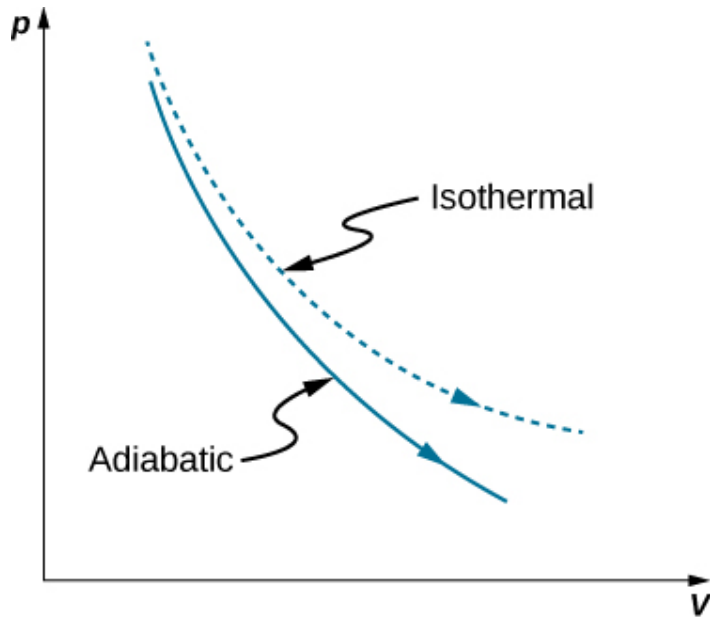
Equation:

$$TV^{\gamma-1} = \text{constant}.$$

A reversible adiabatic expansion of an ideal gas is represented on the pV diagram of [\[link\]](#). The slope of the curve at any point is

Equation:

$$\frac{dp}{dV} = \frac{d}{dV} \left(\frac{\text{constant}}{V^{\gamma}} \right) = -\gamma \frac{p}{V}.$$



Quasi-static adiabatic and isothermal expansions of an ideal gas.

The dashed curve shown on this pV diagram represents an isothermal expansion where T (and therefore pV) is constant. The slope of this curve is useful when we consider the second law of thermodynamics in the next chapter. This slope is

Equation:

$$\frac{dp}{dV} = \frac{d}{dV} \frac{nRT}{V} = -\frac{p}{V}.$$

Because $\gamma > 1$, the isothermal curve is not as steep as that for the adiabatic expansion.

Example:**Compression of an Ideal Gas in an Automobile Engine**

Gasoline vapor is injected into the cylinder of an automobile engine when the piston is in its expanded position. The temperature, pressure, and volume of the resulting gas-air mixture are 20°C , $1.00 \times 10^5 \text{ N/m}^2$, and 240 cm^3 , respectively. The mixture is then compressed adiabatically to a volume of 40 cm^3 . Note that in the actual operation of an automobile engine, the compression is not quasi-static, although we are making that assumption here.

(a) What are the pressure and temperature of the mixture after the compression?

(b) How much work is done by the mixture during the compression?

Strategy

Because we are modeling the process as a quasi-static adiabatic compression of an ideal gas, we have $pV^\gamma = \text{constant}$ and $pV = nRT$. The work needed can

then be evaluated with $W = \int_{V_1}^{V_2} p dV$.

Solution

a. For an adiabatic compression we have

Equation:

$$p_2 = p_1 \left(\frac{V_1}{V_2} \right)^\gamma,$$

so after the compression, the pressure of the mixture is

Equation:

$$p_2 = (1.00 \times 10^5 \text{ N/m}^2) \left(\frac{240 \times 10^{-6} \text{ m}^3}{40 \times 10^{-6} \text{ m}^3} \right)^{1.40} = 1.23 \times 10^6 \text{ N/m}^2.$$

From the ideal gas law, the temperature of the mixture after the compression is

Equation:

$$\begin{aligned}
 T_2 &= \left(\frac{p_2 V_2}{p_1 V_1} \right) T_1 \\
 &= \frac{(1.23 \times 10^6 \text{ N/m}^2)(40 \times 10^{-6} \text{ m}^3)}{(1.00 \times 10^5 \text{ N/m}^2)(240 \times 10^{-6} \text{ m}^3)} \cdot 293 \text{ K} \\
 &= 600 \text{ K} = 328^\circ \text{C}.
 \end{aligned}$$

b. The work done by the mixture during the compression is

Equation:

$$W = \int_{V_1}^{V_2} p dV.$$

With the adiabatic condition of [\[link\]](#), we may write p as K/V^γ , where $K = p_1 V_1^\gamma = p_2 V_2^\gamma$. The work is therefore

Equation:

$$\begin{aligned}
 W &= \int_{V_1}^{V_2} \frac{K}{V^\gamma} dV \\
 &= \frac{K}{1-\gamma} \left(\frac{1}{V_2^{\gamma-1}} - \frac{1}{V_1^{\gamma-1}} \right) \\
 &= \frac{1}{1-\gamma} \left(\frac{p_2 V_2^\gamma}{V_2^{\gamma-1}} - \frac{p_1 V_1^\gamma}{V_1^{\gamma-1}} \right) \\
 &= \frac{1}{1-\gamma} (p_2 V_2 - p_1 V_1) \\
 &= \frac{1}{1-1.40} [(1.23 \times 10^6 \text{ N/m}^2)(40 \times 10^{-6} \text{ m}^3) \\
 &\quad - (1.00 \times 10^5 \text{ N/m}^2)(240 \times 10^{-6} \text{ m}^3)] \\
 &= -63 \text{ J}.
 \end{aligned}$$

Significance

The negative sign on the work done indicates that the piston does work on the gas-air mixture. The engine would not work if the gas-air mixture did work on the piston.

Summary

- A quasi-static adiabatic expansion of an ideal gas produces a steeper pV curve than that of the corresponding isotherm.
- A realistic expansion can be adiabatic but rarely quasi-static.

Key Equations

Equation of state for a closed system	$f(p, V, T) = 0$
Net work for a finite change in volume	$W = \int_{V_1}^{V_2} p dV$
Internal energy of a system (average total energy)	$E_{\text{int}} = \sum_i (K_i + U_i),$
Internal energy of a monatomic ideal gas	$E_{\text{int}} = nN_A \left(\frac{3}{2} k_B T \right) = \frac{3}{2} nRT$
First law of thermodynamics	$\Delta E_{\text{int}} = Q - W$
Molar heat capacity at constant pressure	$C_p = C_V + R$
Ratio of molar heat capacities	$\gamma = C_p / C_V$
Condition for an ideal gas in a quasi-static adiabatic process	$pV^\gamma = \text{constant}$

Conceptual Questions

Exercise:

Problem: Is it possible for γ to be smaller than unity?

Solution:

No, it is always greater than 1.

Exercise:

Problem: Would you expect γ to be larger for a gas or a solid? Explain.

Exercise:**Problem:**

There is no change in the internal energy of an ideal gas undergoing an isothermal process since the internal energy depends only on the temperature. Is it therefore correct to say that an isothermal process is the same as an adiabatic process for an ideal gas? Explain your answer.

Solution:

An adiabatic process has a change in temperature but no heat flow. The isothermal process has no change in temperature but has heat flow.

Exercise:**Problem:**

Does a gas do any work when it expands adiabatically? If so, what is the source of the energy needed to do this work?

Problems**Exercise:****Problem:**

A monatomic ideal gas undergoes a quasi-static adiabatic expansion in which its volume is doubled. How is the pressure of the gas changed?

Solution:

pressure decreased by 0.31 times the original pressure

Exercise:**Problem:**

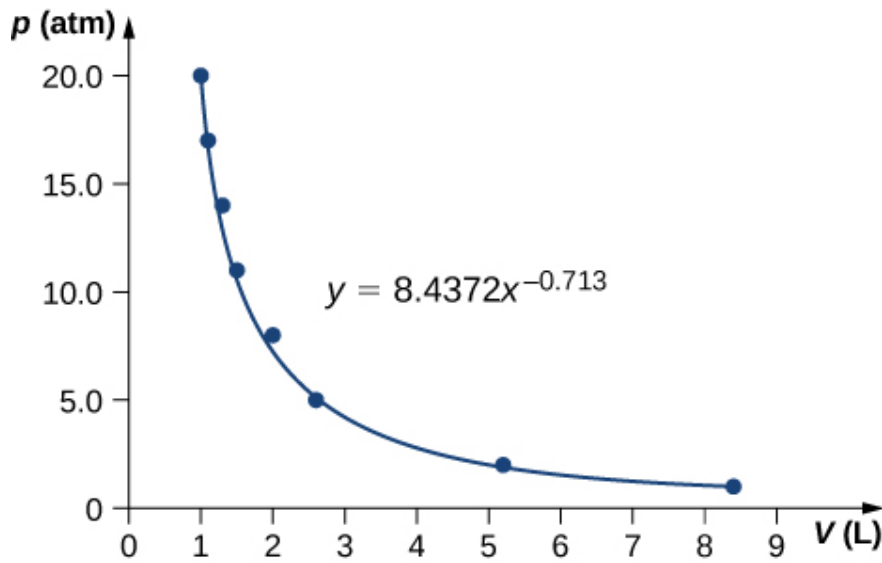
An ideal gas has a pressure of 0.50 atm and a volume of 10 L. It is compressed adiabatically and quasi-statically until its pressure is 3.0 atm and its volume is 2.8 L. Is the gas monatomic, diatomic, or polyatomic?

Exercise:**Problem:**

Pressure and volume measurements of a dilute gas undergoing a quasi-static adiabatic expansion are shown below. Plot $\ln p$ vs. V and determine γ for this gas from your graph.

P (atm)	V (L)
20.0	1.0
17.0	1.1
14.0	1.3
11.0	1.5
8.0	2.0
5.0	2.6
2.0	5.2
1.0	8.4

Solution:



;

$$\gamma = 0.713$$

Exercise:

Problem:

An ideal monatomic gas at 300 K expands adiabatically and reversibly to twice its volume. What is its final temperature?

Exercise:

Problem:

An ideal diatomic gas at 80 K is slowly compressed adiabatically and reversibly to half its volume. What is its final temperature?

Solution:

106 K

Exercise:

Problem:

An ideal diatomic gas at 80 K is slowly compressed adiabatically to one-third its original volume. What is its final temperature?

Exercise:**Problem:**

Compare the change in internal energy of an ideal gas for a quasi-static adiabatic expansion with that for a quasi-static isothermal expansion. What happens to the temperature of an ideal gas in an adiabatic expansion?

Solution:

An adiabatic expansion has less work done and no heat flow, thereby a lower internal energy comparing to an isothermal expansion which has both heat flow and work done. Temperature decreases during adiabatic expansion.

Exercise:**Problem:**

The temperature of n moles of an ideal gas changes from T_1 to T_2 in a quasi-static adiabatic transition. Show that the work done by the gas is given by

$$W = \frac{nR}{\gamma-1}(T_1 - T_2).$$

Exercise:**Problem:**

A dilute gas expands quasi-statically to three times its initial volume. Is the final gas pressure greater for an isothermal or an adiabatic expansion? Does your answer depend on whether the gas is monatomic, diatomic, or polyatomic?

Solution:

Isothermal has a greater final pressure and does not depend on the type of gas.

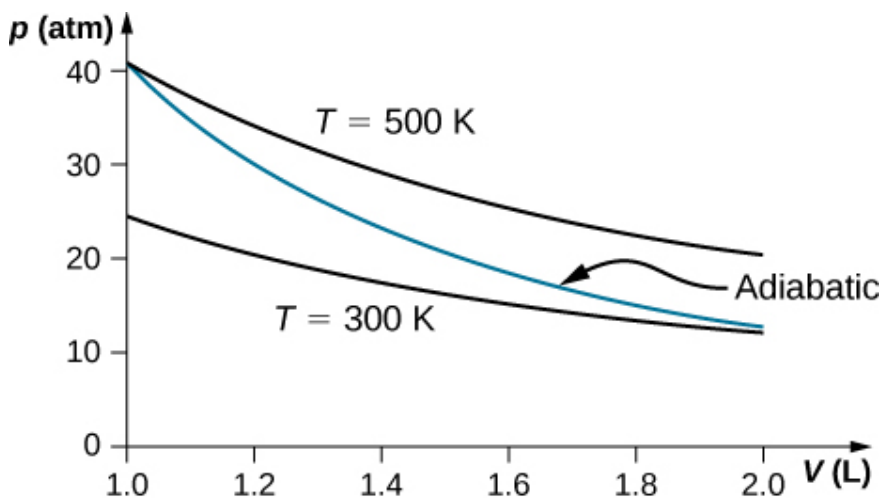
Exercise:

Problem:

(a) An ideal gas expands adiabatically from a volume of $2.0 \times 10^{-3} \text{ m}^3$ to $2.5 \times 10^{-3} \text{ m}^3$. If the initial pressure and temperature were $5.0 \times 10^5 \text{ Pa}$ and 300 K, respectively, what are the final pressure and temperature of the gas? Use $\gamma = 5/3$ for the gas. (b) In an isothermal process, an ideal gas expands from a volume of $2.0 \times 10^{-3} \text{ m}^3$ to $2.5 \times 10^{-3} \text{ m}^3$. If the initial pressure and temperature were $5.0 \times 10^5 \text{ Pa}$ and 300 K, respectively, what are the final pressure and temperature of the gas?

Exercise:**Problem:**

On an adiabatic process of an ideal gas pressure, volume and temperature change such that pV^γ is constant with $\gamma = 5/3$ for monatomic gas such as helium and $\gamma = 7/5$ for diatomic gas such as hydrogen at room temperature. Use numerical values to plot two isotherms of 1 mol of helium gas using ideal gas law and two adiabatic processes mediating between them. Use $T_1 = 500 \text{ K}$, $V_1 = 1 \text{ L}$, and $T_2 = 300 \text{ K}$ for your plot.

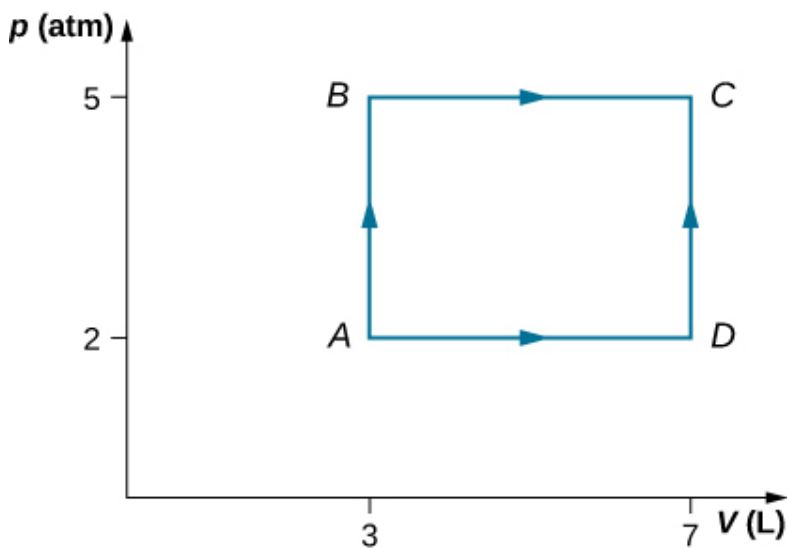
Solution:**Exercise:**

Problem:

Two moles of a monatomic ideal gas such as helium is compressed adiabatically and reversibly from a state (3 atm, 5 L) to a state with pressure 4 atm. (a) Find the volume and temperature of the final state. (b) Find the temperature of the initial state of the gas. (c) Find the work done by the gas in the process. (d) Find the change in internal energy of the gas in the process.

Additional Problems**Exercise:****Problem:**

Consider the process shown below. During steps AB and BC , 3600 J and 2400 J of heat, respectively, are added to the system. (a) Find the work done in each of the processes AB , BC , AD , and DC . (b) Find the internal energy change in processes AB and BC . (c) Find the internal energy difference between states C and A . (d) Find the total heat added in the ADC process. (e) From the information given, can you find the heat added in process AD ? Why or why not?



Solution:

a. $W_{AB} = 0$, $W_{BC} = 2026 \text{ J}$, $W_{AD} = 810.4 \text{ J}$, $W_{DC} = 0$; b. $\Delta E_{AB} = 3600 \text{ J}$, $\Delta E_{BC} = 374 \text{ J}$; c. $\Delta E_{AC} = 3974 \text{ J}$; d. $Q_{ADC} = 4784 \text{ J}$; e. No, because heat was added for both parts AD and DC . There is not enough information to figure out how much is from each segment of the path.

Exercise:

Problem:

A car tire contains 0.0380 m^3 of air at a pressure of $2.20 \times 10^5 \text{ Pa}$ (about 32 psi). How much more internal energy does this gas have than the same volume has at zero gauge pressure (which is equivalent to normal atmospheric pressure)?

Exercise:

Problem:

A helium-filled toy balloon has a gauge pressure of 0.200 atm and a volume of 10.0 L. How much greater is the internal energy of the helium in the balloon than it would be at zero gauge pressure?

Solution:

300 J

Exercise:

Problem:

Steam to drive an old-fashioned steam locomotive is supplied at a constant gauge pressure of $1.75 \times 10^6 \text{ N/m}^2$ (about 250 psi) to a piston with a 0.200-m radius. (a) By calculating $p\Delta V$, find the work done by the steam when the piston moves 0.800 m. Note that this is the net work output, since gauge pressure is used. (b) Now find the amount of work by calculating the force exerted times the distance traveled. Is the answer the same as in part (a)?

Exercise:

Problem:

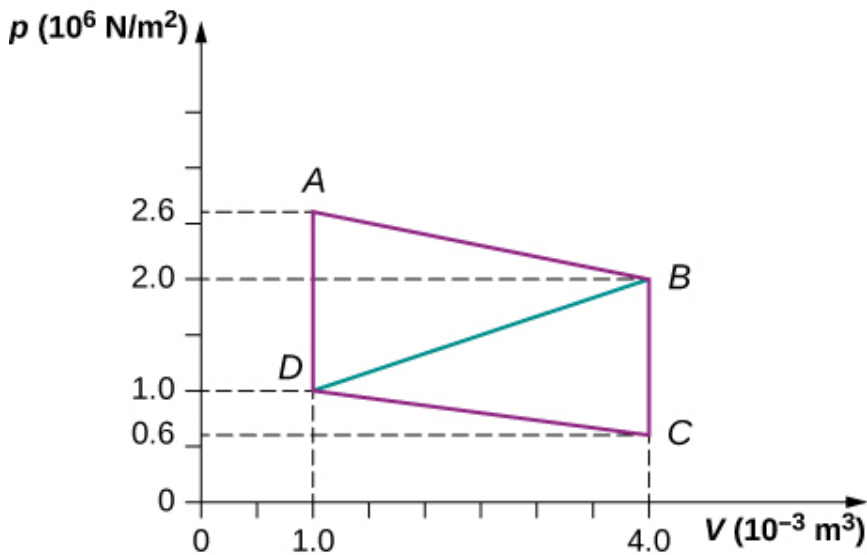
A hand-driven tire pump has a piston with a 2.50-cm diameter and a maximum stroke of 30.0 cm. (a) How much work do you do in one stroke if the average gauge pressure is $2.4 \times 10^5 \text{ N/m}^2$ (about 35 psi)? (b) What average force do you exert on the piston, neglecting friction and gravitational force?

Solution:

a. 59.5 J; b. 170 N

Exercise:**Problem:**

Calculate the net work output of a heat engine following path *ABCD*A as shown below.

**Exercise:****Problem:**

What is the net work output of a heat engine that follows path *ABDA* in the preceding problem with a straight line from *B* to *D*? Why is the work output less than for path *ABCD*A?

Solution:

$$2.4 \times 10^3 \text{ J}$$

Exercise:**Problem:**

Five moles of a monatomic ideal gas in a cylinder at 27°C is expanded isothermally from a volume of 5 L to 10 L. (a) What is the change in internal energy? (b) How much work was done on the gas in the process? (c) How much heat was transferred to the gas?

Exercise:**Problem:**

Four moles of a monatomic ideal gas in a cylinder at 27°C is expanded at constant pressure equal to 1 atm until its volume doubles. (a) What is the change in internal energy? (b) How much work was done by the gas in the process? (c) How much heat was transferred to the gas?

Solution:

a. 15,000 J; b. 10,000 J; c. 25,000 J

Exercise:**Problem:**

Helium gas is cooled from 20°C to 10°C by expanding from 40 atm to 1 atm. If there is 1.4 mol of helium, (a) What is the final volume of helium? (b) What is the change in internal energy?

Exercise:**Problem:**

In an adiabatic process, oxygen gas in a container is compressed along a path that can be described by the following pressure in atm as a function of volume V , with $V_0 = 1\text{ L}$: $p = (3.0 \text{ atm})(V/V_0)^{-1.2}$. The initial and final volumes during the process were 2 L and 1.5 L, respectively. Find the amount of work done on the gas.

Solution:

78 J

Exercise:**Problem:**

A cylinder containing three moles of a monatomic ideal gas is heated at a constant pressure of 2 atm. The temperature of the gas changes from 300 K to 350 K as a result of the expansion. Find work done (a) on the gas; and (b) by the gas.

Exercise:**Problem:**

A cylinder containing three moles of nitrogen gas is heated at a constant pressure of 2 atm. The temperature of the gas changes from 300 K to 350 K as a result of the expansion. Find work done (a) on the gas, and (b) by the gas by using van der Waals equation of state instead of ideal gas law.

Solution:

A cylinder containing three moles of nitrogen gas is heated at a constant pressure of 2 atm. a. -1220 J; b. $+1220$ J

Exercise:**Problem:**

Two moles of a monatomic ideal gas such as helium is compressed adiabatically and reversibly from a state (3 atm, 5 L) to a state with a pressure of 4 atm. (a) Find the volume and temperature of the final state. (b) Find the temperature of the initial state. (c) Find work done by the gas in the process. (d) Find the change in internal energy in the process. Assume $C_V = 5R$ and $C_p = C_V + R$ for the diatomic ideal gas in the conditions given.

Exercise:

Problem:

An insulated vessel contains 1.5 moles of argon at 2 atm. The gas initially occupies a volume of 5 L. As a result of the adiabatic expansion the pressure of the gas is reduced to 1 atm. (a) Find the volume and temperature of the final state. (b) Find the temperature of the gas in the initial state. (c) Find the work done by the gas in the process. (d) Find the change in the internal energy of the gas in the process.

Solution:

a. 7.6 L, 61.6 K; b. 81.3 K; c. $3.63 \text{ L} \cdot \text{atm} = 367 \text{ J}$; d. -367 J

Challenge Problems**Exercise:****Problem:**

One mole of an ideal monatomic gas occupies a volume of $1.0 \times 10^{-2} \text{ m}^3$ at a pressure of $2.0 \times 10^5 \text{ N/m}^2$. (a) What is the temperature of the gas? (b) The gas undergoes a quasi-static adiabatic compression until its volume is decreased to $5.0 \times 10^{-3} \text{ m}^3$. What is the new gas temperature? (c) How much work is done on the gas during the compression? (d) What is the change in the internal energy of the gas?

Exercise:**Problem:**

One mole of an ideal gas is initially in a chamber of volume $1.0 \times 10^{-2} \text{ m}^3$ and at a temperature of 27°C . (a) How much heat is absorbed by the gas when it slowly expands isothermally to twice its initial volume? (b) Suppose the gas is slowly transformed to the same final state by first decreasing the pressure at constant volume and then expanding it isobarically. What is the heat transferred for this case? (c) Calculate the heat transferred when the gas is transformed quasi-statically to the same final state by expanding it isobarically, then decreasing its pressure at constant volume.

Solution:

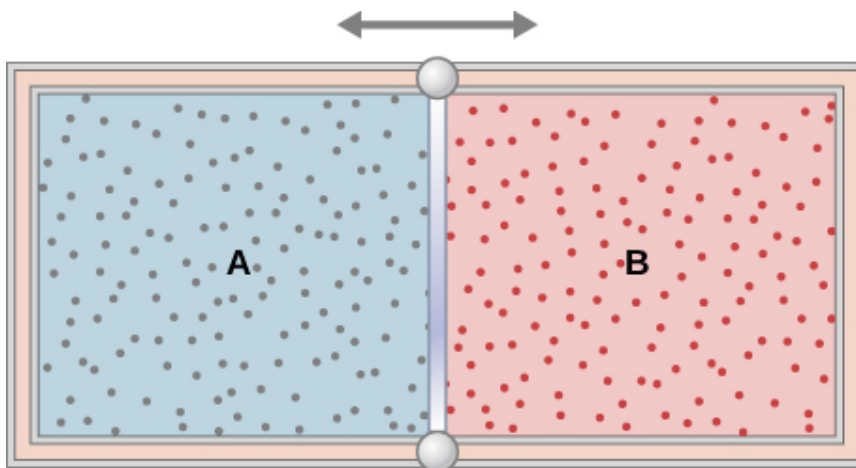
a. 1700 J; b. 1200 J; c. 2400 J

Exercise:**Problem:**

A bullet of mass 10 g is traveling horizontally at 200 m/s when it strikes and embeds in a pendulum bob of mass 2.0 kg. (a) How much mechanical energy is dissipated in the collision? (b) Assuming that C_v for the bob plus bullet is $3R$, calculate the temperature increase of the system due to the collision. Take the molecular mass of the system to be 200 g/mol.

Exercise:**Problem:**

The insulated cylinder shown below is closed at both ends and contains an insulating piston that is free to move on frictionless bearings. The piston divides the chamber into two compartments containing gases A and B. Originally, each compartment has a volume of $5.0 \times 10^{-2} \text{ m}^3$ and contains a monatomic ideal gas at a temperature of 0°C and a pressure of 1.0 atm. (a) How many moles of gas are in each compartment? (b) Heat Q is slowly added to A so that it expands and B is compressed until the pressure of both gases is 3.0 atm. Use the fact that the compression of B is adiabatic to determine the final volume of both gases. (c) What are their final temperatures? (d) What is the value of Q ?



Solution:

a. 2.2 mol; b. $V_A = 2.6 \times 10^{-2} \text{ m}^3$, $V_B = 7.4 \times 10^{-2} \text{ m}^3$; c. $T_A = 1220 \text{ K}$, $T_B = 430 \text{ K}$; d. 30,500 J

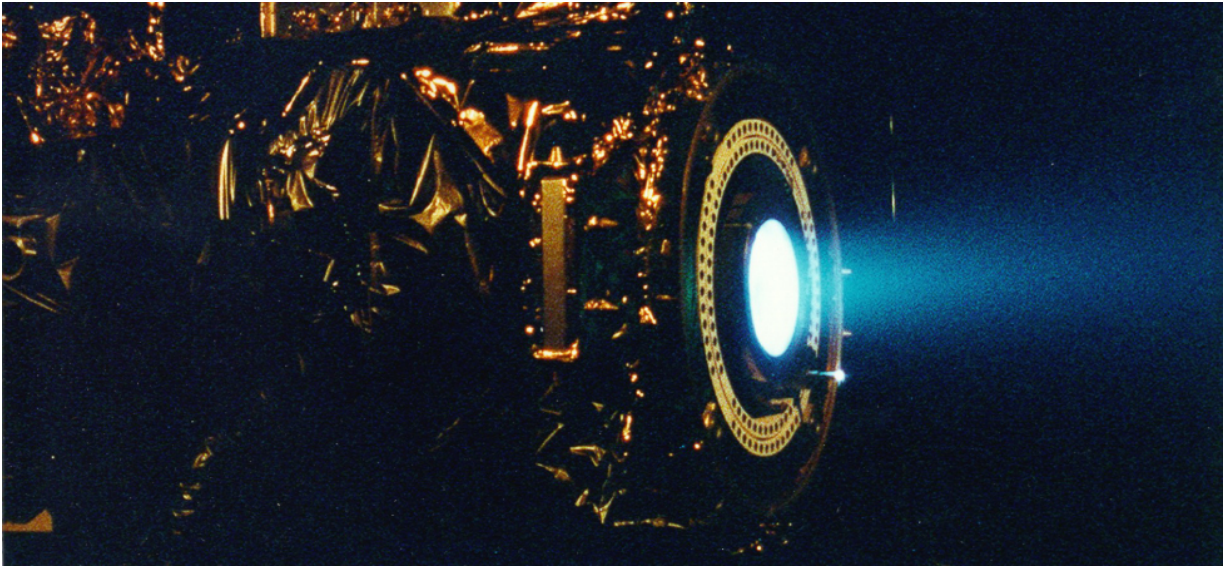
Exercise:**Problem:**

In a diesel engine, the fuel is ignited without a spark plug. Instead, air in a cylinder is compressed adiabatically to a temperature above the ignition temperature of the fuel; at the point of maximum compression, the fuel is injected into the cylinder. Suppose that air at 20°C is taken into the cylinder at a volume V_1 and then compressed adiabatically and quasi-statically to a temperature of 600°C and a volume V_2 . If $\gamma = 1.4$, what is the ratio V_1/V_2 ? (Note: In an operating diesel engine, the compression is not quasi-static.)

Introduction

class="introduction"

A xenon ion engine from the Jet Propulsion Laboratory shows the faint blue glow of charged atoms emitted from the engine. The ion propulsion engine is the first nonchemical propulsion to be used as the primary means of propelling a spacecraft.
(credit: modification of work by NASA/JPL)



According to the first law of thermodynamics, the only processes that can occur are those that conserve energy. But this cannot be the only restriction imposed by nature, because many seemingly possible thermodynamic processes that would conserve energy do not occur. For example, when two bodies are in thermal contact, heat never flows from the colder body to the warmer one, even though this is not forbidden by the first law. So some other thermodynamic principles must be controlling the behavior of physical systems.

One such principle is the *second law of thermodynamics*, which limits the use of energy within a source. Energy cannot arbitrarily pass from one object to another, just as we cannot transfer heat from a cold object to a hot one without doing any work. We cannot unmix cream from coffee without a chemical process that changes the physical characteristics of the system or its environment. We cannot use internal energy stored in the air to propel a car, or use the energy of the ocean to run a ship, without disturbing something around that object.

In the chapter covering the first law of thermodynamics, we started our discussion with a joke by C. P. Snow stating that the first law means “you can’t win.” He paraphrased the second law as “you can’t break even, except on a very cold day.” Unless you are at zero kelvin, you cannot convert 100% of thermal energy into work. We start by discussing spontaneous

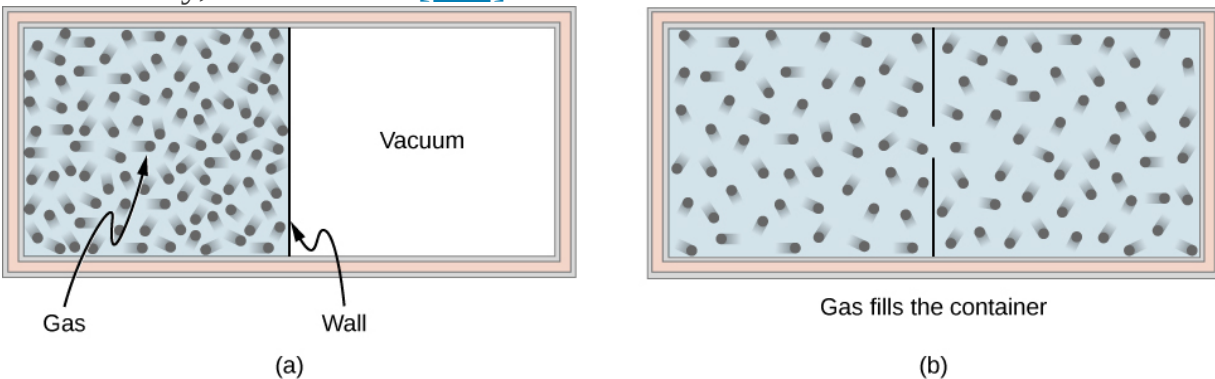
processes and explain why some processes require work to occur even if energy would have been conserved.

Reversible and Irreversible Processes

By the end of this section, you will be able to:

- Define reversible and irreversible processes
- State the second law of thermodynamics via an irreversible process

Consider an ideal gas that is held in half of a thermally insulated container by a wall in the middle of the container. The other half of the container is under vacuum with no molecules inside. Now, if we remove the wall in the middle quickly, the gas expands and fills up the entire container immediately, as shown in [\[link\]](#).



A gas expanding from half of a container to the entire container (a) before and (b) after the wall in the middle is removed.

Because half of the container is under vacuum before the gas expands there, we do not expect any work to be done by the system—that is, $W = 0$ —because no force from the vacuum is exerted on the gas during the expansion. If the container is thermally insulated from the rest of the environment, we do not expect any heat transfer to the system either, so $Q = 0$. Then the first law of thermodynamics leads to the change of the internal energy of the system,

Equation:

$$\Delta E_{\text{int}} = Q - W = 0.$$

For an ideal gas, if the internal energy doesn't change, then the temperature stays the same. Thus, the equation of state of the ideal gas gives us the final pressure of the gas, $p = nRT/V = p_0/2$, where p_0 is the pressure of the gas before the expansion. The volume is doubled and the pressure is halved, but nothing else seems to have changed during the expansion.

All of this discussion is based on what we have learned so far and makes sense. Here is what puzzles us: Can all the molecules go backward to the original half of the container in some future time? Our intuition tells us that this is going to be very unlikely, even though nothing we have learned so far prevents such an event from happening, regardless of how small the probability is. What we are really asking is whether the expansion into the vacuum half of the container is *reversible*.

A **reversible process** is a process in which the system and environment can be restored to exactly the same initial states that they were in before the process occurred, if we go backward along the path of the process. The necessary condition for a reversible process is therefore the quasi-static requirement. Note that it is quite easy to restore a system to its original state; the hard part is to have its environment restored to its original state at the same time. For example, in the example of an ideal gas expanding into vacuum to twice its original volume, we can easily push it back with a piston and restore its temperature and pressure by removing some heat from the gas. The problem is that we cannot do it without changing something in its surroundings, such as dumping some heat there.

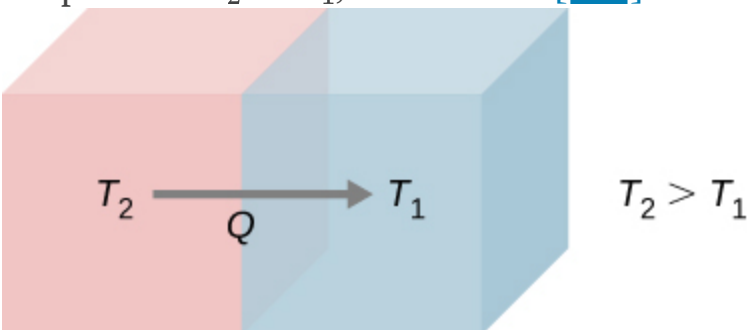
A reversible process is truly an ideal process that rarely happens. We can make certain processes close to reversible and therefore use the consequences of the corresponding reversible processes as a starting point or reference. In reality, almost all processes are irreversible, and some properties of the environment are altered when the properties of the system are restored. The expansion of an ideal gas, as we have just outlined, is irreversible because the process is not even quasi-static, that is, not in an equilibrium state at any moment of the expansion.

From the microscopic point of view, a particle described by Newton's second law can go backward if we flip the direction of time. But this is not

the case, in practical terms, in a macroscopic system with more than 10^{23} particles or molecules, where numerous collisions between these molecules tend to erase any trace of memory of the initial trajectory of each of the particles. For example, we can actually estimate the chance for all the particles in the expanded gas to go back to the original half of the container, but the current age of the universe is still not long enough for it to happen even once.

An **irreversible process** is what we encounter in reality almost all the time. The system and its environment cannot be restored to their original states at the same time. Because this is what happens in nature, it is also called a natural process. The sign of an irreversible process comes from the finite gradient between the states occurring in the actual process. For example, when heat flows from one object to another, there is a finite temperature difference (gradient) between the two objects. More importantly, at any given moment of the process, the system most likely is not at equilibrium or in a well-defined state. This phenomenon is called **irreversibility**.

Let us see another example of irreversibility in thermal processes. Consider two objects in thermal contact: one at temperature T_1 and the other at temperature $T_2 > T_1$, as shown in [\[link\]](#).



Spontaneous heat flow from an object at higher temperature T_2 to another at lower temperature T_1 .

We know from common personal experience that heat flows from a hotter object to a colder one. For example, when we hold a few pieces of ice in

our hands, we feel cold because heat has left our hands into the ice. The opposite is true when we hold one end of a metal rod while keeping the other end over a fire. Based on all of the experiments that have been done on spontaneous heat transfer, the following statement summarizes the governing principle:

Note:

Second Law of Thermodynamics (Clausius statement)

Heat never flows spontaneously from a colder object to a hotter object.

This statement turns out to be one of several different ways of stating the second law of thermodynamics. The form of this statement is credited to German physicist Rudolf Clausius (1822–1888) and is referred to as the **Clausius statement of the second law of thermodynamics**. The word “spontaneously” here means no other effort has been made by a third party, or one that is neither the hotter nor colder object. We will introduce some other major statements of the second law and show that they imply each other. In fact, all the different statements of the second law of thermodynamics can be shown to be equivalent, and all lead to the irreversibility of spontaneous heat flow between macroscopic objects of a very large number of molecules or particles.

Both isothermal and adiabatic processes sketched on a pV graph (discussed in [The First Law of Thermodynamics](#)) are reversible in principle because the system is always at an equilibrium state at any point of the processes and can go forward or backward along the given curves. Other idealized processes can be represented by pV curves; [\[link\]](#) summarizes the most common reversible processes.

Process	Constant Quantity and Resulting Fact
Isobaric	Constant pressure $W = p\Delta V$
Isochoric	Constant volume $W = 0$
Isothermal	Constant temperature $\Delta T = 0$
Adiabatic	No heat transfer $Q = 0$

Summary of Simple Thermodynamic Processes

Summary

- A reversible process is one in which both the system and its environment can return to exactly the states they were in by following the reverse path.
- An irreversible process is one in which the system and its environment cannot return together to exactly the states that they were in.
- The irreversibility of any natural process results from the second law of thermodynamics.

Conceptual Questions

Exercise:

Problem:

State an example of a process that occurs in nature that is as close to reversible as it can be.

Solution:

Some possible solutions are frictionless movement; restrained compression or expansion; energy transfer as heat due to infinitesimal temperature nonuniformity; electric current flow through a zero

resistance; restrained chemical reaction; and mixing of two samples of the same substance at the same state.

Problems

Exercise:

Problem:

A tank contains 111.0 g chlorine gas (Cl_2), which is at temperature 82.0°C and absolute pressure $5.70 \times 10^5 \text{ Pa}$. The temperature of the air outside the tank is 20.0°C . The molar mass of Cl_2 is 70.9 g/mol. (a) What is the volume of the tank? (b) What is the internal energy of the gas? (c) What is the work done by the gas if the temperature and pressure inside the tank drop to 31.0°C and $3.80 \times 10^5 \text{ Pa}$, respectively, due to a leak?

Exercise:

Problem:

A mole of ideal monatomic gas at 0°C and 1.00 atm is warmed up to expand isobarically to triple its volume. How much heat is transferred during the process?

Solution:

$$11.0 \times 10^3 \text{ J}$$

Exercise:

Problem:

A mole of an ideal gas at pressure 4.00 atm and temperature 298 K expands isothermally to double its volume. What is the work done by the gas?

Exercise:

Problem:

After a free expansion to quadruple its volume, a mole of ideal diatomic gas is compressed back to its original volume adiabatically and then cooled down to its original temperature. What is the minimum heat removed from the gas in the final step to restoring its state?

Solution:

$$4.5 pV_0$$

Glossary

Clausius statement of the second law of thermodynamics

heat never flows spontaneously from a colder object to a hotter object

irreversibility

phenomenon associated with a natural process

irreversible process

process in which neither the system nor its environment can be restored to their original states at the same time

reversible process

process in which both the system and the external environment theoretically can be returned to their original states

Heat Engines

By the end of this section, you will be able to:

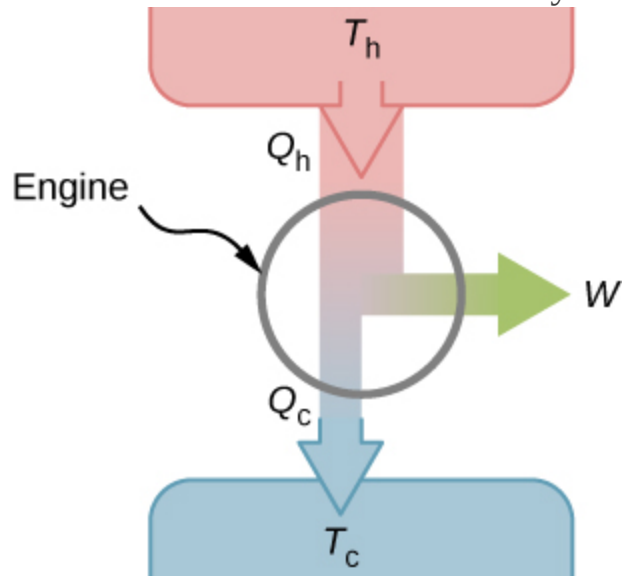
- Describe the function and components of a heat engine
- Explain the efficiency of an engine
- Calculate the efficiency of an engine for a given cycle of an ideal gas

A **heat engine** is a device used to extract heat from a source and then convert it into mechanical work that is used for all sorts of applications. For example, a steam engine on an old-style train can produce the work needed for driving the train. Several questions emerge from the construction and application of heat engines. For example, what is the maximum percentage of the heat extracted that can be used to do work? This turns out to be a question that can only be answered through the second law of thermodynamics.

The second law of thermodynamics can be formally stated in several ways. One statement presented so far is about the direction of spontaneous heat flow, known as the Clausius statement. A couple of other statements are based on heat engines. *Whenever we consider heat engines and associated devices such as refrigerators and heat pumps, we do not use the normal sign convention for heat and work.* For convenience, we assume that the symbols Q_h , Q_c , and W represent only the amounts of heat transferred and work delivered, regardless what the givers or receivers are. Whether heat is entering or leaving a system and work is done to or by a system are indicated by proper signs in front of the symbols and by the directions of arrows in diagrams.

It turns out that we need more than one heat source/sink to construct a heat engine. We will come back to this point later in the chapter, when we compare different statements of the second law of thermodynamics. For the moment, we assume that a heat engine is constructed between a heat source (high-temperature reservoir or hot reservoir) and a heat sink (low-temperature reservoir or cold reservoir), represented schematically in [\[link\]](#). The engine absorbs heat Q_h from a heat source (**hot reservoir**) of Kelvin temperature T_h , uses some of that energy to produce useful work W , and then discards the remaining energy as heat Q_c into a heat sink (**cold**

reservoir) of Kelvin temperature T_c . Power plants and internal combustion engines are examples of heat engines. Power plants use steam produced at high temperature to drive electric generators, while exhausting heat to the atmosphere or a nearby body of water in the role of the heat sink. In an internal combustion engine, a hot gas-air mixture is used to push a piston, and heat is exhausted to the nearby atmosphere in a similar manner.



A schematic representation of a heat engine. Energy flows from the hot reservoir to the cold reservoir while doing work.

Actual heat engines have many different designs. Examples include internal combustion engines, such as those used in most cars today, and external combustion engines, such as the steam engines used in old steam-engine trains. [\[link\]](#) shows a photo of a nuclear power plant in operation. The atmosphere around the reactors acts as the cold reservoir, and the heat generated from the nuclear reaction provides the heat from the hot reservoir.



The heat exhausted from a nuclear power plant goes to the cooling towers, where it is released into the atmosphere.

Heat engines operate by carrying a *working substance* through a cycle. In a steam power plant, the working substance is water, which starts as a liquid, becomes vaporized, is then used to drive a turbine, and is finally condensed back into the liquid state. As is the case for all working substances in cyclic processes, once the water returns to its initial state, it repeats the same sequence.

For now, we assume that the cycles of heat engines are reversible, so there is no energy loss to friction or other irreversible effects. Suppose that the engine of [\[link\]](#) goes through one complete cycle and that Q_h , Q_c , and W represent the heats exchanged and the work done for that cycle. Since the initial and final states of the system are the same, $\Delta E_{\text{int}} = 0$ for the cycle. We therefore have from the first law of thermodynamics,

Equation:

$$W = Q - \Delta E_{\text{int}} = (Q_h - Q_c) - 0,$$

so that

Note:

Equation:

$$W = Q_h - Q_c.$$

The most important measure of a heat engine is its **efficiency (e)**, which is simply “what we get out” divided by “what we put in” during each cycle, as defined by $e = W_{\text{out}}/Q_{\text{in}}$.

With a heat engine working between two heat reservoirs, we get out W and put in Q_h , so the efficiency of the engine is

Note:

Equation:

$$e = \frac{W}{Q_h} = 1 - \frac{Q_c}{Q_h}.$$

Here, we used [\[link\]](#), $W = Q_h - Q_c$, in the final step of this expression for the efficiency.

Example:

A Lawn Mower

A lawn mower is rated to have an efficiency of 25.0% and an average power of 3.00 kW. What are (a) the average work and (b) the minimum heat discharge into the air by the lawn mower in one minute of use?

Strategy

From the average power—that is, the rate of work production—we can figure out the work done in the given elapsed time. Then, from the

efficiency given, we can figure out the minimum heat discharge $Q_c = Q_h(1 - e)$ with $Q_h = Q_c + W$.

Solution

- a. The average work delivered by the lawn mower is

Equation:

$$W = P\Delta t = 3.00 \times 10^3 \times 60 \times 1.00 \text{ J} = 180 \text{ kJ}.$$

- b. The minimum heat discharged into the air is given by

Equation:

$$Q_c = Q_h(1 - e) = (Q_c + W)(1 - e),$$

which leads to

Equation:

$$Q_c = W(1/e - 1) = 180 \times (1/0.25 - 1) \text{ kJ} = 540 \text{ kJ}.$$

Significance

As the efficiency rises, the minimum heat discharged falls. This helps our environment and atmosphere by not having as much waste heat expelled.

Summary

- The work done by a heat engine is the difference between the heat absorbed from the hot reservoir and the heat discharged to the cold reservoir, that is, $W = Q_h - Q_c$.
- The ratio of the work done by the engine and the heat absorbed from the hot reservoir provides the efficiency of the engine, that is, $e = W/Q_h = 1 - Q_c/Q_h$.

Conceptual Questions

Exercise:

Problem:

Explain in practical terms why efficiency is defined as W/Q_h .

Problems**Exercise:****Problem:**

An engine is found to have an efficiency of 0.40. If it does 200 J of work per cycle, what are the corresponding quantities of heat absorbed and discharged?

Exercise:**Problem:**

In performing 100.0 J of work, an engine discharges 50.0 J of heat. What is the efficiency of the engine?

Solution:

0.667

Exercise:**Problem:**

An engine with an efficiency of 0.30 absorbs 500 J of heat per cycle. (a) How much work does it perform per cycle? (b) How much heat does it discharge per cycle?

Exercise:**Problem:**

It is found that an engine discharges 100.0 J while absorbing 125.0 J each cycle of operation. (a) What is the efficiency of the engine? (b) How much work does it perform per cycle?

Solution:

a. 0.200; b. 25 J

Exercise:**Problem:**

The temperature of the cold reservoir of the engine is 300 K. It has an efficiency of 0.30 and absorbs 500 J of heat per cycle. (a) How much work does it perform per cycle? (b) How much heat does it discharge per cycle?

Exercise:**Problem:**

An engine absorbs three times as much heat as it discharges. The work done by the engine per cycle is 50 J. Calculate (a) the efficiency of the engine, (b) the heat absorbed per cycle, and (c) the heat discharged per cycle.

Solution:

a. 0.67; b. 75 J; c. 25 J

Exercise:**Problem:**

A coal power plant consumes 100,000 kg of coal per hour and produces 500 MW of power. If the heat of combustion of coal is 30 MJ/kg, what is the efficiency of the power plant?

Glossary

cold reservoir

sink of heat used by a heat engine

efficiency (e)

output work from the engine over the input heat to the engine from the hot reservoir

heat engine
device that converts heat into work

hot reservoir
source of heat used by a heat engine

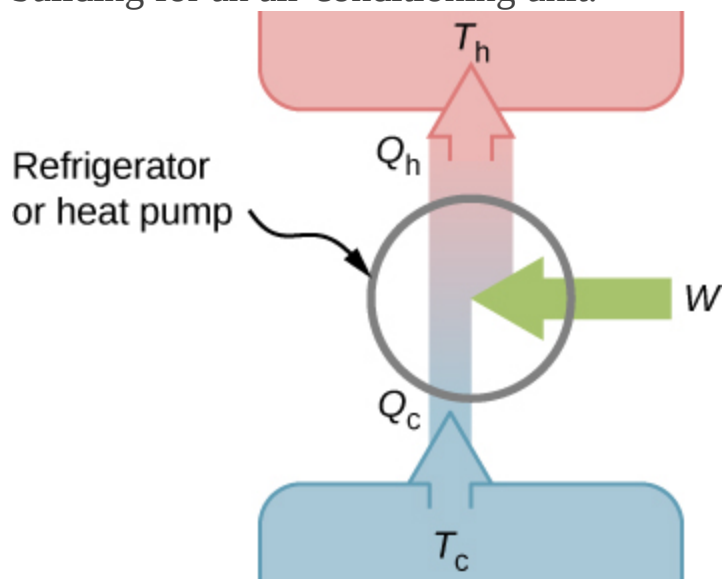
Refrigerators and Heat Pumps

By the end of this section, you will be able to:

- Describe a refrigerator and a heat pump and list their differences
- Calculate the performance coefficients of simple refrigerators and heat pumps

The cycles we used to describe the engine in the preceding section are all reversible, so each sequence of steps can just as easily be performed in the opposite direction. In this case, the engine is known as a refrigerator or a heat pump, depending on what is the focus: the heat removed from the cold reservoir or the heat dumped to the hot reservoir. Either a refrigerator or a heat pump is an engine running in reverse. For a **refrigerator**, the focus is on removing heat from a specific area. For a **heat pump**, the focus is on dumping heat to a specific area.

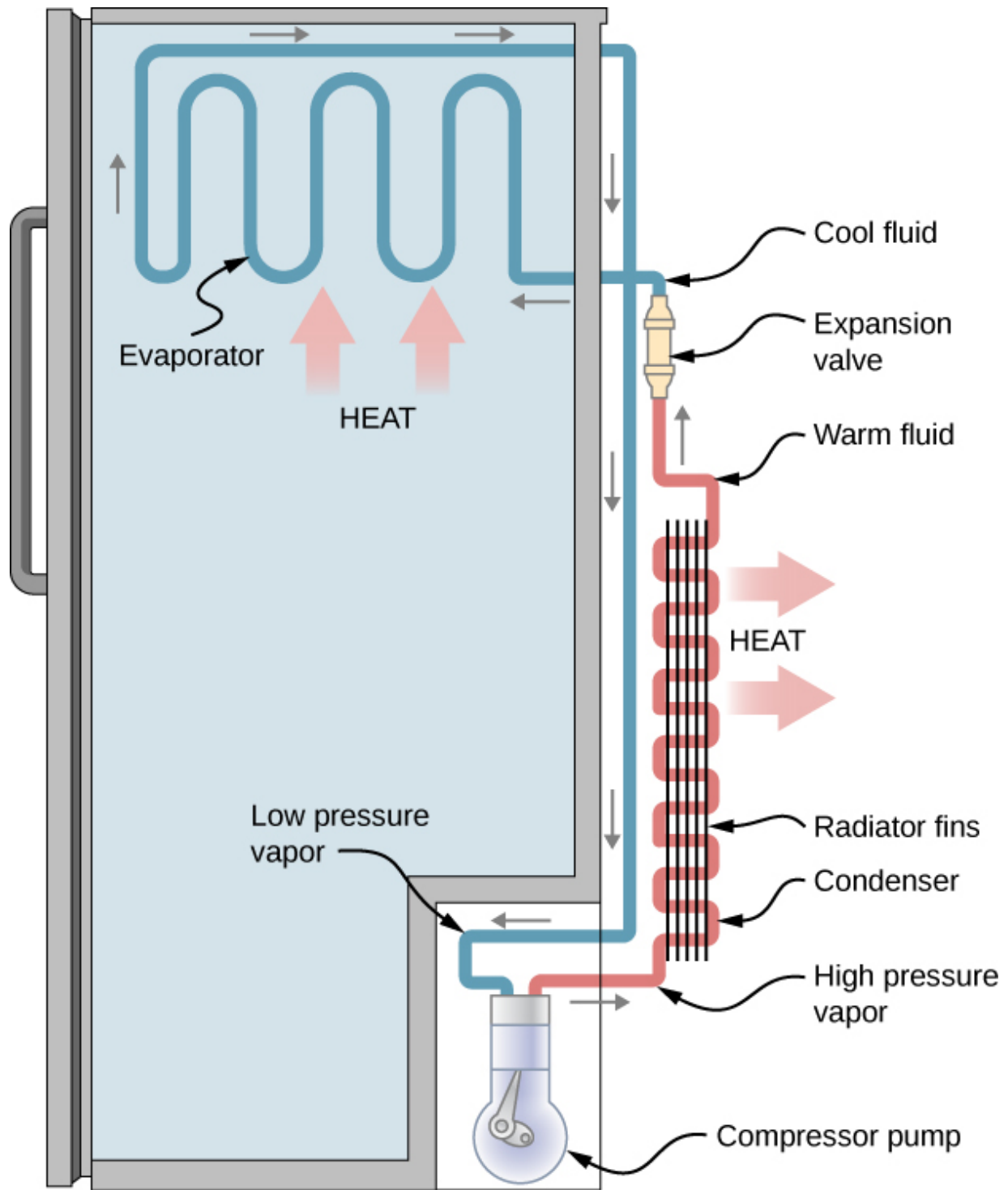
We first consider a refrigerator ([\[link\]](#)). The purpose of this engine is to remove heat from the cold reservoir, which is the space inside the refrigerator for an actual household refrigerator or the space inside a building for an air-conditioning unit.



A schematic representation of a refrigerator (or a heat pump). The

arrow next to work (W) indicates work being put into the system.

A refrigerator (or heat pump) absorbs heat Q_c from the cold reservoir at Kelvin temperature T_c and discards heat Q_h to the hot reservoir at Kelvin temperature T_h , while work W is done on the engine's working substance, as shown by the arrow pointing toward the system in the figure. A household refrigerator removes heat from the food within it while exhausting heat to the surrounding air. The required work, for which we pay in our electricity bill, is performed by the motor that moves a coolant through the coils. A schematic sketch of a household refrigerator is given in [\[link\]](#).



A schematic diagram of a household refrigerator. A coolant with a boiling temperature below the freezing point of water is sent through the cycle (clockwise in this diagram). The coolant extracts heat from the refrigerator at the evaporator, causing coolant to vaporize. It is then

compressed and sent through the condenser, where it exhausts heat to the outside.

The effectiveness or **coefficient of performance** K_R of a refrigerator is measured by the heat removed from the cold reservoir divided by the work done by the working substance cycle by cycle:

Note:

Equation:

$$K_R = \frac{Q_c}{W} = \frac{Q_c}{Q_h - Q_c}.$$

Note that we have used the condition of energy conservation, $W = Q_h - Q_c$, in the final step of this expression.

The effectiveness or coefficient of performance K_P of a heat pump is measured by the heat dumped to the hot reservoir divided by the work done to the engine on the working substance cycle by cycle:

Note:

Equation:

$$K_P = \frac{Q_h}{W} = \frac{Q_h}{Q_h - Q_c}.$$

Once again, we use the energy conservation condition $W = Q_h - Q_c$ to obtain the final step of this expression.

Summary

- A refrigerator or a heat pump is a heat engine run in reverse.
- The focus of a refrigerator is on removing heat from the cold reservoir with a coefficient of performance K_R .
- The focus of a heat pump is on dumping heat to the hot reservoir with a coefficient of performance K_P .

Conceptual Questions

Exercise:

Problem:

If the refrigerator door is left open, what happens to the temperature of the kitchen?

Solution:

The temperature increases since the heat output behind the refrigerator is greater than the cooling from the inside of the refrigerator.

Exercise:

Problem:

Is it possible for the efficiency of a reversible engine to be greater than 1.0? Is it possible for the coefficient of performance of a reversible refrigerator to be less than 1.0?

Problems

Exercise:

Problem:

A refrigerator has a coefficient of performance of 3.0. (a) If it requires 200 J of work per cycle, how much heat per cycle does it remove the cold reservoir? (b) How much heat per cycle is discarded to the hot reservoir?

Solution:

a. 600 J; b. 800 J

Exercise:**Problem:**

During one cycle, a refrigerator removes 500 J from a cold reservoir and discharges 800 J to its hot reservoir. (a) What is its coefficient of performance? (b) How much work per cycle does it require to operate?

Exercise:**Problem:**

If a refrigerator discards 80 J of heat per cycle and its coefficient of performance is 6.0, what are (a) the quantity of heat it removes per cycle from a cold reservoir and (b) the amount of work per cycle required for its operation?

Solution:

a. 69 J; b. 11 J

Exercise:**Problem:**

A refrigerator has a coefficient of performance of 3.0. (a) If it requires 200 J of work per cycle, how much heat per cycle does it remove the cold reservoir? (b) How much heat per cycle is discarded to the hot reservoir?

Glossary

coefficient of performance

measure of effectiveness of a refrigerator or heat pump

heat pump

device that delivers heat to a hot reservoir

refrigerator

device that removes heat from a cold reservoir

Statements of the Second Law of Thermodynamics

By the end of this section, you will be able to:

- Contrast the second law of thermodynamics statements according to Kelvin and Clausius formulations
- Interpret the second of thermodynamics via irreversibility

Earlier in this chapter, we introduced the Clausius statement of the second law of thermodynamics, which is based on the irreversibility of spontaneous heat flow. As we remarked then, the second law of thermodynamics can be stated in several different ways, and all of them can be shown to imply the others. In terms of heat engines, the second law of thermodynamics may be stated as follows:

Note:

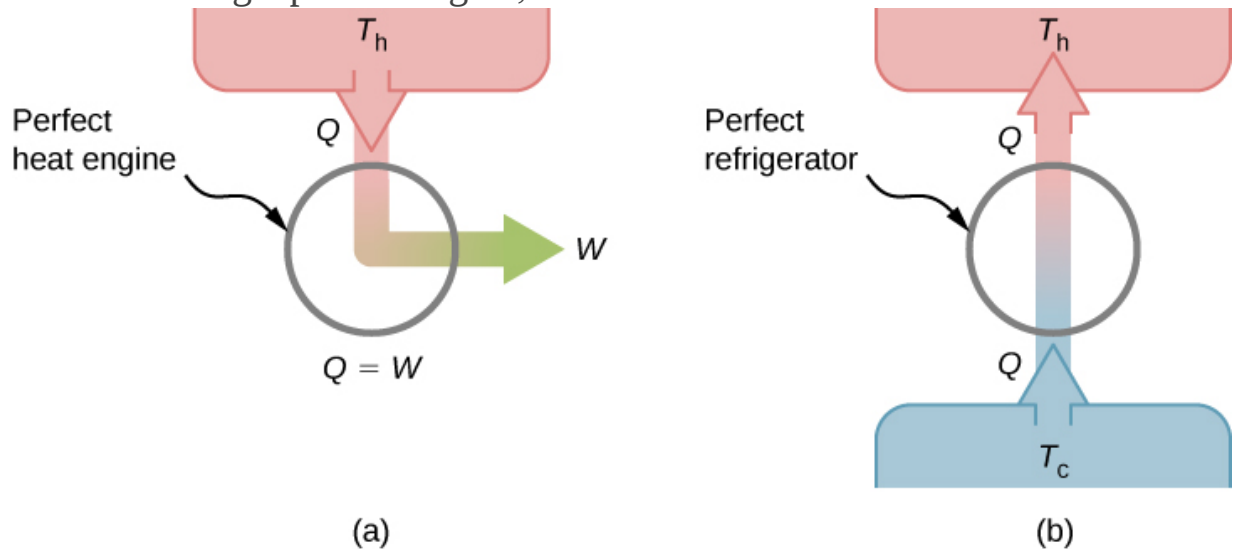
Second Law of Thermodynamics (Kelvin statement)

It is impossible to convert the heat from a single source into work without any other effect.

This is known as the **Kelvin statement of the second law of thermodynamics**. This statement describes an unattainable “**perfect engine**,” as represented schematically in [\[link\]](#)(a). Note that “without any other effect” is a very strong restriction. For example, an engine can absorb heat and turn it all into work, *but not if it completes a cycle*. Without completing a cycle, the substance in the engine is not in its original state and therefore an “other effect” has occurred. Another example is a chamber of gas that can absorb heat from a heat reservoir and do work isothermally against a piston as it expands. However, if the gas were returned to its initial state (that is, made to complete a cycle), it would have to be compressed and heat would have to be extracted from it.

The Kelvin statement is a manifestation of a well-known engineering problem. Despite advancing technology, we are not able to build a heat

engine that is 100% efficient. The first law does not exclude the possibility of constructing a perfect engine, but the second law forbids it.



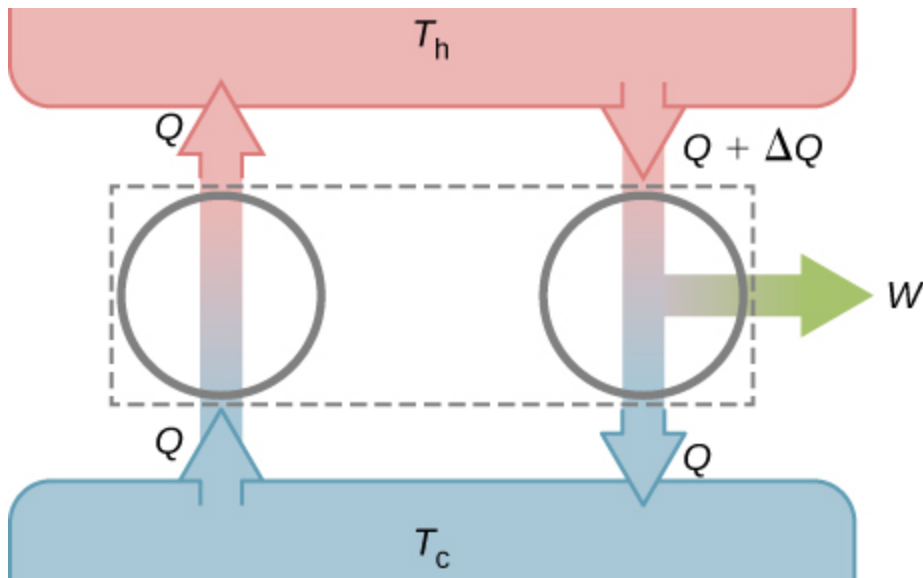
(a) A “perfect heat engine” converts all input heat into work. (b) A “perfect refrigerator” transports heat from a cold reservoir to a hot reservoir without work input. Neither of these devices is achievable in reality.

We can show that the Kelvin statement is equivalent to the Clausius statement if we view the two objects in the Clausius statement as a cold reservoir and a hot reservoir. Thus, the Clausius statement becomes: *It is impossible to construct a refrigerator that transfers heat from a cold reservoir to a hot reservoir without aid from an external source.* The Clausius statement is related to the everyday observation that heat never flows spontaneously from a cold object to a hot object. *Heat transfer in the direction of increasing temperature always requires some energy input.* A “**perfect refrigerator**,” shown in [\[link\]](#)(b), which works without such external aid, is impossible to construct.

To prove the equivalence of the Kelvin and Clausius statements, we show that if one statement is false, it necessarily follows that the other statement is also false. Let us first assume that the Clausius statement is false, so that the perfect refrigerator of [\[link\]](#)(b) does exist. The refrigerator removes heat

Q from a cold reservoir at a temperature T_c and transfers all of it to a hot reservoir at a temperature T_h . Now consider a real heat engine working in the same temperature range. It extracts heat $Q + \Delta Q$ from the hot reservoir, does work W , and discards heat Q to the cold reservoir. From the first law, these quantities are related by $W = (Q + \Delta Q) - Q = \Delta Q$.

Suppose these two devices are combined as shown in [\[link\]](#). The net heat removed from the hot reservoir is ΔQ , no net heat transfer occurs to or from the cold reservoir, and work W is done on some external body. Since $W = \Delta Q$, the combination of a perfect refrigerator and a real heat engine is itself a perfect heat engine, thereby contradicting the Kelvin statement. Thus, if the Clausius statement is false, the Kelvin statement must also be false.

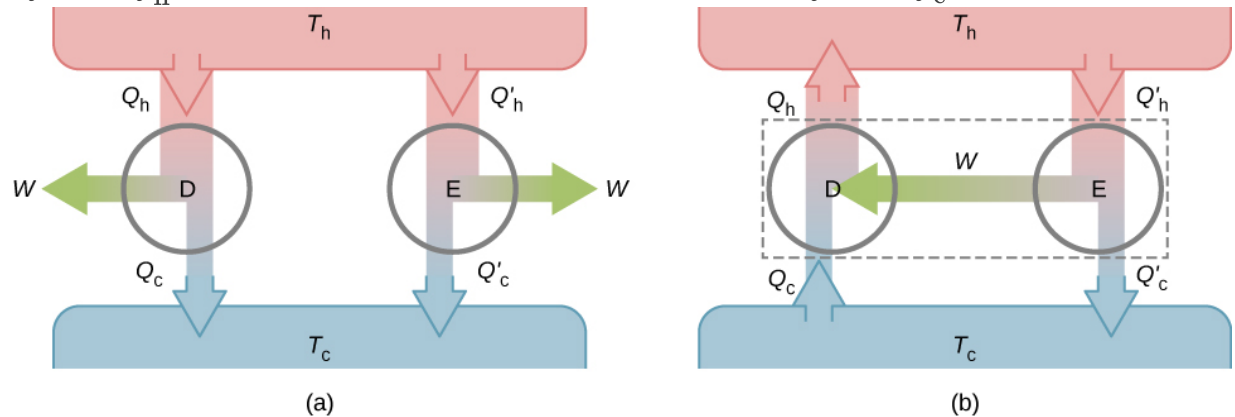


Combining a perfect refrigerator and a real heat engine yields a perfect heat engine because

$$W = \Delta Q.$$

Using the second law of thermodynamics, we now prove two important properties of heat engines operating between two heat reservoirs. The first property is that *any reversible engine operating between two reservoirs has a greater efficiency than any irreversible engine operating between the same two reservoirs.*

The second property to be demonstrated is that *all reversible engines operating between the same two reservoirs have the same efficiency*. To show this, we start with the two engines D and E of [\[link\]](#)(a), which are operating between two common heat reservoirs at temperatures T_h and T_c . First, we assume that D is a reversible engine and that E is a hypothetical irreversible engine that has a higher efficiency than D. If both engines perform the same amount of work W per cycle, it follows from [\[link\]](#) that $Q_h > Q'_h$. It then follows from the first law that $Q_c > Q'_c$.



(a) Two uncoupled engines D and E working between the same reservoirs. (b) The coupled engines, with D working in reverse.

Suppose the cycle of D is reversed so that it operates as a refrigerator, and the two engines are coupled such that the work output of E is used to drive D, as shown in [\[link\]](#)(b). Since $Q_h > Q'_h$ and $Q_c > Q'_c$, the net result of each cycle is equivalent to a spontaneous transfer of heat from the cold reservoir to the hot reservoir, a process the second law does not allow. The original assumption must therefore be wrong, and it is impossible to construct an irreversible engine such that E is more efficient than the reversible engine D.

Now it is quite easy to demonstrate that the efficiencies of all reversible engines operating between the same reservoirs are equal. Suppose that D and E are both reversible engines. If they are coupled as shown in [\[link\]](#)(b), the efficiency of E cannot be greater than the efficiency of D, or the second law would be violated. If both engines are then reversed, the same

reasoning implies that the efficiency of D cannot be greater than the efficiency of E. Combining these results leads to the conclusion that all reversible engines working between the same two reservoirs have the same efficiency.

Note:

Exercise:

Problem:

Check Your Understanding What is the efficiency of a perfect heat engine? What is the coefficient of performance of a perfect refrigerator?

Solution:

A perfect heat engine would have $Q_c = 0$, which would lead to $e = 1 - Q_c/Q_h = 1$. A perfect refrigerator would need zero work, that is, $W = 0$, which leads to $K_R = Q_c/W \rightarrow \infty$.

Note:

Exercise:

Problem:

Check Your Understanding Show that $Q_h - Q'_h = Q_c - Q'_c$ for the hypothetical engine of [\[link\]](#)(b).

Solution:

From the engine on the right, we have $W = Q'_h - Q'_c$. From the refrigerator on the right, we have $Q_h = Q_c + W$. Thus, $W = Q'_h - Q'_c = Q_h - Q_c$.

Summary

- The Kelvin statement of the second law of thermodynamics: It is impossible to convert the heat from a single source into work without any other effect.
- The Kelvin statement and Clausius statement of the second law of thermodynamics are equivalent.

Conceptual Questions

Exercise:

Problem:

In the text, we showed that if the Clausius statement is false, the Kelvin statement must also be false. Now show the reverse, such that if the Kelvin statement is false, it follows that the Clausius statement is false.

Solution:

If we combine a perfect engine and a real refrigerator with the engine converting heat Q from the hot reservoir into work $W = Q$ to drive the refrigerator, then the heat dumped to the hot reservoir by the refrigerator will be $W + \Delta Q$, resulting in a perfect refrigerator transferring heat ΔQ from the cold reservoir to hot reservoir without any other effect.

Exercise:

Problem:

Why don't we operate ocean liners by extracting heat from the ocean or operate airplanes by extracting heat from the atmosphere?

Exercise:

Problem:

Discuss the practical advantages and disadvantages of heat pumps and electric heating.

Solution:

Heat pumps can efficiently extract heat from the ground to heat on cooler days or pull heat out of the house on warmer days. The disadvantage of heat pumps are that they are more costly than alternatives, require maintenance, and will not work efficiently when temperature differences between the inside and outside are very large. Electric heating is much cheaper to purchase than a heat pump; however, it may be more costly to run depending on the electric rates and amount of usage.

Exercise:**Problem:**

The energy output of a heat pump is greater than the energy used to operate the pump. Why doesn't this statement violate the first law of thermodynamics?

Exercise:**Problem:**

Speculate as to why nuclear power plants are less efficient than fossil-fuel plants based on temperature arguments.

Solution:

A nuclear reactor needs to have a lower temperature to operate, so its efficiency will not be as great as a fossil-fuel plant. This argument does not take into consideration the amount of energy per reaction: Nuclear power has a far greater energy output than fossil fuels.

Exercise:

Problem:

An ideal gas goes from state (p_i, V_i) to state (p_f, V_f) when it is allowed to expand freely. Is it possible to represent the actual process on a pV diagram? Explain.

Glossary

Kelvin statement of the second law of thermodynamics

it is impossible to convert the heat from a single source into work without any other effect

perfect engine

engine that can convert heat into work with 100% efficiency

perfect refrigerator (heat pump)

refrigerator (heat pump) that can remove (dump) heat without any input of work

The Carnot Cycle

- Describe the Carnot cycle with the roles of all four processes involved
- Outline the Carnot principle and its implications
- Demonstrate the equivalence of the Carnot principle and the second law of thermodynamics

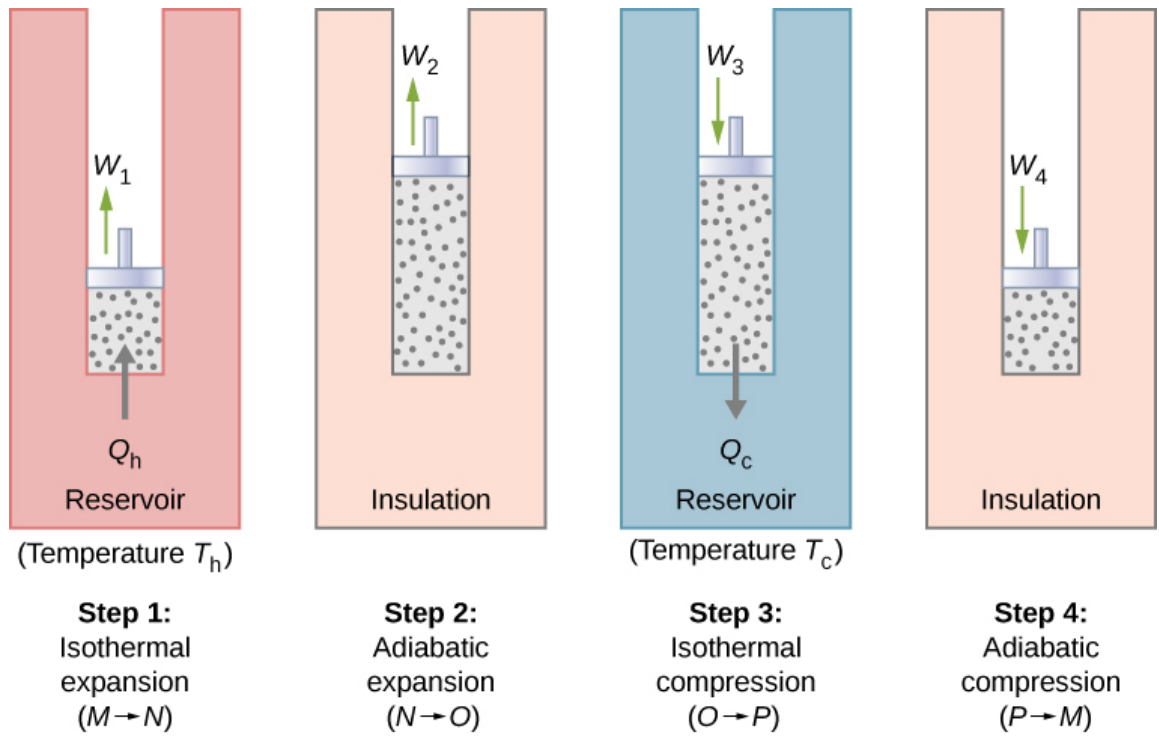
In the early 1820s, Sadi Carnot (1786–1832), a French engineer, became interested in improving the efficiencies of practical heat engines. In 1824, his studies led him to propose a hypothetical working cycle with the highest possible efficiency between the same two reservoirs, known now as the **Carnot cycle**. An engine operating in this cycle is called a **Carnot engine**. The Carnot cycle is of special importance for a variety of reasons. At a practical level, this cycle represents a reversible model for the steam power plant and the refrigerator or heat pump. Yet, it is also very important theoretically, for it plays a major role in the development of another important statement of the second law of thermodynamics. Finally, because only two reservoirs are involved in its operation, it can be used along with the second law of thermodynamics to define an absolute temperature scale that is truly independent of any substance used for temperature measurement.

With an ideal gas as the working substance, the steps of the Carnot cycle, as represented by [\[link\]](#), are as follows.

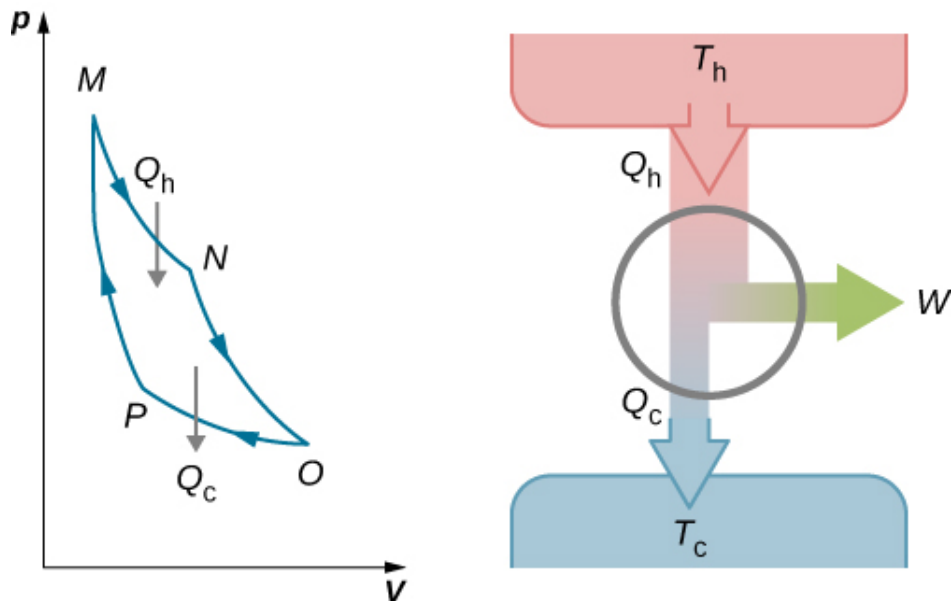
1. *Isothermal expansion*. The gas is placed in thermal contact with a heat reservoir at a temperature T_h . The gas absorbs heat Q_h from the heat reservoir and is allowed to expand isothermally, doing work W_1 . Because the internal energy E_{int} of an ideal gas is a function of the temperature only, the change of the internal energy is zero, that is, $\Delta E_{\text{int}} = 0$ during this isothermal expansion. With the first law of thermodynamics, $\Delta E_{\text{int}} = Q - W$, we find that the heat absorbed by the gas is

Equation:

$$Q_h = W_1 = nRT_h \ln \frac{V_N}{V_M}.$$



The four processes of the Carnot cycle. The working substance is assumed to be an ideal gas whose thermodynamic path $MNOP$ is represented in [\[link\]](#).



The total work done by the gas in the Carnot cycle is

shown and given by the area enclosed by the loop $MNOPM$.

2. *Adiabatic expansion.* The gas is thermally isolated and allowed to expand further, doing work W_2 . Because this expansion is adiabatic, the temperature of the gas falls—in this case, from T_h to T_c . From $pV^\gamma = \text{constant}$ and the equation of state for an ideal gas, $pV = nRT$, we have

Equation:

$$TV^{\gamma-1} = \text{constant},$$

so that

Equation:

$$T_h V_N^{\gamma-1} = T_c V_O^{\gamma-1}.$$

3. *Isothermal compression.* The gas is placed in thermal contact with a cold reservoir at temperature T_c and compressed isothermally. During this process, work W_3 is done on the gas and it gives up heat Q_c to the cold reservoir. The reasoning used in step 1 now yields

Equation:

$$Q_c = nRT_c \ln \frac{V_O}{V_P},$$

where Q_c is the heat dumped to the cold reservoir by the gas.

4. *Adiabatic compression.* The gas is thermally isolated and returned to its initial state by compression. In this process, work W_4 is done on the gas. Because the compression is adiabatic, the temperature of the gas rises—from T_c to T_h in this particular case. The reasoning of step 2 now gives

Equation:

$$T_c V_P^{\gamma-1} = T_h V_M^{\gamma-1}.$$

The total work done by the gas in the Carnot cycle is given by

Equation:

$$W = W_1 + W_2 - W_3 - W_4.$$

This work is equal to the area enclosed by the loop shown in the pV diagram of [\[link\]](#). Because the initial and final states of the system are the same, the change of the internal energy of the gas in the cycle must be zero, that is, $\Delta E_{\text{int}} = 0$. The first law of thermodynamics then gives

Equation:

$$W = Q - \Delta E_{\text{int}} = (Q_{\text{h}} - Q_{\text{c}}) - 0,$$

and

Equation:

$$W = Q_{\text{h}} - Q_{\text{c}}.$$

To find the efficiency of this engine, we first divide Q_{c} by Q_{h} :

Equation:

$$\frac{Q_{\text{c}}}{Q_{\text{h}}} = \frac{T_{\text{c}}}{T_{\text{h}}} \frac{\ln V_{\text{O}}/V_{\text{P}}}{\ln V_{\text{N}}/V_{\text{M}}}.$$

When the adiabatic constant from step 2 is divided by that of step 4, we find

Equation:

$$\frac{V_{\text{O}}}{V_{\text{P}}} = \frac{V_{\text{N}}}{V_{\text{M}}}.$$

Substituting this into the equation for $Q_{\text{c}}/Q_{\text{h}}$, we obtain

Equation:

$$\frac{Q_{\text{c}}}{Q_{\text{h}}} = \frac{T_{\text{c}}}{T_{\text{h}}}.$$

Finally, with [\[link\]](#), we find that the efficiency of this ideal gas Carnot engine is given by

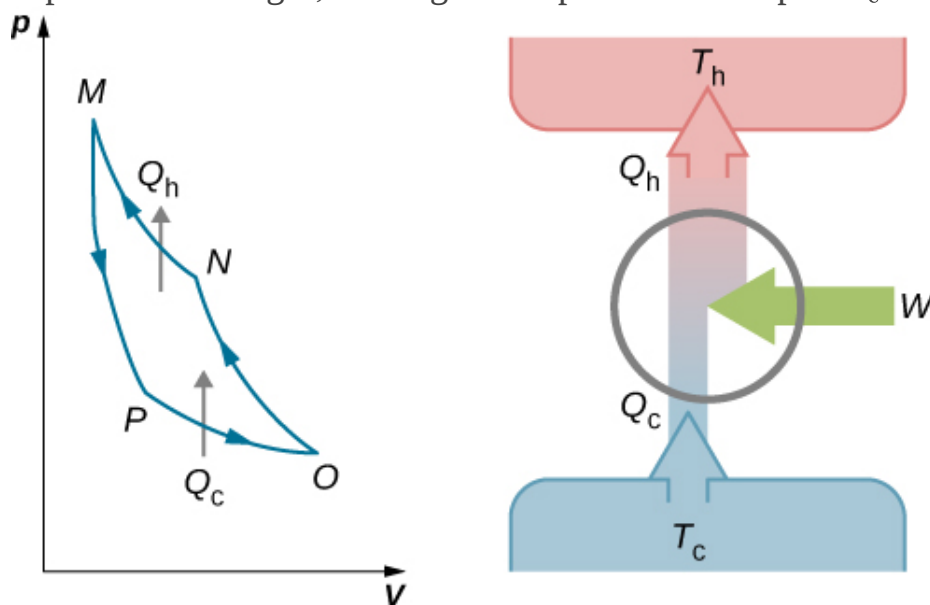
Note:

Equation:

$$e = 1 - \frac{T_c}{T_h}.$$

An engine does not necessarily have to follow a Carnot engine cycle. All engines, however, have the same *net* effect, namely the absorption of heat from a hot reservoir, the production of work, and the discarding of heat to a cold reservoir. This leads us to ask: Do all reversible cycles operating between the same two reservoirs have the same efficiency? The answer to this question comes from the second law of thermodynamics discussed earlier: *All reversible engine cycles produce exactly the same efficiency.* Also, as you might expect, all real engines operating between two reservoirs are less efficient than reversible engines operating between the same two reservoirs. This too is a consequence of the second law of thermodynamics shown earlier.

The cycle of an ideal gas Carnot refrigerator is represented by the pV diagram of [\[link\]](#). It is a Carnot engine operating in reverse. The refrigerator extracts heat Q_c from a cold-temperature reservoir at T_c when the ideal gas expands isothermally. The gas is then compressed adiabatically until its temperature reaches T_h , after which an isothermal compression of the gas results in heat Q_h being discarded to a high-temperature reservoir at T_h . Finally, the cycle is completed by an adiabatic expansion of the gas, causing its temperature to drop to T_c .



The work done on the gas in one cycle of the Carnot refrigerator is shown and given by the area enclosed by the loop *MPONM*.

The work done on the ideal gas is equal to the area enclosed by the path of the pV diagram. From the first law, this work is given by

Equation:

$$W = Q_h - Q_c.$$

An analysis just like the analysis done for the Carnot engine gives

Equation:

$$\frac{Q_c}{T_c} = \frac{Q_h}{T_h}.$$

When combined with [\[link\]](#), this yields

Note:

Equation:

$$K_R = \frac{T_c}{T_h - T_c}$$

for the coefficient of performance of the ideal-gas Carnot refrigerator. Similarly, we can work out the coefficient of performance for a Carnot heat pump as

Note:

Equation:

$$K_P = \frac{Q_h}{Q_h - Q_c} = \frac{T_h}{T_h - T_c}.$$

We have just found equations representing the efficiency of a Carnot engine and the coefficient of performance of a Carnot refrigerator or a Carnot heat pump, assuming an ideal gas for the working substance in both devices. However, these equations are more general than their derivations imply. We will soon show that they are both valid no matter what the working substance is.

Carnot summarized his study of the Carnot engine and Carnot cycle into what is now known as **Carnot's principle**:

Note:

Carnot's Principle

No engine working between two reservoirs at constant temperatures can have a greater efficiency than a reversible engine.

This principle can be viewed as another statement of the second law of thermodynamics and can be shown to be equivalent to the Kelvin statement and the Clausius statement.

Example:

The Carnot Engine

A Carnot engine has an efficiency of 0.60 and the temperature of its cold reservoir is 300 K. (a) What is the temperature of the hot reservoir? (b) If the engine does 300 J of work per cycle, how much heat is removed from the high-temperature reservoir per cycle? (c) How much heat is exhausted to the low-temperature reservoir per cycle?

Strategy

From the temperature dependence of the thermal efficiency of the Carnot engine, we can find the temperature of the hot reservoir. Then, from the definition of the efficiency, we can find the heat removed when the work done by the engine is

given. Finally, energy conservation will lead to how much heat must be dumped to the cold reservoir.

Solution

- a. From $e = 1 - T_c/T_h$ we have

Equation:

$$0.60 = 1 - \frac{300 \text{ K}}{T_h},$$

so that the temperature of the hot reservoir is

Equation:

$$T_h = \frac{300 \text{ K}}{1 - 0.60} = 750 \text{ K}.$$

- b. By definition, the efficiency of the engine is $e = W/Q$, so that the heat removed from the high-temperature reservoir per cycle is

Equation:

$$Q_h = \frac{W}{e} = \frac{300 \text{ J}}{0.60} = 500 \text{ J}.$$

- c. From the first law, the heat exhausted to the low-temperature reservoir per cycle by the engine is

Equation:

$$Q_c = Q_h - W = 500 \text{ J} - 300 \text{ J} = 200 \text{ J}.$$

Significance

A Carnot engine has the maximum possible efficiency of converting heat into work between two reservoirs, but this does not necessarily mean it is 100% efficient. As the difference in temperatures of the hot and cold reservoir increases, the efficiency of a Carnot engine increases.

Example:

A Carnot Heat Pump

Imagine a Carnot heat pump operates between an outside temperature of $0\text{ }^{\circ}\text{C}$ and an inside temperature of $20.0\text{ }^{\circ}\text{C}$. What is the work needed if the heat delivered to the inside of the house is 30.0 kJ ?

Strategy

Because the heat pump is assumed to be a Carnot pump, its performance coefficient is given by $K_P = Q_h/W = T_h/(T_h - T_c)$. Thus, we can find the work W from the heat delivered Q_h .

Solution

The work needed is obtained from

Equation:

$$W = Q_h/K_P = Q_h(T_h - T_c)/T_h = 30\text{ kJ} \times (293\text{ K} - 273\text{ K})/293\text{ K} = 2\text{ kJ}.$$

Significance

We note that this work depends not only on the heat delivered to the house but also on the temperatures outside and inside. The dependence on the temperature outside makes them impractical to use in areas where the temperature is much colder outside than room temperature.

In terms of energy costs, the heat pump is a very economical means for heating buildings ([link](#)). Contrast this method with turning electrical energy directly into heat with resistive heating elements. In this case, one unit of electrical energy furnishes at most only one unit of heat. Unfortunately, heat pumps have problems that do limit their usefulness. They are quite expensive to purchase compared to resistive heating elements, and, as the performance coefficient for a Carnot heat pump shows, they become less effective as the outside temperature decreases. In fact, below about $-10\text{ }^{\circ}\text{C}$, the heat they furnish is less than the energy used to operate them.



A photograph of a heat pump (large box) located outside a house. This heat pump is located in a warm climate area, like the southern United States, since it would be far too inefficient located in the northern half of the United States. (credit: modification of work by Peter Stevens)

Note:

Exercise:

Problem:

Check Your Understanding A Carnot engine operates between reservoirs at $400\text{ }^{\circ}\text{C}$ and $30\text{ }^{\circ}\text{C}$. (a) What is the efficiency of the engine? (b) If the engine does 5.0 J of work per cycle, how much heat per cycle does it absorb from the high-temperature reservoir? (c) How much heat per cycle does it exhaust to the cold-temperature reservoir? (d) What temperatures at the cold reservoir would give the minimum and maximum efficiency?

Solution:

a. $e = 1 - T_c/T_h = 0.55$; b. $Q_h = eW = 9.1 \text{ J}$; c. $Q_c = Q_h - W = 4.1 \text{ J}$; d. -273°C and 400°C

Note:

Exercise:

Problem:

Check Your Understanding A Carnot refrigerator operates between two heat reservoirs whose temperatures are 0°C and 25°C . (a) What is the coefficient of performance of the refrigerator? (b) If 200 J of work are done on the working substance per cycle, how much heat per cycle is extracted from the cold reservoir? (c) How much heat per cycle is discarded to the hot reservoir?

Solution:

a. $K_R = T_c/(T_h - T_c) = 10.9$; b. $Q_c = K_RW = 2.18 \text{ kJ}$; c. $Q_h = Q_c + W = 2.38 \text{ kJ}$

Summary

- The Carnot cycle is the most efficient engine for a reversible cycle designed between two reservoirs.
- The Carnot principle is another way of stating the second law of thermodynamics.

Conceptual Questions

Exercise:

Problem:

To increase the efficiency of a Carnot engine, should the temperature of the hot reservoir be raised or lowered? What about the cold reservoir?

Solution:

In order to increase the efficiency, the temperature of the hot reservoir should be raised, and the cold reservoir should be lowered as much as possible. This can be seen in [\[link\]](#).

Exercise:

Problem: How could you design a Carnot engine with 100% efficiency?

Exercise:

Problem: What type of processes occur in a Carnot cycle?

Solution:

adiabatic and isothermal processes

Problems**Exercise:****Problem:**

The temperature of the cold and hot reservoirs between which a Carnot refrigerator operates are $-73\text{ }^{\circ}\text{C}$ and $270\text{ }^{\circ}\text{C}$, respectively. Which is its coefficient of performance?

Solution:

1.58

Exercise:**Problem:**

Suppose a Carnot refrigerator operates between T_c and T_h . Calculate the amount of work required to extract 1.0 J of heat from the cold reservoir if (a) $T_c = 7\text{ }^{\circ}\text{C}$, $T_h = 27\text{ }^{\circ}\text{C}$; (b) $T_c = -73\text{ }^{\circ}\text{C}$, $T_h = 27\text{ }^{\circ}\text{C}$; (c) $T_c = -173\text{ }^{\circ}\text{C}$, $T_h = 27\text{ }^{\circ}\text{C}$; and (d) $T_c = -273\text{ }^{\circ}\text{C}$, $T_h = 27\text{ }^{\circ}\text{C}$.

Exercise:

Problem:

A Carnot engine operates between reservoirs at 600 and 300 K. If the engine absorbs 100 J per cycle at the hot reservoir, what is its work output per cycle?

Solution:

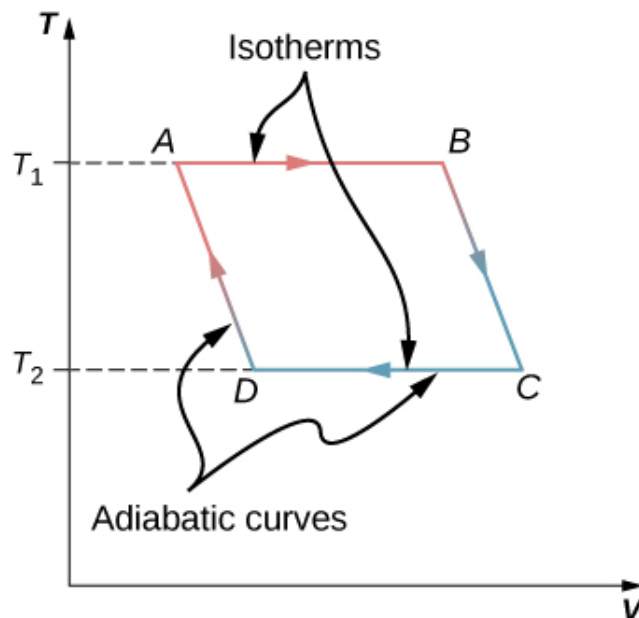
50 J

Exercise:**Problem:**

A 500-W motor operates a Carnot refrigerator between $-5\text{ }^{\circ}\text{C}$ and $30\text{ }^{\circ}\text{C}$. (a) What is the amount of heat per second extracted from the inside of the refrigerator? (b) How much heat is exhausted to the outside air per second?

Exercise:

Problem: Sketch a Carnot cycle on a temperature-volume diagram.

Solution:

Exercise:**Problem:**

A Carnot heat pump operates between 0°C and 20°C . How much heat is exhausted into the interior of a house for every 1.0 J of work done by the pump?

Exercise:**Problem:**

An engine operating between heat reservoirs at 20°C and 200°C extracts 1000 J per cycle from the hot reservoir. (a) What is the maximum possible work that engine can do per cycle? (b) For this maximum work, how much heat is exhausted to the cold reservoir per cycle?

Solution:

a. 381 J; b. 619 J

Exercise:**Problem:**

Suppose a Carnot engine can be operated between two reservoirs as either a heat engine or a refrigerator. How is the coefficient of performance of the refrigerator related to the efficiency of the heat engine?

Exercise:**Problem:**

A Carnot engine is used to measure the temperature of a heat reservoir. The engine operates between the heat reservoir and a reservoir consisting of water at its triple point. (a) If 400 J per cycle are removed from the heat reservoir while 200 J per cycle are deposited in the triple-point reservoir, what is the temperature of the heat reservoir? (b) If 400 J per cycle are removed from the triple-point reservoir while 200 J per cycle are deposited in the heat reservoir, what is the temperature of the heat reservoir?

Solution:

a. 546 K; b. 137 K

Exercise:**Problem:**

What is the minimum work required of a refrigerator if it is to extract 50 J per cycle from the inside of a freezer at $-10\text{ }^{\circ}\text{C}$ and exhaust heat to the air at $25\text{ }^{\circ}\text{C}$?

Glossary**Carnot cycle**

cycle that consists of two isotherms at the temperatures of two reservoirs and two adiabatic processes connecting the isotherms

Carnot engine

Carnot heat engine, refrigerator, or heat pump that operates on a Carnot cycle

Carnot principle

principle governing the efficiency or performance of a heat device operating on a Carnot cycle: any reversible heat device working between two reservoirs must have the same efficiency or performance coefficient, greater than that of an irreversible heat device operating between the same two reservoirs

Entropy

By the end of this section you will be able to:

- Describe the meaning of entropy
- Calculate the change of entropy for some simple processes

The second law of thermodynamics is best expressed in terms of a *change* in the thermodynamic variable known as **entropy**, which is represented by the symbol S . Entropy, like internal energy, is a state function. This means that when a system makes a transition from one state into another, the change in entropy ΔS is independent of path and depends only on the thermodynamic variables of the two states.

We first consider ΔS for a system undergoing a reversible process at a constant temperature. In this case, the change in entropy of the system is given by

Note:

Equation:

$$\Delta S = \frac{Q}{T},$$

where Q is the heat exchanged by the system kept at a temperature T (in kelvin). If the system absorbs heat—that is, with $Q > 0$ —the entropy of the system increases. As an example, suppose a gas is kept at a constant temperature of 300 K while it absorbs 10 J of heat in a reversible process. Then from [\[link\]](#), the entropy change of the gas is

Equation:

$$\Delta S = \frac{10 \text{ J}}{300 \text{ K}} = 0.033 \text{ J/K}.$$

Similarly, if the gas loses 5.0 J of heat; that is, $Q = -5.0 \text{ J}$, at temperature $T = 200 \text{ K}$, we have the entropy change of the system given by

Equation:

$$\Delta S = \frac{-5.0 \text{ J}}{200 \text{ K}} = -0.025 \text{ J/K}.$$

Example:

Entropy Change of Melting Ice

Heat is slowly added to a 50-g chunk of ice at 0°C until it completely melts into water at the same temperature. What is the entropy change of the ice?

Strategy

Because the process is slow, we can approximate it as a reversible process. The temperature is a constant, and we can therefore use [\[link\]](#) in the calculation.

Solution

The ice is melted by the addition of heat:

Equation:

$$Q = mL_f = 50 \text{ g} \times 335 \text{ J/g} = 16.8 \text{ kJ}.$$

In this reversible process, the temperature of the ice-water mixture is fixed at 0°C or 273 K . Now from $\Delta S = Q/T$, the entropy change of the ice is

Equation:

$$\Delta S = \frac{16.8 \text{ kJ}}{273 \text{ K}} = 61.5 \text{ J/K}$$

when it melts to water at 0°C .

Significance

During a phase change, the temperature is constant, allowing us to use [\[link\]](#) to solve this problem. The same equation could also be used if we changed from a liquid to a gas phase, since the temperature does not change during that process either.

The change in entropy of a system for an arbitrary, reversible transition for which the temperature is not necessarily constant is defined by modifying $\Delta S = Q/T$. Imagine a system making a transition from state A to B in small, discrete steps. The temperatures associated with these states are T_A and T_B , respectively. During each step of the transition, the system exchanges heat ΔQ_i reversibly at a temperature T_i . This can be accomplished experimentally by placing the system in thermal contact with a large number of heat reservoirs of varying temperatures T_i , as illustrated in [\[link\]](#). The change in entropy for each step is $\Delta S_i = Q_i/T_i$. The net change in entropy of the system for the transition is

Equation:

$$\Delta S = S_B - S_A = \sum_i \Delta S_i = \sum_i \frac{\Delta Q_i}{T_i}.$$

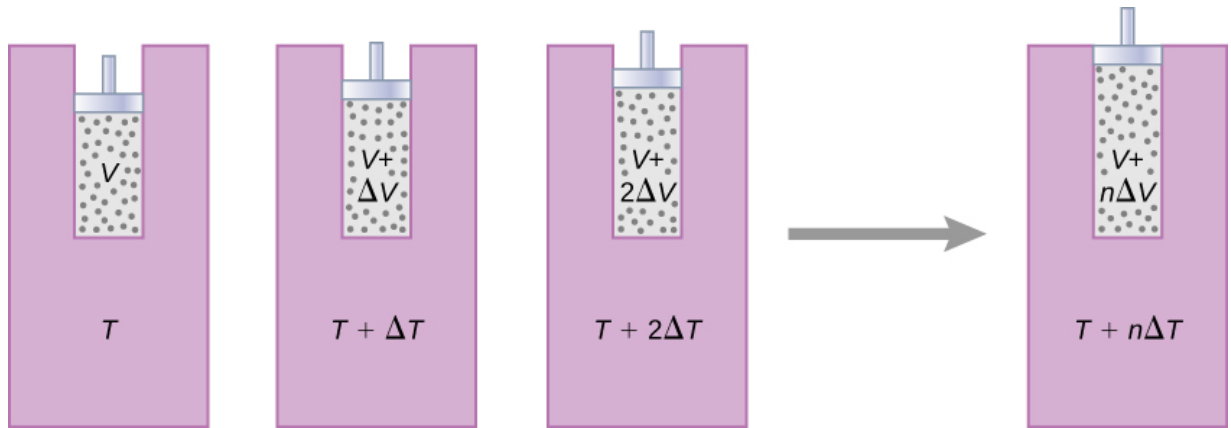
We now take the limit as $\Delta Q_i \rightarrow 0$, and the number of steps approaches infinity. Then, replacing the summation by an integral, we obtain

Note:

Equation:

$$\Delta S = S_B - S_A = \int_A^B \frac{dQ}{T},$$

where the integral is taken between the initial state A and the final state B . This equation is valid only if the transition from A to B is reversible.



The gas expands at constant pressure as its temperature is increased in small steps through the use of a series of heat reservoirs.

As an example, let us determine the net entropy change of a reversible engine while it undergoes a single Carnot cycle. In the adiabatic steps 2 and 4 of the cycle shown in [\[link\]](#), no heat exchange takes place, so

$\Delta S_2 = \Delta S_4 = \int dQ/T = 0$. In step 1, the engine absorbs heat Q_h at a temperature T_h , so its entropy change is $\Delta S_1 = Q_h/T_h$. Similarly, in step 3, $\Delta S_3 = -Q_c/T_c$. The net entropy change of the engine in one cycle of operation is then

Equation:

$$\Delta S_E = \Delta S_1 + \Delta S_2 + \Delta S_3 + \Delta S_4 = \frac{Q_h}{T_h} - \frac{Q_c}{T_c}.$$

However, we know that for a Carnot engine,

Equation:

$$\frac{Q_h}{T_h} = \frac{Q_c}{T_c},$$

so

Equation:

$$\Delta S_E = 0.$$

There is no net change in the entropy of the Carnot engine over a complete cycle. Although this result was obtained for a particular case, its validity can be shown to be far more general: There is no net change in the entropy of a system undergoing any complete reversible cyclic process.

Mathematically, we write this statement as

Note:**Equation:**

$$\oint dS = \oint \frac{dQ}{T} = 0$$

where \oint represents the integral over a *closed reversible path*.

We can use [\[link\]](#) to show that the entropy change of a system undergoing a reversible process between two given states is path independent. An arbitrary, closed path for a reversible cycle that passes through the states A and B is shown in [\[link\]](#). From [\[link\]](#), $\oint dS = 0$ for this closed path. We may split this integral into two segments, one along I, which leads from A to B , the other along II, which leads from B to A . Then

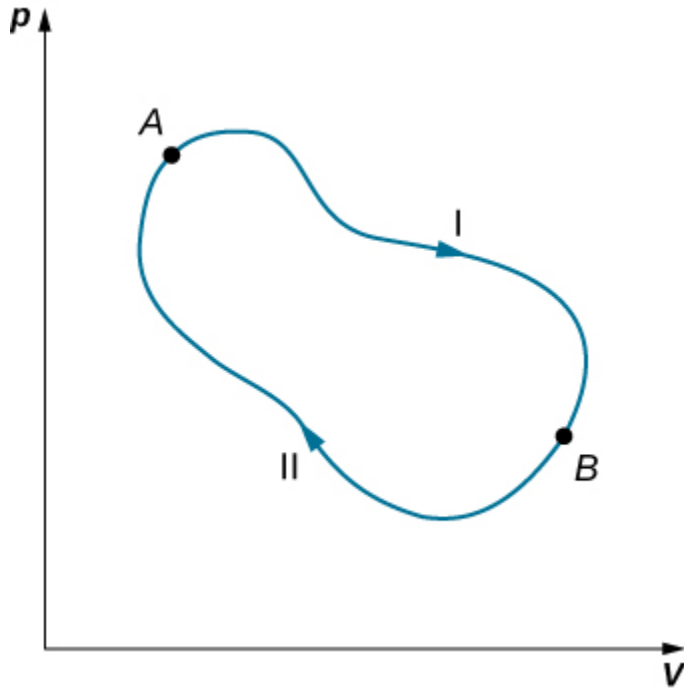
Equation:

$$\left[\int_A^B dS \right]_{\text{I}} + \left[\int_B^A dS \right]_{\text{II}} = 0.$$

Since the process is reversible,

Equation:

$$\left[\int_A^B dS \right]_{\text{I}} = \left[\int_A^B dS \right]_{\text{II}}.$$



The closed loop passing through states *A* and *B* represents a reversible cycle.

Hence, the entropy change in going from *A* to *B* is the same for paths I and II. Since paths I and II are arbitrary, reversible paths, the entropy change in a transition between two equilibrium states is the same for all the reversible processes joining these states. Entropy, like internal energy, is therefore a state function.

What happens if the process is irreversible? When the process is irreversible, we expect the entropy of a closed system, or the system and its

environment (the universe), to increase. Therefore we can rewrite this expression as

Note:

Equation:

$$\Delta S \geq 0,$$

where S is the total entropy of the closed system or the entire universe, and the equal sign is for a reversible process. The fact is the **entropy statement of the second law of thermodynamics**:

Note:

Second Law of Thermodynamics (Entropy statement)

The entropy of a closed system and the entire universe never decreases.

We can show that this statement is consistent with the Kelvin statement, the Clausius statement, and the Carnot principle.

Example:

Entropy Change of a System during an Isobaric Process

Determine the entropy change of an object of mass m and specific heat c that is cooled rapidly (and irreversibly) at constant pressure from T_h to T_c .

Strategy

The process is clearly stated as an irreversible process; therefore, we cannot simply calculate the entropy change from the actual process. However, because entropy of a system is a function of state, we can imagine a reversible process that starts from the same initial state and ends

at the given final state. Then, the entropy change of the system is given by [\[link\]](#), $\Delta S = \int_A^B dQ/T$.

Solution

To replace this rapid cooling with a process that proceeds reversibly, we imagine that the hot object is put into thermal contact with successively cooler heat reservoirs whose temperatures range from T_h to T_c .

Throughout the substitute transition, the object loses infinitesimal amounts of heat dQ , so we have

Equation:

$$\Delta S = \int_{T_h}^{T_c} \frac{dQ}{T}.$$

From the definition of heat capacity, an infinitesimal exchange dQ for the object is related to its temperature change dT by

Equation:

$$dQ = mc dT.$$

Substituting this dQ into the expression for ΔS , we obtain the entropy change of the object as it is cooled at constant pressure from T_h to T_c :

Equation:

$$\Delta S = \int_{T_h}^{T_c} \frac{mc dT}{T} = mc \ln \frac{T_c}{T_h}.$$

Note that $\Delta S < 0$ here because $T_c < T_h$. In other words, the object has lost some entropy. But if we count whatever is used to remove the heat from the object, we would still end up with $\Delta S_{\text{universe}} > 0$ because the process is irreversible.

Significance

If the temperature changes during the heat flow, you must keep it inside the integral to solve for the change in entropy. If, however, the temperature is constant, you can simply calculate the entropy change as the heat flow divided by the temperature.

Example:**Stirling Engine**

The steps of a reversible Stirling engine are as follows. For this problem, we will use 0.0010 mol of a monatomic gas that starts at a temperature of 133 °C and a volume of 0.10 m³, which will be called point A. Then it goes through the following steps:

1. Step *AB*: isothermal expansion at 133 °C from 0.10 m³ to 0.20 m³
2. Step *BC*: isochoric cooling to 33 °C
3. Step *CD*: isothermal compression at 33 °C from 0.20 m³ to 0.10 m³
4. Step *DA*: isochoric heating back to 133 °C and 0.10 m³

- (a) Draw the pV diagram for the Stirling engine with proper labels.
(b) Fill in the following table.

Step	W (J)	Q (J)	ΔS (J/K)
Step <i>AB</i>			
Step <i>BC</i>			
Step <i>CD</i>			
Step <i>DA</i>			
Complete cycle			

- (c) How does the efficiency of the Stirling engine compare to the Carnot engine working within the same two heat reservoirs?

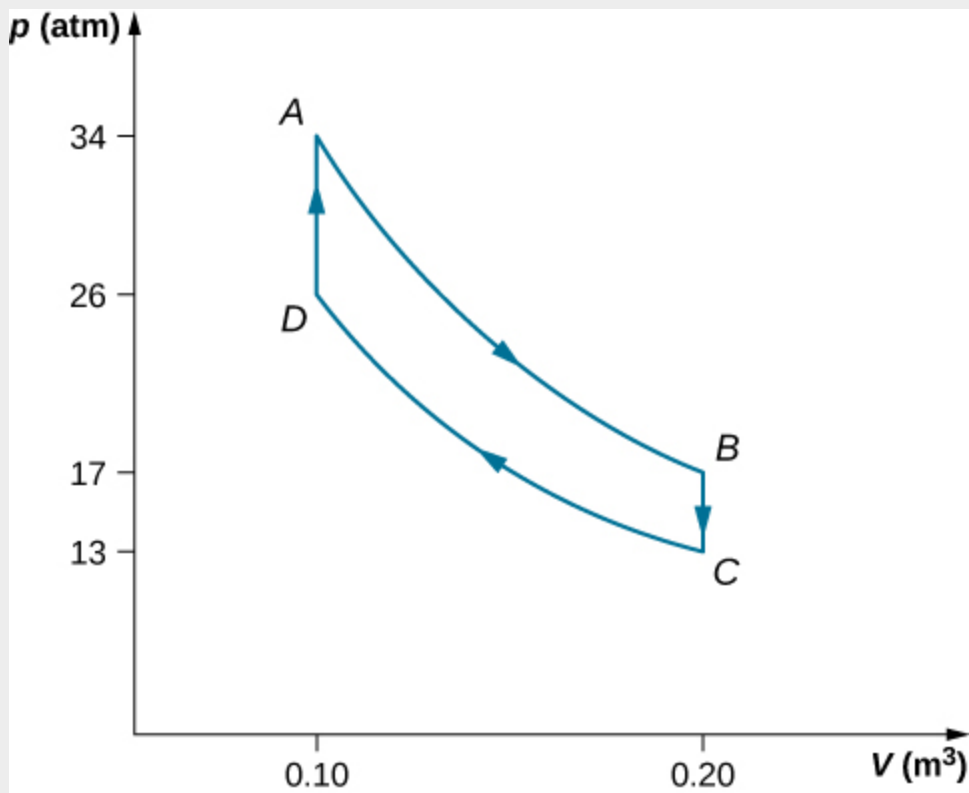
Strategy

Using the ideal gas law, calculate the pressure at each point so that they can be labeled on the pV diagram. Isothermal work is calculated using

$W = nRT \ln \left(\frac{V_2}{V_1} \right)$, and an isochoric process has no work done. The heat flow is calculated from the first law of thermodynamics, $Q = \Delta E_{\text{int}} - W$ where $\Delta E_{\text{int}} = \frac{3}{2}nR\Delta T$ for monatomic gasses. Isothermal steps have a change in entropy of Q/T , whereas isochoric steps have $\Delta S = \frac{3}{2}nR \ln \left(\frac{T_2}{T_1} \right)$. The efficiency of a heat engine is calculated by using $e_{\text{Stir}} = W/Q_{\text{h}}$.

Solution

a. The graph is shown below.



b. The completed table is shown below.

Step	W (J)	Q (J)	ΔS (J/K)
Step <i>AB</i> Isotherm	2.3	2.3	0.0057
Step <i>BC</i> Isochoric	0	−1.2	0.0035
Step <i>CD</i> Isotherm	−1.8	−1.8	−0.0059
Step <i>DA</i> Isochoric	0	1.2	−0.0035
Complete cycle	0.5	0.5	~ 0

c. The efficiency of the Stirling heat engine is

Equation:

$$e_{\text{Stir}} = W/Q_h = (Q_{AB} + Q_{CD})/(Q_{AB} + Q_{DA}) = 0.5/4.5 = 0.11.$$

If this were a Carnot engine operating between the same heat reservoirs, its efficiency would be

Equation:

$$e_{\text{Car}} = 1 - \left(\frac{T_c}{T_h} \right) = 0.25.$$

Therefore, the Carnot engine would have a greater efficiency than the Stirling engine.

Significance

In the early days of steam engines, accidents would occur due to the high pressure of the steam in the boiler. Robert Stirling developed an engine in 1816 that did not use steam and therefore was safer. The Stirling engine was commonly used in the nineteenth century, but developments in steam and internal combustion engines have made it difficult to broaden the use of the Stirling engine.

The Stirling engine uses compressed air as the working substance, which passes back and forth between two chambers with a porous plug, called the

regenerator, which is made of material that does not conduct heat as well. In two of the steps, pistons in the two chambers move in phase.

Summary

- The change in entropy for a reversible process at constant temperature is equal to the heat divided by the temperature. The entropy change of a system under a reversible process is given by $\Delta S = \int_A^B dQ/T$.
- A system's change in entropy between two states is independent of the reversible thermodynamic path taken by the system when it makes a transition between the states.

Conceptual Questions

Exercise:

Problem:

Does the entropy increase for a Carnot engine for each cycle?

Exercise:

Problem:

Is it possible for a system to have an entropy change if it neither absorbs nor emits heat during a reversible transition? What happens if the process is irreversible?

Solution:

Entropy will not change if it is a reversible transition but will change if the process is irreversible.

Problems

Exercise:**Problem:**

Two hundred joules of heat are removed from a heat reservoir at a temperature of 200 K. What is the entropy change of the reservoir?

Solution:

-1 J/K

Exercise:**Problem:**

In an isothermal reversible expansion at 27°C , an ideal gas does 20 J of work. What is the entropy change of the gas?

Exercise:**Problem:**

An ideal gas at 300 K is compressed isothermally to one-fifth its original volume. Determine the entropy change per mole of the gas.

Solution:

-13 J/(K mole)

Exercise:**Problem:**

What is the entropy change of 10 g of steam at 100°C when it condenses to water at the same temperature?

Exercise:

Problem:

A metal rod is used to conduct heat between two reservoirs at temperatures T_h and T_c , respectively. When an amount of heat Q flows through the rod from the hot to the cold reservoir, what is the net entropy change of the rod, the hot reservoir, the cold reservoir, and the universe?

Solution:

$$-\frac{Q}{T_h}, \frac{Q}{T_c}, Q \left(\frac{1}{T_c} - \frac{1}{T_h} \right)$$

Exercise:**Problem:**

For the Carnot cycle of [\[link\]](#), what is the entropy change of the hot reservoir, the cold reservoir, and the universe?

Exercise:**Problem:**

A 5.0-kg piece of lead at a temperature of 600°C is placed in a lake whose temperature is 15°C . Determine the entropy change of (a) the lead piece, (b) the lake, and (c) the universe.

Solution:

a. -709 J/K ; b. 1300 J/K ; c. 591 J/K

Exercise:**Problem:**

One mole of an ideal gas doubles its volume in a reversible isothermal expansion. (a) What is the change in entropy of the gas? (b) If 1500 J of heat are added in this process, what is the temperature of the gas?

Exercise:

Problem:

One mole of an ideal monatomic gas is confined to a rigid container. When heat is added reversibly to the gas, its temperature changes from T_1 to T_2 . (a) How much heat is added? (b) What is the change in entropy of the gas?

Solution:

a. $Q = nR\Delta T$; b. $S = nR \ln(T_2/T_1)$

Exercise:**Problem:**

(a) A 5.0-kg rock at a temperature of 20°C is dropped into a shallow lake also at 20°C from a height of $1.0 \times 10^3\text{ m}$. What is the resulting change in entropy of the universe? (b) If the temperature of the rock is 100°C when it is dropped, what is the change of entropy of the universe? Assume that air friction is negligible (not a good assumption) and that $c = 860\text{ J/kg} \cdot \text{K}$ is the specific heat of the rock.

Glossary**entropy**

state function of the system that changes when heat is transferred between the system and the environment

entropy statement of the second law of thermodynamics

entropy of a closed system or the entire universe never decreases

Entropy on a Microscopic Scale

By the end of this section you will be able to:

- Interpret the meaning of entropy at a microscopic scale
- Calculate a change in entropy for an irreversible process of a system and contrast with the change in entropy of the universe
- Explain the third law of thermodynamics

We have seen how entropy is related to heat exchange at a particular temperature. In this section, we consider entropy from a statistical viewpoint. Although the details of the argument are beyond the scope of this textbook, it turns out that entropy can be related to how disordered or randomized a system is—the more it is disordered, the higher is its entropy. For example, a new deck of cards is very ordered, as the cards are arranged numerically by suit. In shuffling this new deck, we randomize the arrangement of the cards and therefore increase its entropy ([\[link\]](#)). Thus, by picking one card off the top of the deck, there would be no indication of what the next selected card will be.



The entropy of a new deck of cards goes up after the dealer shuffles them. (credit: “Rommel SK”/YouTube)

The second law of thermodynamics requires that the entropy of the universe increase in any irreversible process. Thus, in terms of order, the second law may be stated as follows:

In any irreversible process, the universe becomes more disordered. For example, the irreversible free expansion of an ideal gas, shown in [\[link\]](#), results in a larger volume for the gas molecules to occupy. A larger volume means more possible arrangements for the same number of atoms, so disorder is also increased. As a result, the entropy of the gas has gone up. The gas in this case is a closed system, and the process is irreversible. Changes in phase also illustrate the connection between entropy and **disorder**.

Example:

Entropy Change of the Universe

Suppose we place 50 g of ice at 0°C in contact with a heat reservoir at 20°C . Heat spontaneously flows from the reservoir to the ice, which melts and eventually reaches a temperature of 20°C . Find the change in entropy of (a) the ice and (b) the universe.

Strategy

Because the entropy of a system is a function of its state, we can imagine two reversible processes for the ice: (1) ice is melted at $0^\circ\text{C}(T_A)$; and (2) melted ice (water) is warmed up from 0°C to $20^\circ\text{C}(T_B)$ under constant pressure. Then, we add the change in entropy of the reservoir when we calculate the change in entropy of the universe.

Solution

- a. From [\[link\]](#), the increase in entropy of the ice is

Equation:

$$\begin{aligned}
\Delta S_{\text{ice}} &= \Delta S_1 + \Delta S_2 \\
&= \frac{mL_f}{T_A} + mc \int_A^B \frac{dT}{T} \\
&= \left(\frac{50 \times 335}{273} + 50 \times 4.19 \times \ln \frac{293}{273} \right) \text{ J/K} \\
&= 76.3 \text{ J/K}.
\end{aligned}$$

b. During this transition, the reservoir gives the ice an amount of heat equal to

Equation:

$$\begin{aligned}
Q &= mL_f + mc(T_B - T_A) \\
&= 50 \times (335 + 4.19 \times 20) \text{ J} \\
&= 2.10 \times 10^4 \text{ J}.
\end{aligned}$$

This leads to a change (decrease) in entropy of the reservoir:

Equation:

$$\Delta S_{\text{reservoir}} = \frac{-Q}{T_B} = -71.7 \text{ J/K}.$$

The increase in entropy of the universe is therefore

Equation:

$$\Delta S_{\text{universe}} = 76.3 \text{ J/K} - 71.7 \text{ J/K} = 4.6 \text{ J/K} > 0.$$

Significance

The entropy of the universe therefore is greater than zero since the ice gains more entropy than the reservoir loses. If we considered only the phase change of the ice into water and not the temperature increase, the entropy change of the ice and reservoir would be the same, resulting in the universe gaining no entropy.

This process also results in a more disordered universe. The ice changes from a solid with molecules located at specific sites to a liquid whose molecules are much freer to move. The molecular arrangement has therefore become more randomized. Although the change in average kinetic energy of the molecules of the heat reservoir is negligible, there is nevertheless a significant decrease in the entropy of the reservoir because it has many more molecules than the melted ice cube. However, the reservoir's decrease in entropy is still not as large as the increase in entropy of the ice. The increased disorder of the ice more than compensates for the increased order of the reservoir, and the entropy of the universe increases by 4.6 J/K.

You might suspect that the growth of different forms of life might be a net ordering process and therefore a violation of the second law. After all, a single cell gathers molecules and eventually becomes a highly structured organism, such as a human being. However, this ordering process is more than compensated for by the disordering of the rest of the universe. The net result is an increase in entropy and an increase in the disorder of the universe.

Note:**Exercise:****Problem:**

Check Your Understanding In [\[link\]](#), the spontaneous flow of heat from a hot object to a cold object results in a net increase in entropy of the universe. Discuss how this result can be related to an increase in disorder of the system.

Solution:

When heat flows from the reservoir to the ice, the internal (mainly kinetic) energy of the ice goes up, resulting in a higher average speed and thus an average greater position variance of the molecules in the ice. The reservoir does become more ordered, but due to its much

larger amount of molecules, it does not offset the change in entropy in the system.

The second law of thermodynamics makes clear that the entropy of the universe never decreases during any thermodynamic process. For any other thermodynamic system, when the process is reversible, the change of the entropy is given by $\Delta S = Q/T$. But what happens if the temperature goes to zero, $T \rightarrow 0$? It turns out this is not a question that can be answered by the second law.

A fundamental issue still remains: Is it possible to cool a system all the way down to zero kelvin? We understand that the system must be at its lowest energy state because lowering temperature reduces the kinetic energy of the constituents in the system. What happens to the entropy of a system at the absolute zero temperature? It turns out the absolute zero temperature is not reachable—at least, not through a finite number of cooling steps. This is a statement of the **third law of thermodynamics**, whose proof requires quantum mechanics that we do not present here. In actual experiments, physicists have continuously pushed that limit downward, with the lowest temperature achieved at about 1×10^{-10} K in a low-temperature lab at the Helsinki University of Technology in 2008.

Like the second law of thermodynamics, the third law of thermodynamics can be stated in different ways. One of the common statements of the third law of thermodynamics is: *The absolute zero temperature cannot be reached through any finite number of cooling steps.*

In other words, the temperature of any given physical system must be finite, that is, $T > 0$. This produces a very interesting question in physics: Do we know how a system would behave if it were at the absolute zero temperature?

The reason a system is unable to reach 0 K is fundamental and requires quantum mechanics to fully understand its origin. But we can certainly ask what happens to the entropy of a system when we try to cool it down to 0

K. Because the amount of heat that can be removed from the system becomes vanishingly small, we expect that the change in entropy of the system along an isotherm approaches zero, that is,

Note:

Equation:

$$\lim_{T \rightarrow 0} (\Delta S)_T = 0.$$

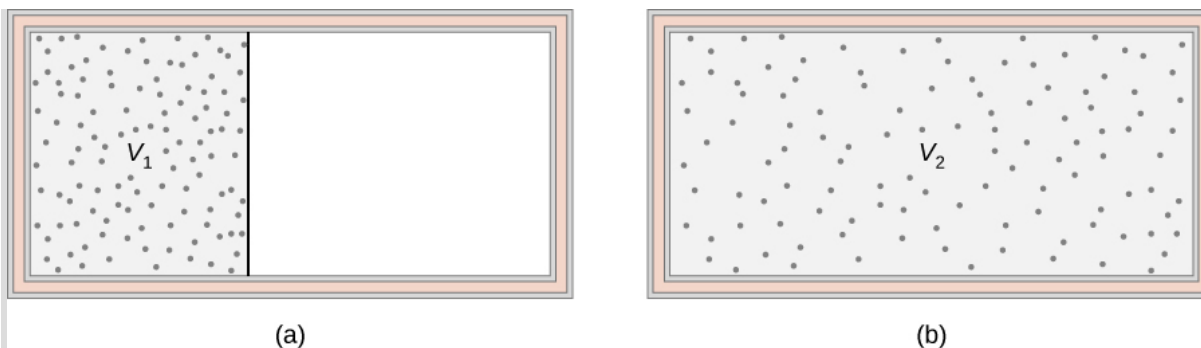
This can be viewed as another statement of the third law, with all the isotherms becoming **isentropic**, or into a reversible ideal adiabat. We can put this expression in words: *A system becomes perfectly ordered when its temperature approaches absolute zero and its entropy approaches its absolute minimum.*

The third law of thermodynamics puts another limit on what can be done when we look for energy resources. If there could be a reservoir at the absolute zero temperature, we could have engines with efficiency of 100%, which would, of course, violate the second law of thermodynamics.

Example:

Entropy Change of an Ideal Gas in Free Expansion

An ideal gas occupies a partitioned volume V_1 inside a box whose walls are thermally insulating, as shown in [\[link\]](#)(a). When the partition is removed, the gas expands and fills the entire volume V_2 of the box, as shown in part (b). What is the entropy change of the universe (the system plus its environment)?



The adiabatic free expansion of an ideal gas from volume V_1 to volume V_2 .

Strategy

The adiabatic free expansion of an ideal gas is an irreversible process. There is no change in the internal energy (and hence temperature) of the gas in such an expansion because no work or heat transfer has happened. Thus, a convenient reversible path connecting the same two equilibrium states is a slow, isothermal expansion from V_1 to V_2 . In this process, the gas could be expanding against a piston while in thermal contact with a heat reservoir, as in step 1 of the Carnot cycle.

Solution

Since the temperature is constant, the entropy change is given by $\Delta S = Q/T$, where

Equation:

$$Q = W = \int_{V_1}^{V_2} p dV$$

because $\Delta E_{\text{int}} = 0$. Now, with the help of the ideal gas law, we have

Equation:

$$Q = nRT \int_{V_1}^{V_2} \frac{dV}{V} = nRT \ln \frac{V_2}{V_1},$$

so the change in entropy of the gas is

Equation:

$$\Delta S = \frac{Q}{T} = nR \ln \frac{V_2}{V_1}.$$

Because $V_2 > V_1$, ΔS is positive, and the entropy of the gas has gone up during the free expansion.

Significance

What about the environment? The walls of the container are thermally insulating, so no heat exchange takes place between the gas and its surroundings. The entropy of the environment is therefore constant during the expansion. The net entropy change of the universe is then simply the entropy change of the gas. Since this is positive, the entropy of the universe increases in the free expansion of the gas.

Example:

Entropy Change during Heat Transfer

Heat flows from a steel object of mass 4.00 kg whose temperature is 400 K to an identical object at 300 K. Assuming that the objects are thermally isolated from the environment, what is the net entropy change of the universe after thermal equilibrium has been reached?

Strategy

Since the objects are identical, their common temperature at equilibrium is 350 K. To calculate the entropy changes associated with their transitions, we substitute the irreversible process of the heat transfer by two isobaric, reversible processes, one for each of the two objects. The entropy change for each object is then given by $\Delta S = mc \ln(T_B/T_A)$.

Solution

Using $c = 450 \text{ J/kg} \cdot \text{K}$, the specific heat of steel, we have for the hotter object

Equation:

$$\begin{aligned} \Delta S_h &= \int_{T_1}^{T_2} \frac{mc dT}{T} = mc \ln \frac{T_2}{T_1} \\ &= (4.00 \text{ kg})(450 \text{ J/kg} \cdot \text{K}) \ln \frac{350 \text{ K}}{400 \text{ K}} = -240 \text{ J/K}. \end{aligned}$$

Similarly, the entropy change of the cooler object is

Equation:

$$\Delta S_c = (4.00 \text{ kg})(450 \text{ J/kg} \cdot \text{K}) \ln \frac{350 \text{ K}}{300 \text{ K}} = 277 \text{ J/K}.$$

The net entropy change of the two objects during the heat transfer is then

Equation:

$$\Delta S_h + \Delta S_c = 37 \text{ J/K}.$$

Significance

The objects are thermally isolated from the environment, so its entropy must remain constant. Thus, the entropy of the universe also increases by 37 J/K.

Note:**Exercise:****Problem:**

Check Your Understanding A quantity of heat Q is absorbed from a reservoir at a temperature T_h by a cooler reservoir at a temperature T_c . What is the entropy change of the hot reservoir, the cold reservoir, and the universe?

Solution:

$$-Q/T_h; Q/T_c; \text{ and } Q(T_h - T_c)/(T_h T_c)$$

Note:**Exercise:**

Problem:

Check Your Understanding A 50-g copper piece at a temperature of $20\text{ }^{\circ}\text{C}$ is placed into a large insulated vat of water at $100\text{ }^{\circ}\text{C}$. (a) What is the entropy change of the copper piece when it reaches thermal equilibrium with the water? (b) What is the entropy change of the water? (c) What is the entropy change of the universe?

Solution:

a. 4.71 J/K ; b. -4.18 J/K ; c. 0.53 J/K

Note:

View this [site](#) to learn about entropy and microstates. Start with a large barrier in the middle and 1000 molecules in only the left chamber. What is the total entropy of the system? Now remove the barrier and let the molecules travel from the left to the right hand side? What is the total entropy of the system now? Lastly, add heat and note what happens to the temperature. Did this increase entropy of the system?

Summary

- Entropy can be related to how disordered a system is—the more it is disordered, the higher is its entropy. In any irreversible process, the universe becomes more disordered.
- According to the third law of thermodynamics, absolute zero temperature is unreachable.

Key Equations

Result of energy conservation	$W = Q_h - Q_c$
Efficiency of a heat engine	$e = \frac{W}{Q_h} = 1 - \frac{Q_c}{Q_h}$
Coefficient of performance of a refrigerator	$K_R = \frac{Q_c}{W} = \frac{Q_c}{Q_h - Q_c}$
Coefficient of performance of a heat pump	$K_P = \frac{Q_h}{W} = \frac{Q_h}{Q_h - Q_c}$
Resulting efficiency of a Carnot cycle	$e = 1 - \frac{T_c}{T_h}$
Performance coefficient of a reversible refrigerator	$K_R = \frac{T_c}{T_h - T_c}$
Performance coefficient of a reversible heat pump	$K_P = \frac{T_h}{T_h - T_c}$
Entropy of a system undergoing a reversible process at a constant temperature	$\Delta S = \frac{Q}{T}$
Change of entropy of a system under a reversible process	$\Delta S = S_B - S_A = \int_A^B dQ/T$
Entropy of a system undergoing any complete reversible cyclic process	$\oint dS = \oint \frac{dQ}{T} = 0$
Change of entropy of a closed system under an irreversible process	$\Delta S \geq 0$
Change in entropy of the system along an isotherm	$\lim_{T \rightarrow 0} (\Delta S)_T = 0$

Conceptual Questions

Exercise:

Problem:

Are the entropy changes of the *systems* in the following processes positive or negative? (a) *water vapor* that condenses on a cold surface; (b) gas in a container that leaks into the surrounding atmosphere; (c) an *ice cube* that melts in a glass of lukewarm water; (d) the *lukewarm water* of part (c); (e) a *real heat engine* performing a cycle; (f) *food* cooled in a refrigerator.

Exercise:

Problem:

Discuss the entropy changes in the systems of Question 21.10 in terms of disorder.

Solution:

Entropy is a function of disorder, so all the answers apply here as well.

Problems

Exercise:

Problem:

A copper rod of cross-sectional area 5.0 cm^2 and length 5.0 m conducts heat from a heat reservoir at 373 K to one at 273 K . What is the time rate of change of the universe's entropy for this process?

Solution:

$$3.78 \times 10^{-3} \text{ W/K}$$

Exercise:**Problem:**

Fifty grams of water at $20\text{ }^{\circ}\text{C}$ is heated until it becomes vapor at $100\text{ }^{\circ}\text{C}$. Calculate the change in entropy of the water in this process.

Exercise:**Problem:**

Fifty grams of water at $0\text{ }^{\circ}\text{C}$ are changed into vapor at $100\text{ }^{\circ}\text{C}$. What is the change in entropy of the water in this process?

Solution:

430 J/K

Exercise:**Problem:**

In an isochoric process, heat is added to 10 mol of monoatomic ideal gas whose temperature increases from 273 to 373 K. What is the entropy change of the gas?

Exercise:**Problem:**

Two hundred grams of water at $0\text{ }^{\circ}\text{C}$ is brought into contact with a heat reservoir at $80\text{ }^{\circ}\text{C}$. After thermal equilibrium is reached, what is the temperature of the water? Of the reservoir? How much heat has been transferred in the process? What is the entropy change of the water? Of the reservoir? What is the entropy change of the universe?

Solution:

$80\text{ }^{\circ}\text{C}$, $80\text{ }^{\circ}\text{C}$, $6.70 \times 10^4\text{ J}$, 215 J/K, -190 J/K , 25 J/K

Exercise:

Problem:

Suppose that the temperature of the water in the previous problem is raised by first bringing it to thermal equilibrium with a reservoir at a temperature of 40 °C and then with a reservoir at 80 °C. Calculate the entropy changes of (a) each reservoir, (b) of the water, and (c) of the universe.

Exercise:**Problem:**

Two hundred grams of water at 0 °C is brought into contact into thermal equilibrium successively with reservoirs at 20 °C, 40 °C, 60 °C, and 80 °C. (a) What is the entropy change of the water? (b) Of the reservoir? (c) What is the entropy change of the universe?

Solution:

$$\Delta S_{\text{H}_2\text{O}} = 215 \text{ J/K}, \Delta S_{\text{R}} = -208 \text{ J/K}, \Delta S_{\text{U}} = 7 \text{ J/K}$$

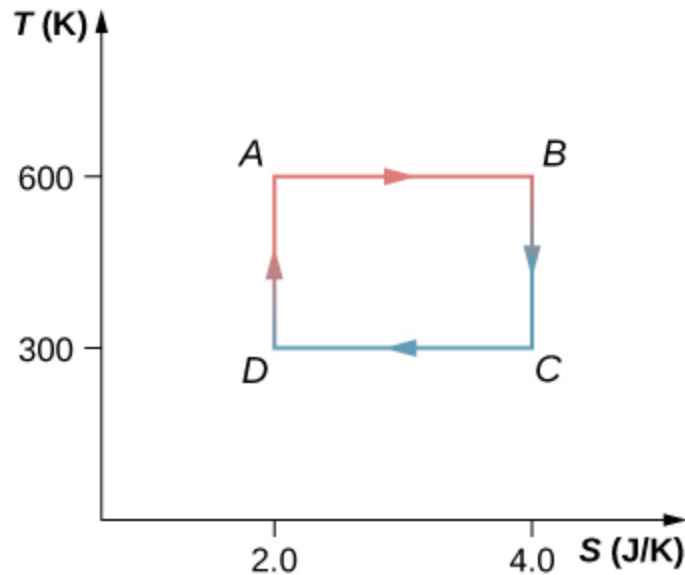
Exercise:**Problem:**

(a) Ten grams of H₂O starts as ice at 0 °C. The ice absorbs heat from the air (just above 0 °C) until all of it melts. Calculate the entropy change of the H₂O, of the air, and of the universe. (b) Suppose that the air in part (a) is at 20 °C rather than 0 °C and that the ice absorbs heat until it becomes water at 20 °C. Calculate the entropy change of the H₂O, of the air, and of the universe. (c) Is either of these processes reversible?

Exercise:

Problem:

The Carnot cycle is represented by the temperature-entropy diagram shown below. (a) How much heat is absorbed per cycle at the high-temperature reservoir? (b) How much heat is exhausted per cycle at the low-temperature reservoir? (c) How much work is done per cycle by the engine? (d) What is the efficiency of the engine?



Solution:

a. 1200 J; b. 600 J; c. 600 J; d. 0.50

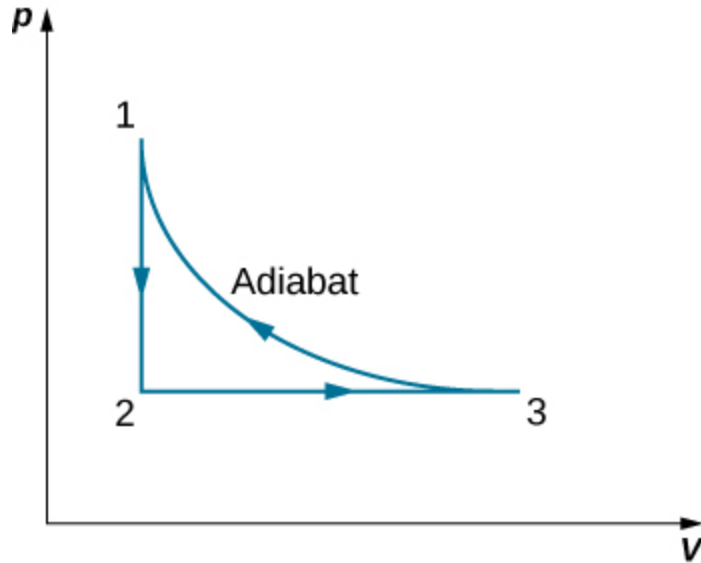
Exercise:**Problem:**

A Carnot engine operating between heat reservoirs at 500 and 300 K absorbs 1500 J per cycle at the high-temperature reservoir. (a) Represent the engine's cycle on a temperature-entropy diagram. (b) How much work per cycle is done by the engine?

Exercise:

Problem:

A monoatomic ideal gas (n moles) goes through a cyclic process shown below. Find the change in entropy of the gas in each step and the total entropy change over the entire cycle.



Solution:

$$\Delta S = nC_V \ln \left(\frac{T_2}{T_1} \right) + nC_p \ln \left(\frac{T_3}{T_2} \right)$$

Exercise:**Problem:**

A Carnot engine has an efficiency of 0.60. When the temperature of its cold reservoir changes, the efficiency drops to 0.55. If initially $T_c = 27^\circ \text{C}$, determine (a) the constant value of T_h and (b) the final value of T_c .

Exercise:

Problem:

A Carnot engine performs 100 J of work while discharging 200 J of heat each cycle. After the temperature of the hot reservoir only is adjusted, it is found that the engine now does 130 J of work while discarding the same quantity of heat. (a) What are the initial and final efficiencies of the engine? (b) What is the fractional change in the temperature of the hot reservoir?

Solution:

a. 0.33, 0.39; b. 0.91

Exercise:**Problem:**

A Carnot refrigerator exhausts heat to the air, which is at a temperature of 25 °C. How much power is used by the refrigerator if it freezes 1.5 g of water per second? Assume the water is at 0 °C.

Additional Problems**Exercise:****Problem:**

A 300-W heat pump operates between the ground, whose temperature is 0 °C, and the interior of a house at 22 °C. What is the maximum amount of heat per hour that the heat pump can supply to the house?

Solution:

$1.45 \times 10^7 \text{ J}$

Exercise:

Problem:

An engineer must design a refrigerator that does 300 J of work per cycle to extract 2100 J of heat per cycle from a freezer whose temperature is -10°C . What is the maximum air temperature for which this condition can be met? Is this a reasonable condition to impose on the design?

Exercise:**Problem:**

A Carnot engine employs 1.5 mol of nitrogen gas as a working substance, which is considered as an ideal diatomic gas with $\gamma = 7/5$ at the working temperatures of the engine. The Carnot cycle goes in the cycle $ABCD$ with AB being an isothermal expansion. The volume at points A and C of the cycle are $5.0 \times 10^{-3} \text{ m}^3$ and 0.15 L, respectively. The engine operates between two thermal baths of temperature 500 K and 300 K. (a) Find the values of volume at B and D . (b) How much heat is absorbed by the gas in the AB isothermal expansion? (c) How much work is done by the gas in the AB isothermal expansion? (d) How much heat is given up by the gas in the CD isothermal expansion? (e) How much work is done by the gas in the CD isothermal compression? (f) How much work is done by the gas in the BC adiabatic expansion? (g) How much work is done by the gas in the DA adiabatic compression? (h) Find the value of efficiency of the engine based on the net work and heat input. Compare this value to the efficiency of a Carnot engine based on the temperatures of the two baths.

Solution:

a. $V_B = 0.042 \text{ m}^3$, $V_D = 0.018 \text{ m}^3$; b. 13,000 J; c. 13,000 J; d. $-8,000 \text{ J}$; e. $-8,000 \text{ J}$; f. 6200 J; g. -6200 J ; h. 39%; with temperatures efficiency is 40%, which is off likely by rounding errors.

Exercise:

Problem:

A 5.0-kg wood block starts with an initial speed of 8.0 m/s and slides across the floor until friction stops it. Estimate the resulting change in entropy of the universe. Assume that everything stays at a room temperature of 20 °C.

Exercise:**Problem:**

A system consisting of 20.0 mol of a monoatomic ideal gas is cooled at constant pressure from a volume of 50.0 L to 10.0 L. The initial temperature was 300 K. What is the change in entropy of the gas?

Solution:

−670 J/K

Exercise:**Problem:**

A glass beaker of mass 400 g contains 500 g of water at 27 °C. The beaker is heated reversibly so that the temperature of the beaker and water rise gradually to 57 °C. Find the change in entropy of the beaker and water together.

Exercise:**Problem:**

A Carnot engine operates between 550 °C and 20 °C baths and produces 300 kJ of energy in each cycle. Find the change in entropy of the (a) hot bath and (b) cold bath, in each Carnot cycle?

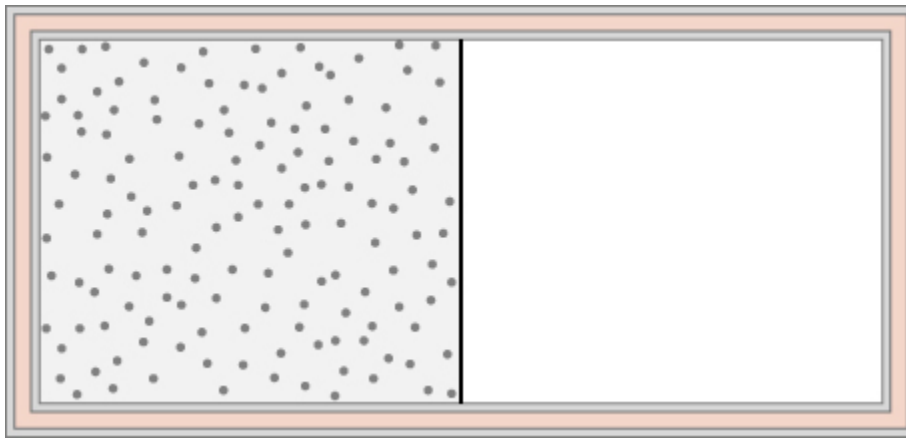
Solution:

a. −570 J/K; b. 570 J/K

Exercise:

Problem:

An ideal gas at temperature T is stored in the left half of an insulating container of volume V using a partition of negligible volume (see below). What is the entropy change per mole of the gas in each of the following cases? (a) The partition is suddenly removed and the gas quickly fills the entire container. (b) A tiny hole is punctured in the partition and after a long period, the gas reaches an equilibrium state such that there is no net flow through the hole. (c) The partition is moved very slowly and adiabatically all the way to the right wall so that the gas finally fills the entire container.

**Exercise:****Problem:**

A 0.50-kg piece of aluminum at 250°C is dropped into 1.0 kg of water at 20°C . After equilibrium is reached, what is the net entropy change of the system?

Solution:

82 J/K

Exercise:

Problem:

Suppose 20 g of ice at 0°C is added to 300 g of water at 60°C . What is the total change in entropy of the mixture after it reaches thermal equilibrium?

Exercise:**Problem:**

A heat engine operates between two temperatures such that the working substance of the engine absorbs 5000 J of heat from the high-temperature bath and discharges 3000 J to the low-temperature bath. The rest of the energy is converted into mechanical energy of the turbine. Find (a) the amount of work produced by the engine and (b) the efficiency of the engine.

Solution:

a. 2000 J; b. 40%

Exercise:**Problem:**

A thermal engine produces 4 MJ of electrical energy while operating between two thermal baths of different temperatures. The working substance of the engine discharges 5 MJ of heat to the cold temperature bath. What is the efficiency of the engine?

Exercise:**Problem:**

A coal power plant consumes 100,000 kg of coal per hour and produces 500 MW of power. If the heat of combustion of coal is 30 MJ/kg, what is the efficiency of the power plant?

Solution:

60%

Exercise:**Problem:**

A Carnot engine operates in a Carnot cycle between a heat source at $550\text{ }^{\circ}\text{C}$ and a heat sink at $20\text{ }^{\circ}\text{C}$. Find the efficiency of the Carnot engine.

Exercise:**Problem:**

A Carnot engine working between two heat baths of temperatures 600 K and 273 K completes each cycle in 5 sec . In each cycle, the engine absorbs 10 kJ of heat. Find the power of the engine.

Solution:

64.4%

Exercise:**Problem:**

A Carnot cycle working between $100\text{ }^{\circ}\text{C}$ and $30\text{ }^{\circ}\text{C}$ is used to drive a refrigerator between $-10\text{ }^{\circ}\text{C}$ and $30\text{ }^{\circ}\text{C}$. How much energy must the Carnot engine produce per second so that the refrigerator is able to discard 10 J of energy per second?

Challenge Problems**Exercise:****Problem:**

(a) An infinitesimal amount of heat is added reversibly to a system. By combining the first and second laws, show that $dU = TdS - dW$. (b) When heat is added to an ideal gas, its temperature and volume change from T_1 and V_1 to T_2 and V_2 . Show that the entropy change of n moles of the gas is given by

$$\Delta S = nC_v \ln \frac{T_2}{T_1} + nR \ln \frac{V_2}{V_1}.$$

Solution:

derive

Exercise:

Problem:

Using the result of the preceding problem, show that for an ideal gas undergoing an adiabatic process, $TV^{\gamma-1}$ is constant.

Exercise:

Problem:

With the help of the two preceding problems, show that ΔS between states 1 and 2 of n moles an ideal gas is given by

$$\Delta S = nC_p \ln \frac{T_2}{T_1} - nR \ln \frac{p_2}{p_1}.$$

Solution:

derive

Exercise:

Problem:

A cylinder contains 500 g of helium at 120 atm and 20 °C. The valve is leaky, and all the gas slowly escapes isothermally into the atmosphere. Use the results of the preceding problem to determine the resulting change in entropy of the universe.

Exercise:

Problem:

A diatomic ideal gas is brought from an initial equilibrium state at $p_1 = 0.50$ atm and $T_1 = 300$ K to a final stage with $p_2 = 0.20$ atm and $T_2 = 500$ K. Use the results of the previous problem to determine the entropy change per mole of the gas.

Solution:

18 J/K

Exercise:**Problem:**

The gasoline internal combustion engine operates in a cycle consisting of six parts. Four of these parts involve, among other things, friction, heat exchange through finite temperature differences, and accelerations of the piston; it is irreversible. Nevertheless, it is represented by the ideal reversible *Otto cycle*, which is illustrated below. The working substance of the cycle is assumed to be air. The six steps of the Otto cycle are as follows:

- i. Isobaric intake stroke (*OA*). A mixture of gasoline and air is drawn into the combustion chamber at atmospheric pressure p_0 as the piston expands, increasing the volume of the cylinder from zero to V_A .
- ii. Adiabatic compression stroke (*AB*). The temperature of the mixture rises as the piston compresses it adiabatically from a volume V_A to V_B .
- iii. Ignition at constant volume (*BC*). The mixture is ignited by a spark. The combustion happens so fast that there is essentially no motion of the piston. During this process, the added heat Q_1 causes the pressure to increase from p_B to p_C at the constant volume $V_B (= V_C)$.
- iv. Adiabatic expansion (*CD*). The heated mixture of gasoline and air expands against the piston, increasing the volume from V_C to V_D .

This is called the *power stroke*, as it is the part of the cycle that delivers most of the power to the crankshaft.

- v. Constant-volume exhaust (*DA*). When the exhaust valve opens, some of the combustion products escape. There is almost no movement of the piston during this part of the cycle, so the volume remains constant at $V_A (= V_D)$. Most of the available energy is lost here, as represented by the heat exhaust Q_2 .
- vi. Isobaric compression (*AO*). The exhaust valve remains open, and the compression from V_A to zero drives out the remaining combustion products.

(a) Using (i) $e = W/Q_1$; (ii) $W = Q_1 - Q_2$; and (iii) $Q_1 = nC_v(T_C - T_B)$, $Q_2 = nC_v(T_D - T_A)$, show that

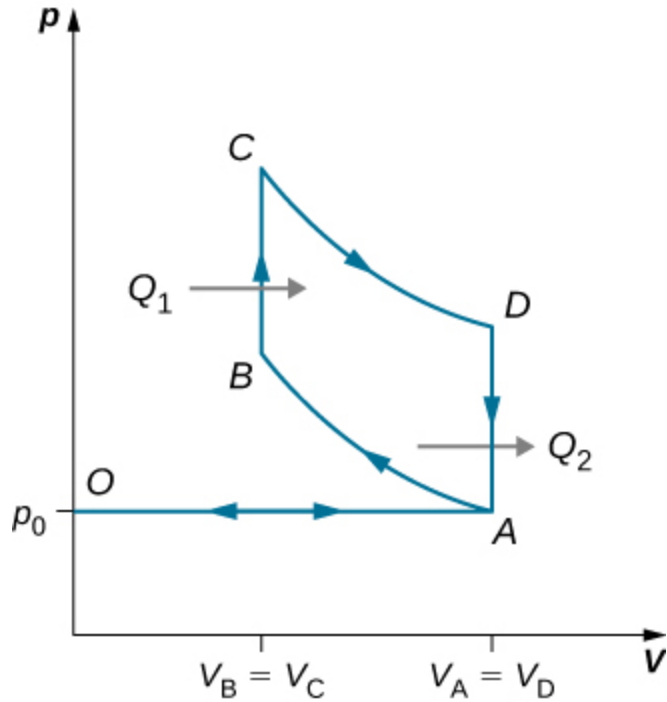
$$e = 1 - \frac{T_D - T_A}{T_C - T_B}.$$

(b) Use the fact that steps (ii) and (iv) are adiabatic to show that

$$e = 1 - \frac{1}{r^{\gamma-1}},$$

where $r = V_A/V_B$. The quantity r is called the *compression ratio* of the engine.

(c) In practice, r is kept less than around 7. For larger values, the gasoline-air mixture is compressed to temperatures so high that it explodes before the finely timed spark is delivered. This *preignition* causes engine knock and loss of power. Show that for $r = 6$ and $\gamma = 1.4$ (the value for air), $e = 0.51$, or an efficiency of 51%. Because of the many irreversible processes, an actual internal combustion engine has an efficiency much less than this ideal value. A typical efficiency for a tuned engine is about 25% to 30%.



Exercise:

Problem:

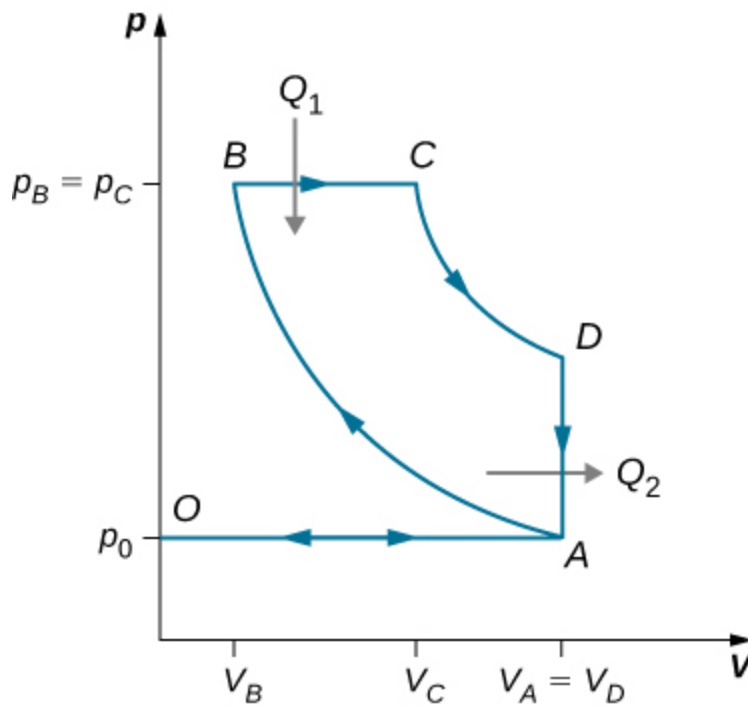
An ideal *diesel* cycle is shown below. This cycle consists of five strokes. In this case, only air is drawn into the chamber during the intake stroke OA . The air is then compressed adiabatically from state A to state B , raising its temperature high enough so that when fuel is added during the power stroke BC , it ignites. After ignition ends at C , there is a further adiabatic power stroke CD . Finally, there is an exhaust at constant volume as the pressure drops from p_D to p_A , followed by a further exhaust when the piston compresses the chamber volume to zero.

(a) Use $W = Q_1 - Q_2$, $Q_1 = nC_p(T_C - T_B)$, and $Q_2 = nC_v(T_D - T_A)$ to show that $e = \frac{W}{Q_1} = 1 - \frac{T_D - T_A}{\gamma(T_C - T_B)}$.

(b) Use the fact that $A \rightarrow B$ and $C \rightarrow D$ are adiabatic to show that

$$e = 1 - \frac{1}{\gamma} \frac{\left(\frac{V_C}{V_D}\right)^\gamma - \left(\frac{V_B}{V_A}\right)^\gamma}{\left(\frac{V_C}{V_D}\right) - \left(\frac{V_B}{V_A}\right)}.$$

(c) Since there is no preignition (remember, the chamber does not contain any fuel during the compression), the compression ratio can be larger than that for a gasoline engine. Typically, $V_A/V_B = 15$ and $V_D/V_C = 5$. For these values and $\gamma = 1.4$, show that $\epsilon = 0.56$, or an efficiency of 56%. Diesel engines actually operate at an efficiency of about 30% to 35% compared with 25% to 30% for gasoline engines.



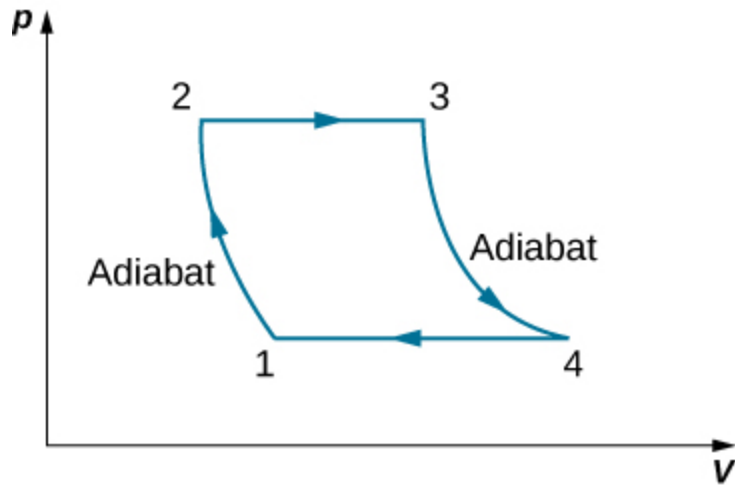
Solution:

proof

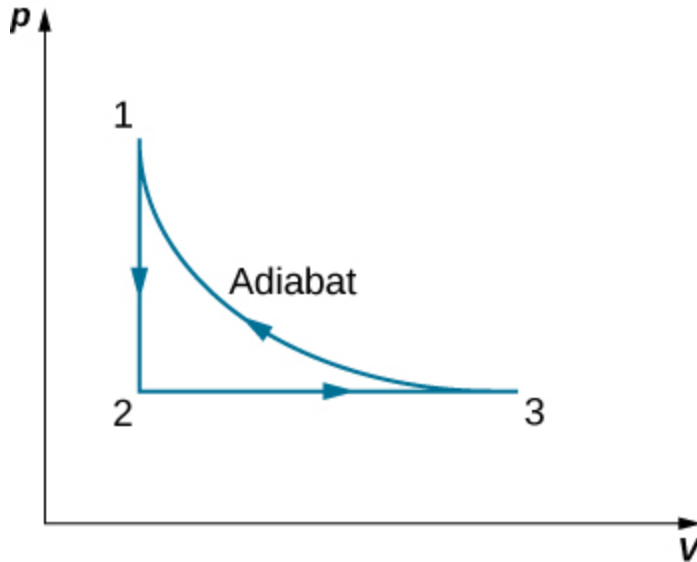
Exercise:

Problem:

Consider an ideal gas Joule cycle, also called the Brayton cycle, shown below. Find the formula for efficiency of the engine using this cycle in terms of P_1 , P_2 , and γ .

**Exercise:****Problem:**

Derive a formula for the coefficient of performance of a refrigerator using an ideal gas as a working substance operating in the cycle shown below in terms of the properties of the three states labeled 1, 2, and 3.



Solution:

$$K_R = \frac{3(p_1 - p_2)V_1}{5p_2V_3 - 3p_1V_1 - p_2V_1}$$

Exercise:

Problem:

Two moles of nitrogen gas, with $\gamma = 7/5$ for ideal diatomic gases, occupies a volume of 10^{-2}m^3 in an insulated cylinder at temperature 300 K. The gas is adiabatically and reversibly compressed to a volume of 5 L. The piston of the cylinder is locked in its place, and the insulation around the cylinder is removed. The heat-conducting cylinder is then placed in a 300-K bath. Heat from the compressed gas leaves the gas, and the temperature of the gas becomes 300 K again. The gas is then slowly expanded at the fixed temperature 300 K until the volume of the gas becomes 10^{-2}m^3 , thus making a complete cycle for the gas. For the entire cycle, calculate (a) the work done by the gas, (b) the heat into or out of the gas, (c) the change in the internal energy of the gas, and (d) the change in entropy of the gas.

Exercise:

Problem:

A Carnot refrigerator, working between $0\text{ }^{\circ}\text{C}$ and $30\text{ }^{\circ}\text{C}$ is used to cool a bucket of water containing 10^{-2} m^3 of water at $30\text{ }^{\circ}\text{C}$ to $5\text{ }^{\circ}\text{C}$ in 2 hours. Find the total amount of work needed.

Solution:

$$W = 110,000\text{ J}$$

Glossary

disorder

measure of order in a system; the greater the disorder is, the higher the entropy

isentropic

reversible adiabatic process where the process is frictionless and no heat is transferred

third law of thermodynamics

absolute zero temperature cannot be reached through any finite number of cooling steps

Entropy and Availability of Energy

Introduction

class="introduction"

Electric
charges
exist all
around us.
They can
cause
objects to be
repelled
from each
other or to
be attracted
to each
other.
(credit:
modification
n of work
by Sean
McGrath)



Back when we were studying Newton's laws, we identified several physical phenomena as forces. We did so based on the effect they had on a physical object: Specifically, they caused the object to accelerate. Later, when we studied impulse and momentum, we expanded this idea to identify a force as any physical phenomenon that changed the momentum of an object. In either case, the result is the same: We recognize a force by the effect that it has on an object.

In [Gravitation](#), we examined the force of gravity, which acts on all objects with mass. In this chapter, we begin the study of the electric force, which acts on all objects with a property called charge. The electric force is much stronger than gravity (in most systems where both appear), but it can be a force of attraction or a force of repulsion, which leads to very different effects on objects. The electric force helps keep atoms together, so it is of fundamental importance in matter. But it also governs most everyday interactions we deal with, from chemical interactions to biological processes.

Electric Charge

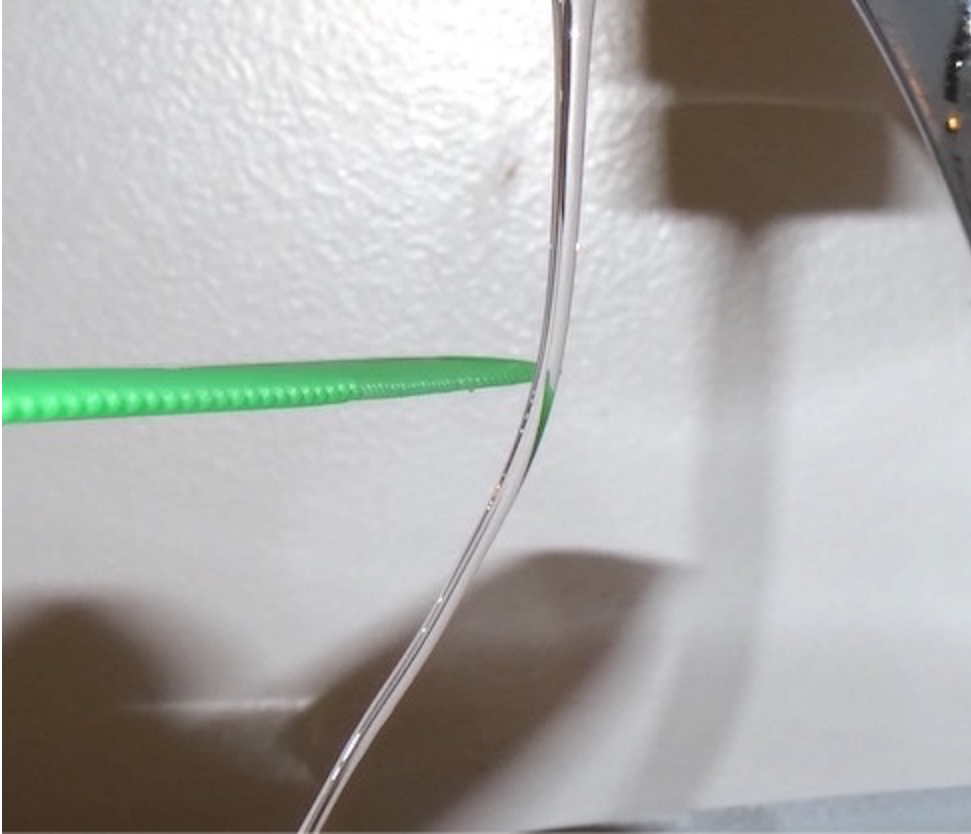
By the end of this section, you will be able to:

- Describe the concept of electric charge
- Explain qualitatively the force electric charge creates

You are certainly familiar with electronic devices that you activate with the click of a switch, from computers to cell phones to television. And you have certainly seen electricity in a flash of lightning during a heavy thunderstorm. But you have also most likely experienced electrical effects in other ways, maybe without realizing that an electric force was involved. Let's take a look at some of these activities and see what we can learn from them about electric charges and forces.

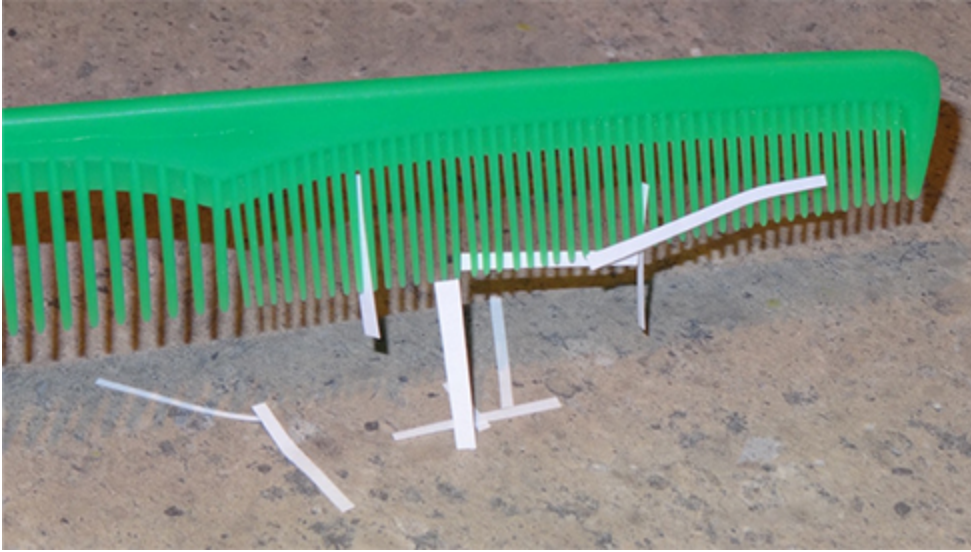
Discoveries

You have probably experienced the phenomenon of **static electricity**: When you first take clothes out of a dryer, many (not all) of them tend to stick together; for some fabrics, they can be very difficult to separate. Another example occurs if you take a woolen sweater off quickly—you can feel (and hear) the static electricity pulling on your clothes, and perhaps even your hair. If you comb your hair on a dry day and then put the comb close to a thin stream of water coming out of a faucet, you will find that the water stream bends toward (is attracted to) the comb ([link](#)).



An electrically charged comb attracts a stream of water from a distance. Note that the water is not touching the comb. (credit: Jane Whitney)

Suppose you bring the comb close to some small strips of paper; the strips of paper are attracted to the comb and even cling to it ([link](#)). In the kitchen, quickly pull a length of plastic cling wrap off the roll; it will tend to cling to most any nonmetallic material (such as plastic, glass, or food). If you rub a balloon on a wall for a few seconds, it will stick to the wall. Probably the most annoying effect of static electricity is getting shocked by a doorknob (or a friend) after shuffling your feet on some types of carpeting.

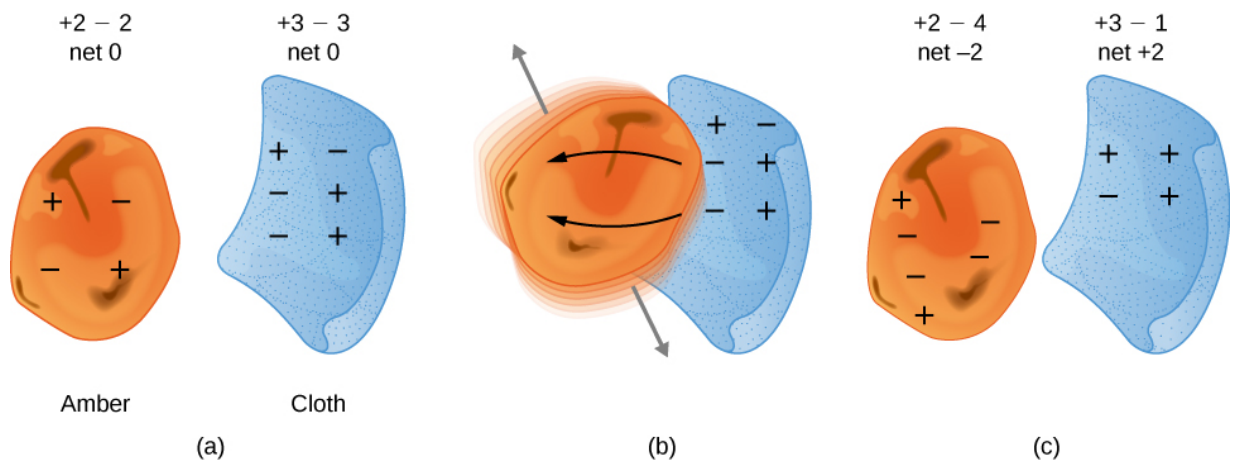


After being used to comb hair, this comb attracts small strips of paper from a distance, without physical contact. Investigation of this behavior helped lead to the concept of the electric force. (credit: Jane Whitney)

Many of these phenomena have been known for centuries. The ancient Greek philosopher Thales of Miletus (624–546 BCE) recorded that when amber (a hard, translucent, fossilized resin from extinct trees) was vigorously rubbed with a piece of fur, a force was created that caused the fur and the amber to be attracted to each other ([\[link\]](#)). Additionally, he found that the rubbed amber would not only attract the fur, and the fur attract the amber, but they both could affect other (nonmetallic) objects, even if not in contact with those objects ([\[link\]](#)).



Borneo amber is mined in Sabah, Malaysia, from shale-sandstone-mudstone veins. When a piece of amber is rubbed with a piece of fur, the amber gains more electrons, giving it a net negative charge. At the same time, the fur, having lost electrons, becomes positively charged.
(credit: “Sebakoamber”/Wikimedia Commons)



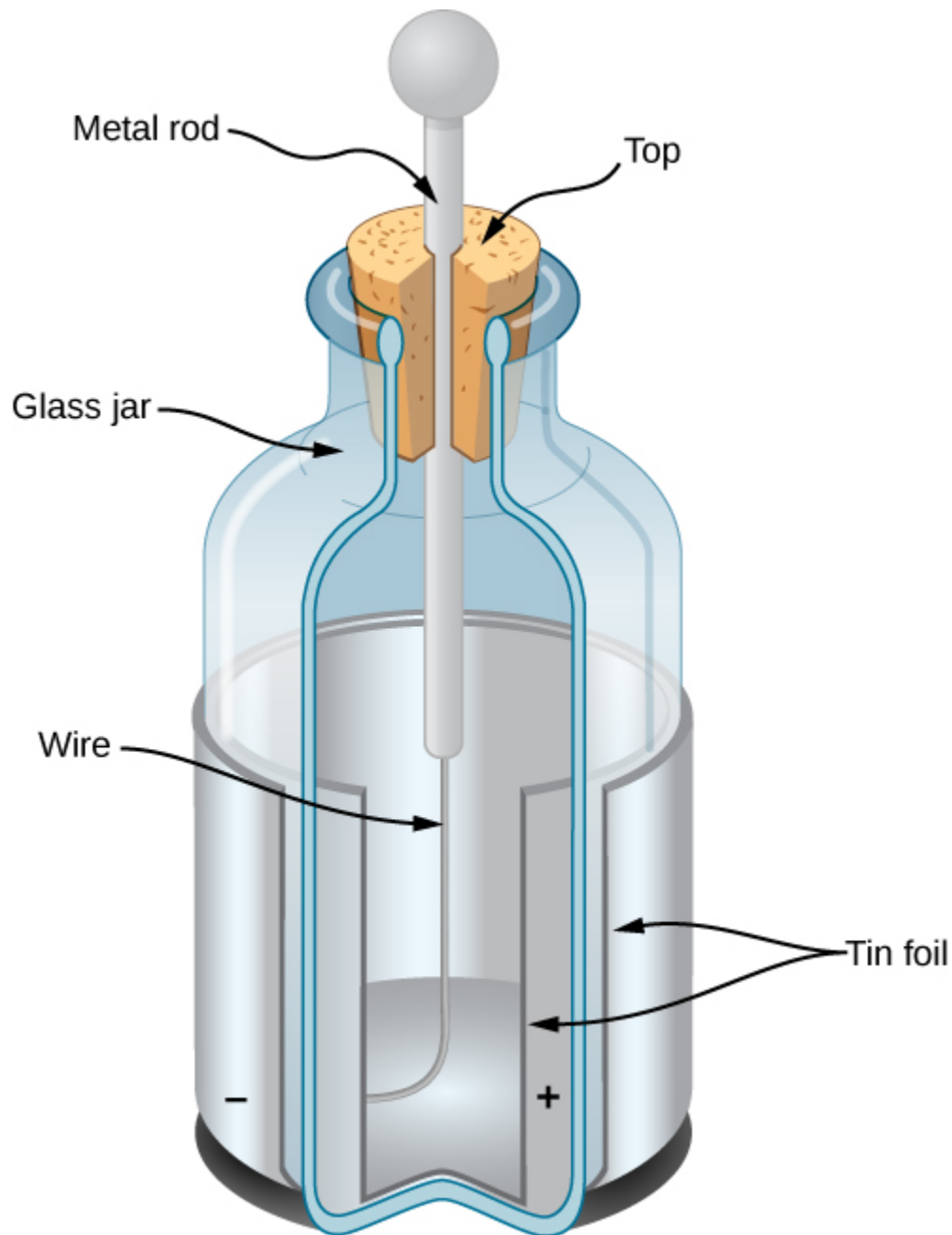
When materials are rubbed together, charges can be separated, particularly if one material has a greater affinity for electrons than another. (a) Both the amber and cloth are originally neutral, with equal positive and negative charges. Only a tiny fraction of the charges are involved, and only a few of them are shown here. (b) When rubbed together, some negative charge is transferred to the amber, leaving the cloth with a net positive charge. (c) When separated, the amber and cloth now have net charges, but the absolute value of the net positive and negative charges will be equal.

The English physicist William Gilbert (1544–1603) also studied this attractive force, using various substances. He worked with amber, and, in addition, he experimented with rock crystal and various precious and semi-precious gemstones. He also experimented with several metals. He found that the metals never exhibited this force, whereas the minerals did. Moreover, although an electrified amber rod would attract a piece of fur, it would repel another electrified amber rod; similarly, two electrified pieces of fur would repel each other.

This suggested there were two types of an electric property; this property eventually came to be called **electric charge**. The difference between the two types of electric charge is in the directions of the electric forces that each type of charge causes: These forces are repulsive when the same type of charge exists on two interacting objects and attractive when the charges are of opposite types. The SI unit of electric charge is the **coulomb** (C), after the French physicist Charles-Augustin de Coulomb (1736–1806).

The most peculiar aspect of this new force is that it does not require physical contact between the two objects in order to cause an acceleration. This is an example of a so-called “long-range” force. (Or, as James Clerk Maxwell later phrased it, “action at a distance.”) With the exception of gravity, all other forces we have discussed so far act only when the two interacting objects actually touch.

The American physicist and statesman Benjamin Franklin found that he could concentrate charge in a “Leyden jar,” which was essentially a glass jar with two sheets of metal foil, one inside and one outside, with the glass between them ([\[link\]](#)). This created a large electric force between the two foil sheets.



A Leyden jar (an early version of what is now called a capacitor) allowed experimenters to store large amounts of electric charge. Benjamin Franklin used such a jar to demonstrate that lightning behaved exactly like the electricity he got from the equipment in his laboratory.

Franklin pointed out that the observed behavior could be explained by supposing that one of the two types of charge remained motionless, while the other type of charge flowed from one piece of foil to the other. He further suggested that an excess of what he called this “electrical fluid” be called “positive electricity” and the deficiency of it be called “negative electricity.” His suggestion, with some minor modifications, is the model we use today. (With the experiments that he was able to do, this was a pure guess; he had no way of actually determining the sign of the moving charge. Unfortunately, he guessed wrong; we now know that the charges that flow are the ones Franklin labeled negative, and the positive charges remain largely motionless. Fortunately, as we’ll see, it makes no practical or theoretical difference which choice we make, as long as we stay consistent with our choice.)

Let’s list the specific observations that we have of this **electric force**:

- The force acts without physical contact between the two objects.
- The force can be either attractive or repulsive: If two interacting objects carry the same sign of charge, the force is repulsive; if the charges are of opposite sign, the force is attractive. These interactions are referred to as **electrostatic repulsion** and **electrostatic attraction**, respectively.
- Not all objects are affected by this force.
- The magnitude of the force decreases (rapidly) with increasing separation distance between the objects.

To be more precise, we find experimentally that the magnitude of the force decreases as the square of the distance between the two interacting objects increases. Thus, for example, when the distance between two interacting

objects is doubled, the force between them decreases to one fourth what it was in the original system. We can also observe that the surroundings of the charged objects affect the magnitude of the force. However, we will explore this issue in a later chapter.

Properties of Electric Charge

In addition to the existence of two types of charge, several other properties of charge have been discovered.

- **Charge is quantized.** This means that electric charge comes in discrete amounts, and there is a smallest possible amount of charge that an object can have. In the SI system, this smallest amount is $e \equiv 1.602 \times 10^{-19} \text{ C}$. No free particle can have less charge than this, and, therefore, the charge on any object—the charge on all objects—must be an integer multiple of this amount. All macroscopic, charged objects have charge because electrons have either been added or taken away from them, resulting in a net charge.
- **The magnitude of the charge is independent of the type.** Phrased another way, the smallest possible positive charge (to four significant figures) is $+1.602 \times 10^{-19} \text{ C}$, and the smallest possible negative charge is $-1.602 \times 10^{-19} \text{ C}$; these values are exactly equal. This is simply how the laws of physics in our universe turned out.
- **Charge is conserved.** Charge can neither be created nor destroyed; it can only be transferred from place to place, from one object to another. Frequently, we speak of two charges “canceling”; this is verbal shorthand. It means that if two objects that have equal and opposite charges are physically close to each other, then the (oppositely directed) forces they apply on some other charged object cancel, for a net force of zero. It is important that you understand that the charges on the objects by no means disappear, however. The net charge of the universe is constant.
- **Charge is conserved in closed systems.** In principle, if a negative charge disappeared from your lab bench and reappeared on the Moon, conservation of charge would still hold. However, this never happens. If the total charge you have in your local system on your lab bench is changing, there will be a measurable flow of charge into or out of the

system. Again, charges can and do move around, and their effects can and do cancel, but the net charge in your local environment (if closed) is conserved. The last two items are both referred to as the **law of conservation of charge**.

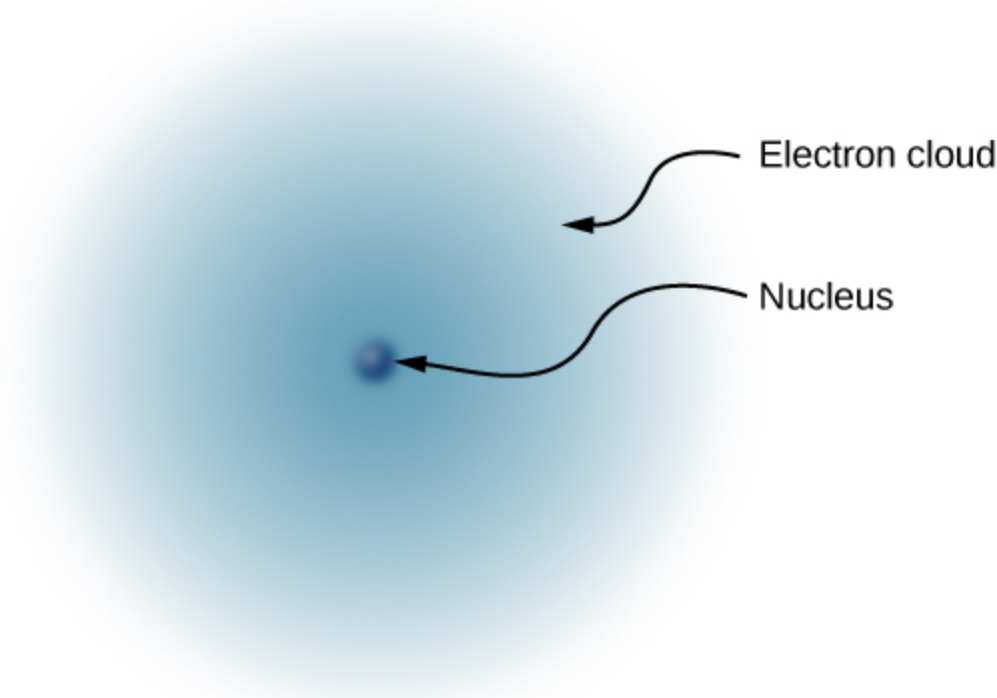
The Source of Charges: The Structure of the Atom

Once it became clear that all matter was composed of particles that came to be called atoms, it also quickly became clear that the constituents of the atom included both positively charged particles and negatively charged particles. The next question was, what are the physical properties of those electrically charged particles?

The negatively charged particle was the first one to be discovered. In 1897, the English physicist J. J. Thomson was studying what was then known as *cathode rays*. Some years before, the English physicist William Crookes had shown that these “rays” were negatively charged, but his experiments were unable to tell any more than that. (The fact that they carried a negative electric charge was strong evidence that these were not rays at all, but particles.) Thomson prepared a pure beam of these particles and sent them through crossed electric and magnetic fields, and adjusted the various field strengths until the net deflection of the beam was zero. With this experiment, he was able to determine the charge-to-mass ratio of the particle. This ratio showed that the mass of the particle was much smaller than that of any other previously known particle—1837 times smaller, in fact. Eventually, this particle came to be called the **electron**.

Since the atom as a whole is electrically neutral, the next question was to determine how the positive and negative charges are distributed within the atom. Thomson himself imagined that his electrons were embedded within a sort of positively charged paste, smeared out throughout the volume of the atom. However, in 1908, the New Zealand physicist Ernest Rutherford showed that the positive charges of the atom existed within a tiny core—called a nucleus—that took up only a very tiny fraction of the overall volume of the atom, but held over 99% of the mass. (See [Linear Momentum and Collisions](#).) In addition, he showed that the negatively charged electrons perpetually orbited about this nucleus, forming a sort of

electrically charged cloud that surrounds the nucleus ([link](#)). Rutherford concluded that the nucleus was constructed of small, massive particles that he named **protons**.

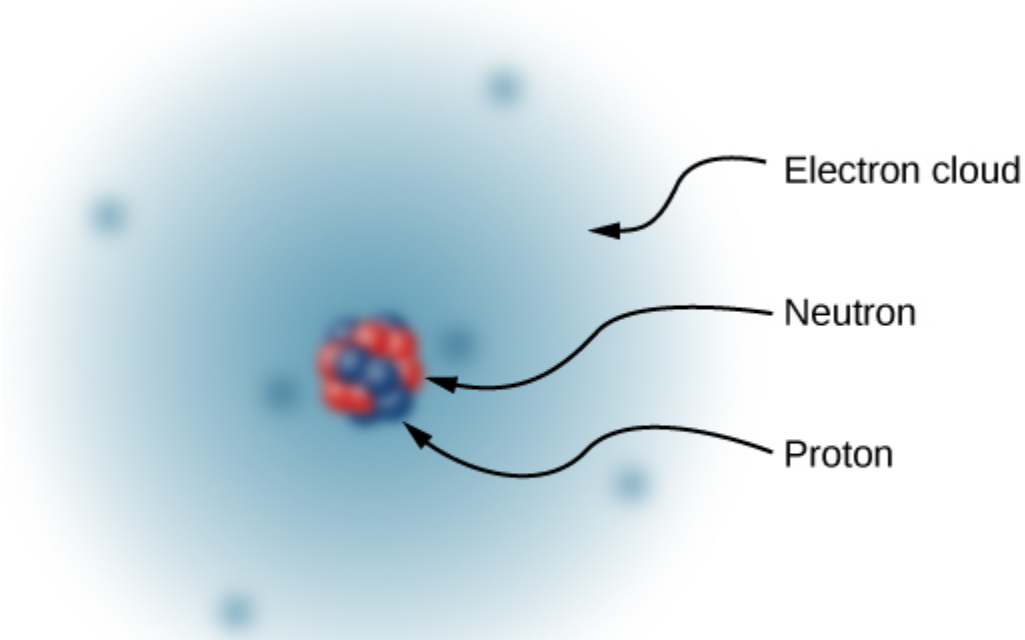


This simplified model of a hydrogen atom shows a positively charged nucleus (consisting, in the case of hydrogen, of a single proton), surrounded by an electron “cloud.” The charge of the electron cloud is equal (and opposite in sign) to the charge of the nucleus, but the electron does not have a definite location in space; hence, its representation here is as a cloud. Normal macroscopic amounts of matter contain immense numbers of atoms and molecules, and, hence, even greater numbers of individual negative and positive charges.

Since it was known that different atoms have different masses, and that ordinarily atoms are electrically neutral, it was natural to suppose that different atoms have different numbers of protons in their nucleus, with an equal number of negatively charged electrons orbiting about the positively charged nucleus, thus making the atoms overall electrically neutral. However, it was soon discovered that although the lightest atom, hydrogen, did indeed have a single proton as its nucleus, the next heaviest atom—helium—has twice the number of protons (two), but *four* times the mass of hydrogen.

This mystery was resolved in 1932 by the English physicist James Chadwick, with the discovery of the **neutron**. The neutron is, essentially, an electrically neutral twin of the proton, with no electric charge, but (nearly) identical mass to the proton. The helium nucleus therefore has two neutrons along with its two protons. (Later experiments were to show that although the neutron is electrically neutral overall, it does have an internal charge *structure*. Furthermore, although the masses of the neutron and the proton are *nearly* equal, they aren't exactly equal: The neutron's mass is very slightly larger than the mass of the proton. That slight mass excess turned out to be of great importance. That, however, is a story that will have to wait until our study of modern physics in [Nuclear Physics](#).)

Thus, in 1932, the picture of the atom was of a small, massive nucleus constructed of a combination of protons and neutrons, surrounded by a collection of electrons whose combined motion formed a sort of negatively charged “cloud” around the nucleus ([\[link\]](#)). In an electrically neutral atom, the total negative charge of the collection of electrons is equal to the total positive charge in the nucleus. The very low-mass electrons can be more or less easily removed or added to an atom, changing the net charge on the atom (though without changing its type). An atom that has had the charge altered in this way is called an **ion**. Positive ions have had electrons removed, whereas negative ions have had excess electrons added. We also use this term to describe molecules that are not electrically neutral.



The nucleus of a carbon atom is composed of six protons and six neutrons. As in hydrogen, the surrounding six electrons do not have definite locations and so can be considered to be a sort of cloud surrounding the nucleus.

The story of the atom does not stop there, however. In the latter part of the twentieth century, many more subatomic particles were discovered in the nucleus of the atom: pions, neutrinos, and quarks, among others. With the exception of the photon, none of these particles are directly relevant to the study of electromagnetism, so we defer further discussion of them until the chapter on particle physics ([Particle Physics and Cosmology](#)).

A Note on Terminology

As noted previously, electric charge is a property that an object can have. This is similar to how an object can have a property that we call mass, a property that we call density, a property that we call temperature, and so on.

Technically, we should always say something like, “Suppose we have a particle that carries a charge of $3\ \mu\text{C}$.” However, it is very common to say instead, “Suppose we have a $3\text{-}\mu\text{C}$ charge.” Similarly, we often say something like, “Six charges are located at the vertices of a regular hexagon.” A charge is not a particle; rather, it is a *property* of a particle. Nevertheless, this terminology is extremely common (and is frequently used in this book, as it is everywhere else). So, keep in the back of your mind what we really mean when we refer to a “charge.”

Summary

- There are only two types of charge, which we call positive and negative. Like charges repel, unlike charges attract, and the force between charges decreases with the square of the distance.
- The vast majority of positive charge in nature is carried by protons, whereas the vast majority of negative charge is carried by electrons. The electric charge of one electron is equal in magnitude and opposite in sign to the charge of one proton.
- An ion is an atom or molecule that has nonzero total charge due to having unequal numbers of electrons and protons.
- The SI unit for charge is the coulomb (C), with protons and electrons having charges of opposite sign but equal magnitude; the magnitude of this basic charge is $e \equiv 1.602 \times 10^{-19}\ \text{C}$
- Both positive and negative charges exist in neutral objects and can be separated by bringing the two objects into physical contact; rubbing the objects together can remove electrons from the bonds in one object and place them on the other object, increasing the charge separation.
- For macroscopic objects, negatively charged means an excess of electrons and positively charged means a depletion of electrons.
- The law of conservation of charge states that the net charge of a closed system is constant.

Conceptual Questions

Exercise:

Problem:

There are very large numbers of charged particles in most objects. Why, then, don't most objects exhibit static electricity?

Solution:

There are mostly equal numbers of positive and negative charges present, making the object electrically neutral.

Exercise:**Problem:**

Why do most objects tend to contain nearly equal numbers of positive and negative charges?

Exercise:**Problem:**

A positively charged rod attracts a small piece of cork. (a) Can we conclude that the cork is negatively charged? (b) The rod repels another small piece of cork. Can we conclude that this piece is positively charged?

Solution:

a. no; b. yes

Exercise:**Problem:**

Two bodies attract each other electrically. Do they both have to be charged? Answer the same question if the bodies repel one another.

Exercise:

Problem:

How would you determine whether the charge on a particular rod is positive or negative?

Solution:

Take an object with a known charge, either positive or negative, and bring it close to the rod. If the known charged object is positive and it is repelled from the rod, the rod is charged positive. If the positively charged object is attracted to the rod, the rod is negatively charged.

Problems**Exercise:****Problem:**

Common static electricity involves charges ranging from nanocoulombs to microcoulombs. (a) How many electrons are needed to form a charge of -2.00 nC ? (b) How many electrons must be removed from a neutral object to leave a net charge of $0.500 \mu\text{C}$?

Solution:

- a. $2.00 \times 10^{-9} \text{ C} \left(\frac{1}{1.602 \times 10^{-19}} \text{ e/C} \right) = 1.248 \times 10^{10} \text{ electrons};$
b. $0.500 \times 10^{-6} \text{ C} \left(\frac{1}{1.602 \times 10^{-19}} \text{ e/C} \right) = 3.121 \times 10^{12} \text{ electrons}$

Exercise:**Problem:**

If 1.80×10^{20} electrons move through a pocket calculator during a full day's operation, how many coulombs of charge moved through it?

Exercise:

Problem:

To start a car engine, the car battery moves 3.75×10^{21} electrons through the starter motor. How many coulombs of charge were moved?

Solution:

$$\frac{3.750 \times 10^{21} \text{ e}}{6.242 \times 10^{18} \text{ e/C}} = -600.8 \text{ C}$$

Exercise:**Problem:**

A certain lightning bolt moves 40.0 C of charge. How many fundamental units of charge is this?

Exercise:**Problem:**

A 2.5-g copper penny is given a charge of -2.0×10^{-9} C. (a) How many excess electrons are on the penny? (b) By what percent do the excess electrons change the mass of the penny?

Solution:

a. $2.0 \times 10^{-9} \text{ C} (6.242 \times 10^{18} \text{ e/C}) = 1.248 \times 10^{10} \text{ e};$

b. $9.109 \times 10^{-31} \text{ kg} (1.248 \times 10^{10} \text{ e}) = 1.137 \times 10^{-20} \text{ kg},$
 $\frac{1.137 \times 10^{-20} \text{ kg}}{2.5 \times 10^{-3} \text{ kg}} = 4.548 \times 10^{-18} \text{ or } 4.545 \times 10^{-16} \%$

Exercise:

Problem:

A 2.5-g copper penny is given a charge of $4.0 \times 10^{-9} \text{ C}$. (a) How many electrons are removed from the penny? (b) If no more than one electron is removed from an atom, what percent of the atoms are ionized by this charging process?

Glossary

coulomb

SI unit of electric charge

electric charge

physical property of an object that causes it to be attracted toward or repelled from another charged object; each charged object generates and is influenced by a force called an electric force

electric force

noncontact force observed between electrically charged objects

electron

particle surrounding the nucleus of an atom and carrying the smallest unit of negative charge

electrostatic attraction

phenomenon of two objects with opposite charges attracting each other

electrostatic repulsion

phenomenon of two objects with like charges repelling each other

ion

atom or molecule with more or fewer electrons than protons

law of conservation of charge

net electric charge of a closed system is constant

neutron

neutral particle in the nucleus of an atom, with (nearly) the same mass as a proton

proton

particle in the nucleus of an atom and carrying a positive charge equal in magnitude to the amount of negative charge carried by an electron

static electricity

buildup of electric charge on the surface of an object; the arrangement of the charge remains constant (“static”)

Conductors, Insulators, and Charging by Induction

By the end of this section, you will be able to:

- Explain what a conductor is
- Explain what an insulator is
- List the differences and similarities between conductors and insulators
- Describe the process of charging by induction

In the preceding section, we said that scientists were able to create electric charge only on nonmetallic materials and never on metals. To understand why this is the case, you have to understand more about the nature and structure of atoms. In this section, we discuss how and why electric charges do—or do not—move through materials ([link](#)). A more complete description is given in a later chapter.



This power adapter uses metal wires and connectors to conduct electricity from the wall socket to a laptop computer. The conducting wires allow electrons to move freely through the cables, which are shielded by rubber and plastic. These materials act as insulators that don't allow electric charge to escape outward. (credit: modification of work by “Evan-Amos”/Wikimedia Commons)

Conductors and Insulators

As discussed in the previous section, electrons surround the tiny nucleus in the form of a (comparatively) vast cloud of negative charge. However, this cloud does have a definite structure to it. Let's consider an atom of the most commonly used conductor, copper.

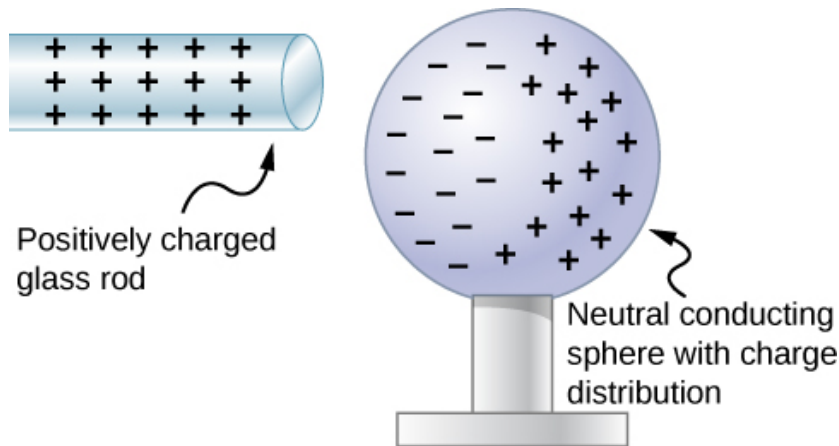
For reasons that will become clear in [Atomic Structure](#), there is an outermost electron that is only loosely bound to the atom's nucleus. It can be easily dislodged; it then moves to a neighboring atom. In a large mass of copper atoms (such as a copper wire or a sheet of copper), these vast numbers of outermost electrons (one per atom) wander from atom to atom, and are the electrons that do the moving when electricity flows. These wandering, or “free,” electrons are called **conduction electrons**, and copper is therefore an excellent **conductor** (of electric charge). All conducting elements have a similar arrangement of their electrons, with one or two conduction electrons. This includes most metals.

Insulators, in contrast, are made from materials that lack conduction electrons; charge flows only with great difficulty, if at all. Even if excess charge is added to an insulating material, it cannot move, remaining indefinitely in place. This is why insulating materials exhibit the electrical attraction and repulsion forces described earlier, whereas conductors do not; any excess charge placed on a conductor would instantly flow away (due to mutual repulsion from existing charges), leaving no excess charge around to create forces. Charge cannot flow along or through an **insulator**, so its electric forces remain for long periods of time. (Charge will dissipate from an insulator, given enough time.) As it happens, amber, fur, and most semi-precious gems are insulators, as are materials like wood, glass, and plastic.

Charging by Induction

Let's examine in more detail what happens in a conductor when an electrically charged object is brought close to it. As mentioned, the conduction electrons in the conductor are able to move with nearly complete freedom. As a result, when a charged insulator (such as a positively charged glass rod) is brought close to the conductor, the (total) charge on the insulator exerts an electric force on the conduction electrons. Since the rod is positively charged, the conduction electrons (which themselves are negatively charged) are attracted, flowing toward the insulator to the near side of the conductor ([link](#)).

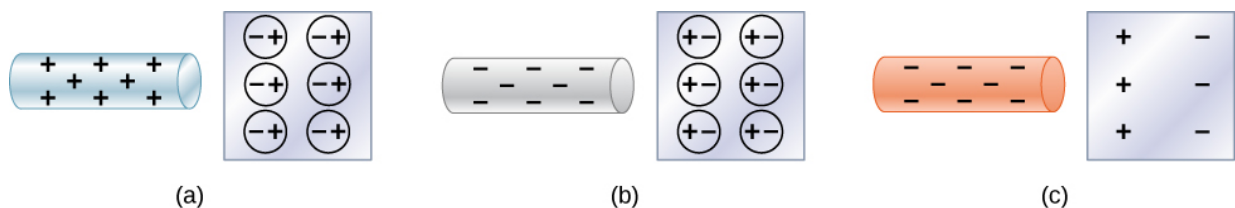
Now, the conductor is still overall electrically neutral; the conduction electrons have changed position, but they are still in the conducting material. However, the conductor now has a charge *distribution*; the near end (the portion of the conductor closest to the insulator) now has more negative charge than positive charge, and the reverse is true of the end farthest from the insulator. The relocation of negative charges to the near side of the conductor results in an overall positive charge in the part of the conductor farthest from the insulator. We have thus created an electric charge distribution where one did not exist before. This process is referred to as *inducing polarization*—in this case, polarizing the conductor. The resulting separation of positive and negative charge is called **polarization**, and a material, or even a molecule, that exhibits polarization is said to be polarized. A similar situation occurs with a negatively charged insulator, but the resulting polarization is in the opposite direction.



Induced polarization. A positively charged glass rod is brought near the left side of the conducting sphere, attracting negative charge and leaving the other side of the sphere positively charged. Although the sphere is overall still electrically neutral, it now has a charge distribution, so it can exert an electric force on other nearby charges. Furthermore, the distribution is such that it will be attracted to the glass rod.

The result is the formation of what is called an electric **dipole**, from a Latin phrase meaning “two ends.” The presence of electric charges on the insulator—and the electric forces they apply to the conduction electrons—creates, or “induces,” the dipole in the conductor.

Neutral objects can be attracted to any charged object. The pieces of straw attracted to polished amber are neutral, for example. If you run a plastic comb through your hair, the charged comb can pick up neutral pieces of paper. [\[link\]](#) shows how the polarization of atoms and molecules in neutral objects results in their attraction to a charged object.

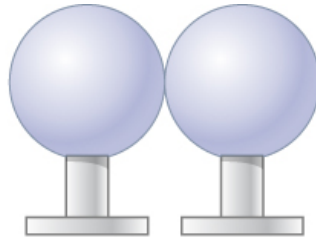


Both positive and negative objects attract a neutral object by polarizing its molecules. (a) A positive object brought near a neutral insulator polarizes its molecules. There is a slight shift in the distribution of the electrons orbiting the

molecule, with unlike charges being brought nearer and like charges moved away. Since the electrostatic force decreases with distance, there is a net attraction. (b) A negative object produces the opposite polarization, but again attracts the neutral object. (c) The same effect occurs for a conductor; since the unlike charges are closer, there is a net attraction.

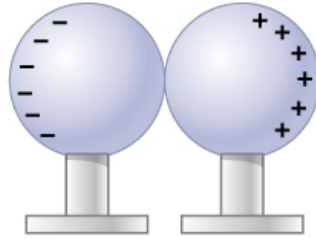
When a charged rod is brought near a neutral substance, an insulator in this case, the distribution of charge in atoms and molecules is shifted slightly. Opposite charge is attracted nearer the external charged rod, while like charge is repelled. Since the electrostatic force decreases with distance, the repulsion of like charges is weaker than the attraction of unlike charges, and so there is a net attraction. Thus, a positively charged glass rod attracts neutral pieces of paper, as will a negatively charged rubber rod. Some molecules, like water, are polar molecules. Polar molecules have a natural or inherent separation of charge, although they are neutral overall. Polar molecules are particularly affected by other charged objects and show greater polarization effects than molecules with naturally uniform charge distributions.

When the two ends of a dipole can be separated, this method of **charging by induction** may be used to create charged objects without transferring charge. In [\[link\]](#), we see two neutral metal spheres in contact with one another but insulated from the rest of the world. A positively charged rod is brought near one of them, attracting negative charge to that side, leaving the other sphere positively charged.



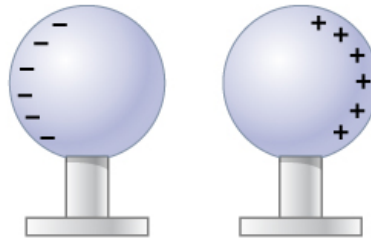
(a)

A charged rod...



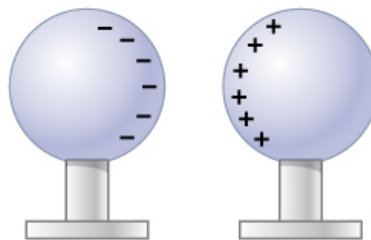
... causes separation of charge

(b)



The spheres are separated.

(c)

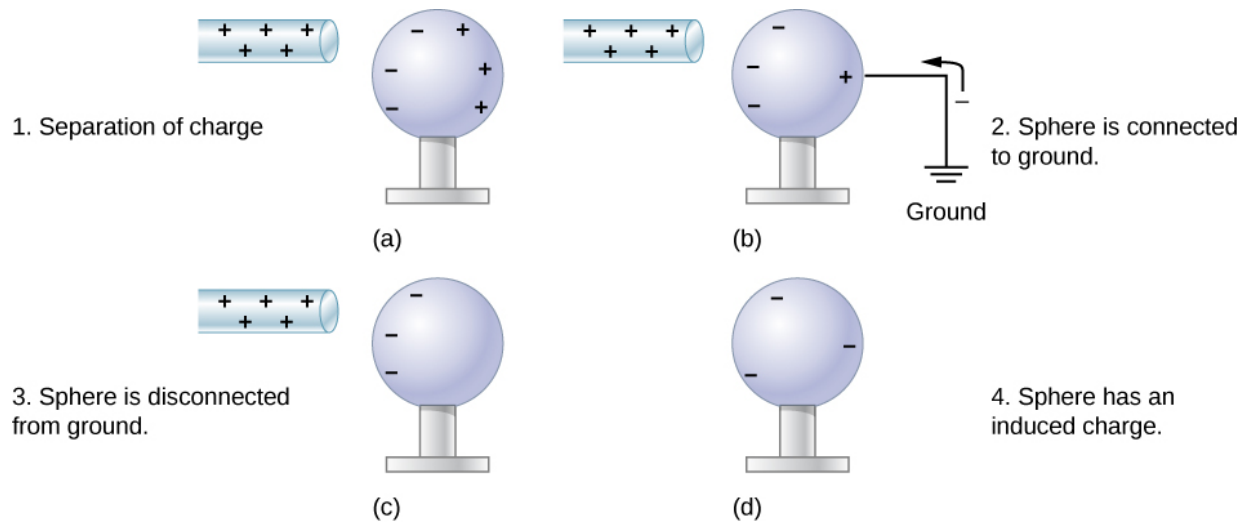


Each sphere is now charged:
one positive, one negative

(d)

Charging by induction. (a) Two uncharged or neutral metal spheres are in contact with each other but insulated from the rest of the world. (b) A positively charged glass rod is brought near the sphere on the left, attracting negative charge and leaving the other sphere positively charged. (c) The spheres are separated before the rod is removed, thus separating negative and positive charges. (d) The spheres retain net charges after the inducing rod is removed—without ever having been touched by a charged object.

Another method of charging by induction is shown in [\[link\]](#). The neutral metal sphere is polarized when a charged rod is brought near it. The sphere is then grounded, meaning that a conducting wire is run from the sphere to the ground. Since Earth is large and most of the ground is a good conductor, it can supply or accept excess charge easily. In this case, electrons are attracted to the sphere through a wire called the ground wire, because it supplies a conducting path to the ground. The ground connection is broken before the charged rod is removed, leaving the sphere with an excess charge opposite to that of the rod. Again, an opposite charge is achieved when charging by induction, and the charged rod loses none of its excess charge.



Charging by induction using a ground connection. (a) A positively charged rod is brought near a neutral metal sphere, polarizing it. (b) The sphere is grounded, allowing electrons to be attracted from Earth's ample supply. (c) The ground connection is broken. (d) The positive rod is removed, leaving the sphere with an induced negative charge.

Summary

- A conductor is a substance that allows charge to flow freely through its atomic structure.
- An insulator holds charge fixed in place.
- Polarization is the separation of positive and negative charges in a neutral object. Polarized objects have their positive and negative charges concentrated in different

areas, giving them a charge distribution.

Conceptual Questions

Exercise:

Problem:

An eccentric inventor attempts to levitate a cork ball by wrapping it with foil and placing a large negative charge on the ball and then putting a large positive charge on the ceiling of his workshop. Instead, while attempting to place a large negative charge on the ball, the foil flies off. Explain.

Exercise:

Problem:

When a glass rod is rubbed with silk, it becomes positive and the silk becomes negative—yet both attract dust. Does the dust have a third type of charge that is attracted to both positive and negative? Explain.

Solution:

No, the dust is attracted to both because the dust particle molecules become polarized in the direction of the silk.

Exercise:

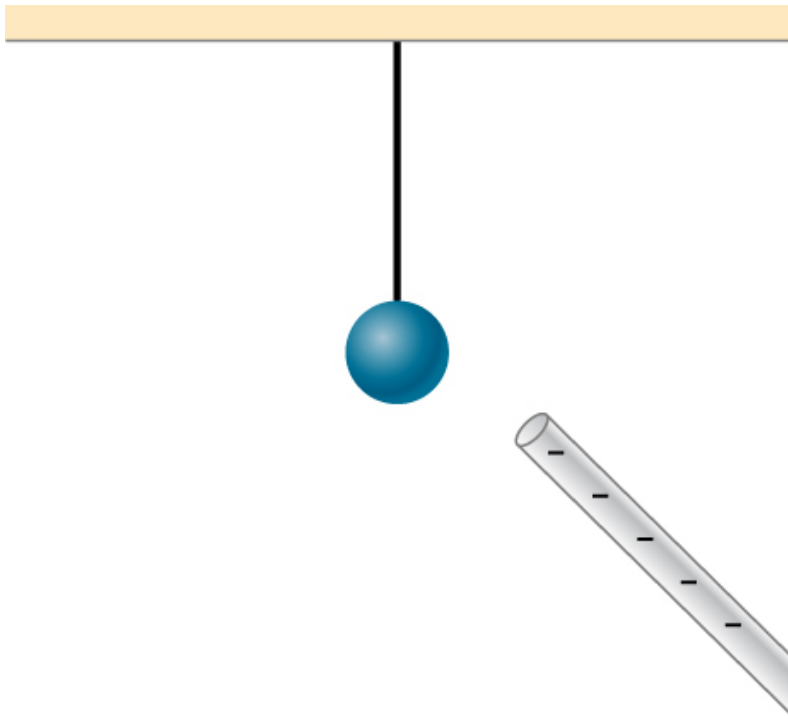
Problem:

Why does a car always attract dust right after it is polished? (Note that car wax and car tires are insulators.)

Exercise:

Problem:

Does the uncharged conductor shown below experience a net electric force?



Solution:

Yes, polarization charge is induced on the conductor so that the positive charge is nearest the charged rod, causing an attractive force.

Exercise:**Problem:**

While walking on a rug, a person frequently becomes charged because of the rubbing between his shoes and the rug. This charge then causes a spark and a slight shock when the person gets close to a metal object. Why are these shocks so much more common on a dry day?

Exercise:

Problem: Compare charging by conduction to charging by induction.

Solution:

Charging by conduction is charging by contact where charge is transferred to the object. Charging by induction first involves producing a polarization charge in the object and then connecting a wire to ground to allow some of the charge to leave the object, leaving the object charged.

Exercise:

Problem:

Small pieces of tissue are attracted to a charged comb. Soon after sticking to the comb, the pieces of tissue are repelled from it. Explain.

Exercise:**Problem:**

Trucks that carry gasoline often have chains dangling from their undercarriages and brushing the ground. Why?

Solution:

This is so that any excess charge is transferred to the ground, keeping the gasoline receptacles neutral. If there is excess charge on the gasoline receptacle, a spark could ignite it.

Exercise:

Problem: Why do electrostatic experiments work so poorly in humid weather?

Exercise:**Problem:**

Why do some clothes cling together after being removed from the clothes dryer? Does this happen if they're still damp?

Solution:

The dryer charges the clothes. If they are damp, the presence of water molecules suppresses the charge.

Exercise:

Problem: Can induction be used to produce charge on an insulator?

Exercise:**Problem:**

Suppose someone tells you that rubbing quartz with cotton cloth produces a third kind of charge on the quartz. Describe what you might do to test this claim.

Solution:

There are only two types of charge, attractive and repulsive. If you bring a charged object near the quartz, only one of these two effects will happen, proving there is not a third kind of charge.

Exercise:

Problem:

A handheld copper rod does not acquire a charge when you rub it with a cloth. Explain why.

Exercise:

Problem:

Suppose you place a charge q near a large metal plate. (a) If q is attracted to the plate, is the plate necessarily charged? (b) If q is repelled by the plate, is the plate necessarily charged?

Solution:

a. No, since a polarization charge is induced. b. Yes, since the polarization charge would produce only an attractive force.

Problems

Exercise:

Problem:

Suppose a speck of dust in an electrostatic precipitator has 1.0000×10^{12} protons in it and has a net charge of -5.00 nC (a very large charge for a small speck). How many electrons does it have?

Solution:

$$5.00 \times 10^{-9} \text{ C} (6.242 \times 10^{18} \text{ e/C}) = 3.121 \times 10^{10} \text{ e};$$
$$3.121 \times 10^{10} \text{ e} + 1.0000 \times 10^{12} \text{ e} = 1.0312 \times 10^{12} \text{ e}$$

Exercise:

Problem:

An amoeba has 1.00×10^{16} protons and a net charge of 0.300 pC. (a) How many fewer electrons are there than protons? (b) If you paired them up, what fraction of the protons would have no electrons?

Exercise:

Problem:

A 50.0-g ball of copper has a net charge of $2.00 \mu\text{C}$. What fraction of the copper's electrons has been removed? (Each copper atom has 29 protons, and copper has an atomic mass of 63.5.)

Solution:

atomic mass of copper atom times $1 \text{ u} = 1.055 \times 10^{-25} \text{ kg}$;

number of copper atoms $= 4.739 \times 10^{23}$ atoms;

number of electrons equals 29 times number of atoms or 1.374×10^{25} electrons;

$$\frac{2.00 \times 10^{-6} \text{ C}(6.242 \times 10^{18} \text{ e/C})}{1.374 \times 10^{25} \text{ e}} = 9.083 \times 10^{-13} \text{ or } 9.083 \times 10^{-11}\%$$

Exercise:**Problem:**

What net charge would you place on a 100-g piece of sulfur if you put an extra electron on 1 in 10^{12} of its atoms? (Sulfur has an atomic mass of 32.1 u.)

Exercise:**Problem:**

How many coulombs of positive charge are there in 4.00 kg of plutonium, given its atomic mass is 244 and that each plutonium atom has 94 protons?

Solution:

$$244.00 \text{ u}(1.66 \times 10^{-27} \text{ kg/u}) = 4.050 \times 10^{-25} \text{ kg};$$

$$\frac{4.00 \text{ kg}}{4.050 \times 10^{-25} \text{ kg}} = 9.877 \times 10^{24} \text{ atoms} \quad 9.877 \times 10^{24}(94) = 9.284 \times 10^{26} \text{ protons}$$

;

$$9.284 \times 10^{26}(1.602 \times 10^{-19} \text{ C/p}) = 1.487 \times 10^8 \text{ C}$$

Glossary

charging by induction

process by which an electrically charged object brought near a neutral object creates a charge separation in that object

conduction electron

electron that is free to move away from its atomic orbit

conductor

material that allows electrons to move separately from their atomic orbits; object with properties that allow charges to move about freely within it

dipole

two equal and opposite charges that are fixed close to each other

insulator

material that holds electrons securely within their atomic orbits

polarization

slight shifting of positive and negative charges to opposite sides of an object

Coulomb's Law

By the end of this section, you will be able to:

- Describe the electric force, both qualitatively and quantitatively
- Calculate the force that charges exert on each other
- Determine the direction of the electric force for different source charges
- Correctly describe and apply the superposition principle for multiple source charges

Experiments with electric charges have shown that if two objects each have electric charge, then they exert an electric force on each other. The magnitude of the force is linearly proportional to the net charge on each object and inversely proportional to the square of the distance between them. (Interestingly, the force does not depend on the mass of the objects.) The direction of the force vector is along the imaginary line joining the two objects and is dictated by the signs of the charges involved.

Let

- q_1, q_2 = the net electric charges of the two objects;
- \vec{r}_{12} = the vector displacement from q_1 to q_2 .

The electric force \vec{F} on one of the charges is proportional to the magnitude of its own charge and the magnitude of the other charge, and is inversely proportional to the square of the distance between them:

Equation:

$$F \propto \frac{q_1 q_2}{r_{12}^2}.$$

This proportionality becomes an equality with the introduction of a proportionality constant. For reasons that will become clear in a later chapter, the proportionality constant that we use is actually a collection of constants. (We discuss this constant shortly.)

Note:

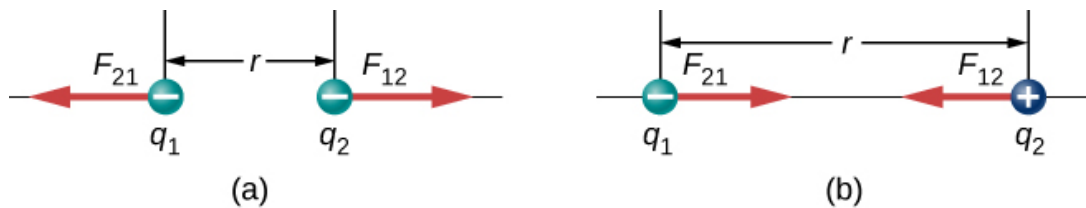
Coulomb's Law

The magnitude of the electric force (or **Coulomb force**) between two electrically charged particles is equal to

Equation:

$$|\mathbf{F}_{12}| = \frac{1}{4\pi\epsilon_0} \frac{|q_1 q_2|}{r_{12}^2}$$

The unit vector \hat{r} has a magnitude of 1 and points along the axis as the charges. If the charges have the same sign, the force is in the same direction as \hat{r} showing a repelling force. If the charges have different signs, the force is in the opposite direction of \hat{r} showing an attracting force. ([\[link\]](#)).



The electrostatic force \vec{F} between point charges q_1 and q_2 separated by a distance r is given by Coulomb's law. Note that Newton's third law (every force exerted creates an equal and opposite force) applies as usual—the force on q_1 is equal in magnitude and opposite in direction to the force it exerts on q_2 . (a) Like charges; (b) unlike charges.

It is important to note that the electric force is not constant; it is a function of the separation distance between the two charges. If either the test charge or the source charge (or both) move, then \vec{r} changes, and therefore so does the force. An immediate consequence of this is that direct application of Newton's laws with this force can be mathematically difficult, depending on the specific problem at hand. It can (usually) be done, but we almost always look for easier methods of calculating whatever physical quantity we are interested in. (Conservation of energy is the most common choice.)

Finally, the new constant ϵ_0 in Coulomb's law is called the *permittivity of free space*, or (better) the **permittivity of vacuum**. It has a very important physical meaning that we will discuss in a later chapter; for now, it is simply an empirical proportionality constant. Its numerical value (to three significant figures) turns out to be

Equation:

$$\epsilon_0 = 8.85 \times 10^{-12} \frac{\text{C}^2}{\text{N} \cdot \text{m}^2}.$$

These units are required to give the force in Coulomb's law the correct units of newtons. Note that in Coulomb's law, the permittivity of vacuum is only part of the proportionality constant. For convenience, we often define a Coulomb's constant:

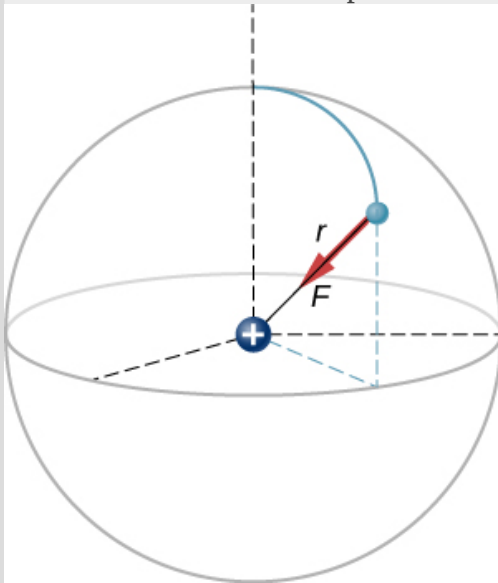
Equation:

$$k_e = \frac{1}{4\pi\epsilon_0} = 8.99 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2}.$$

Example:

The Force on the Electron in Hydrogen

A hydrogen atom consists of a single proton and a single electron. The proton has a charge of $+e$ and the electron has $-e$. In the “ground state” of the atom, the electron orbits the proton at most probable distance of $5.29 \times 10^{-11} \text{ m}$ ([link](#)). Calculate the electric force on the electron due to the proton.



A schematic depiction of a hydrogen atom, showing the force on the electron. This depiction is only to enable us to calculate the force; the hydrogen atom does not really look like this. Recall [link](#).

Strategy

For the purposes of this example, we are treating the electron and proton as two point particles, each with an electric charge, and we are told the distance between them; we are asked to calculate the force on the electron. We thus use Coulomb’s law.

Solution

Our two charges and the distance between them are,

Equation:

$$\begin{aligned}q_1 &= +e = +1.602 \times 10^{-19} \text{ C} \\q_2 &= -e = -1.602 \times 10^{-19} \text{ C} \\r &= 5.29 \times 10^{-11} \text{ m}.\end{aligned}$$

The magnitude of the force on the electron is

Equation:

$$F = \frac{1}{4\pi\epsilon_0} \frac{|e|^2}{r^2} = \frac{1}{4\pi \left(8.85 \times 10^{-12} \frac{\text{C}^2}{\text{N}\cdot\text{m}^2} \right)} \frac{(1.602 \times 10^{-19} \text{ C})^2}{(5.29 \times 10^{-11} \text{ m})^2} = 8.25 \times 10^{-8} \text{ N}.$$

As for the direction, since the charges on the two particles are opposite, the force is attractive; the force on the electron points radially directly toward the proton, everywhere in the electron's orbit. The force is thus expressed as

Equation:

$$\vec{\mathbf{F}} = (8.25 \times 10^{-8} \text{ N}) \hat{\mathbf{r}}.$$

Significance

This is a three-dimensional system, so the electron (and therefore the force on it) can be anywhere in an imaginary spherical shell around the proton. In this “classical” model of the hydrogen atom, the electrostatic force on the electron points in the inward centripetal direction, thus maintaining the electron's orbit. But note that the quantum mechanical model of hydrogen (discussed in [Quantum Mechanics](#)) is utterly different.

Note:

Exercise:

Problem:

Check Your Understanding What would be different if the electron also had a positive charge?

Solution:

The force would point outward.

Multiple Source Charges

The analysis that we have done for two particles can be extended to an arbitrary number of particles; we simply repeat the analysis, two charges at a time. Specifically, we ask the

question: Given N charges (which we refer to as source charge), what is the net electric force that they exert on some other point charge (which we call the test charge)? Note that we use these terms because we can think of the test charge being used to test the strength of the force provided by the source charges.

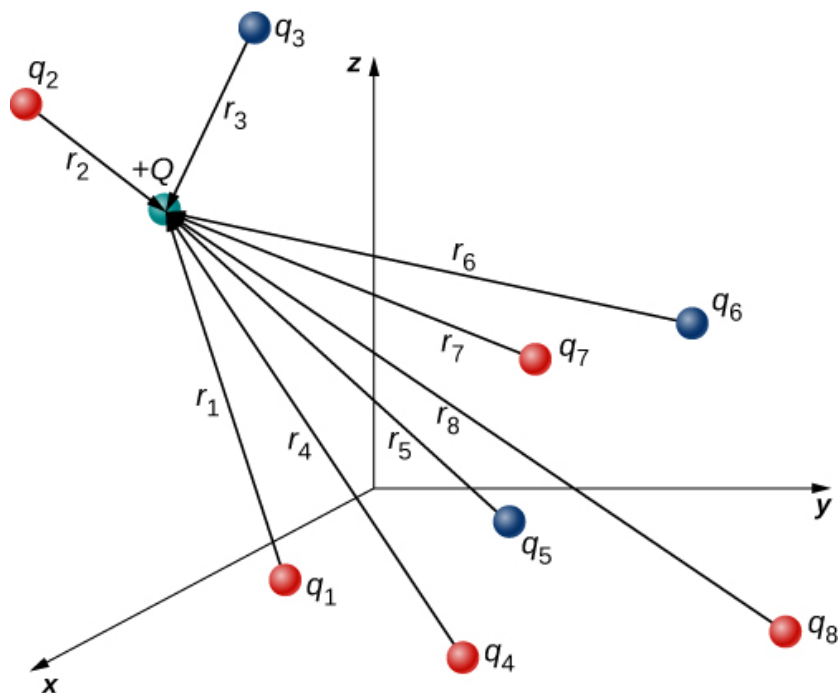
Like all forces that we have seen up to now, the net electric force on our test charge is simply the vector sum of each individual electric force exerted on it by each of the individual source charges. Thus, we can calculate the net force on the test charge Q by calculating the force on it from each source charge, taken one at a time, and then adding all those forces together (as vectors). This ability to simply add up individual forces in this way is referred to as the **principle of superposition**, and is one of the more important features of the electric force. In mathematical form, this becomes

Note:

Equation:

$$\vec{\mathbf{F}}(r) = \frac{1}{4\pi\epsilon_0} Q \sum_{i=1}^N \frac{q_i}{r_i^2} \hat{\mathbf{r}}_i.$$

In this expression, Q represents the charge of the particle that is experiencing the electric force $\vec{\mathbf{F}}$, and is located at $\vec{\mathbf{r}}$ from the origin; the q_i 's are the N source charges, and the vectors $\vec{\mathbf{r}}_i = r_i \hat{\mathbf{r}}_i$ are the displacements from the position of the i th charge to the position of Q . Each of the N unit vectors points directly from its associated source charge toward the test charge. All of this is depicted in [\[link\]](#). Please note that there is no physical difference between Q and q_i ; the difference in labels is merely to allow clear discussion, with Q being the charge we are determining the force on.



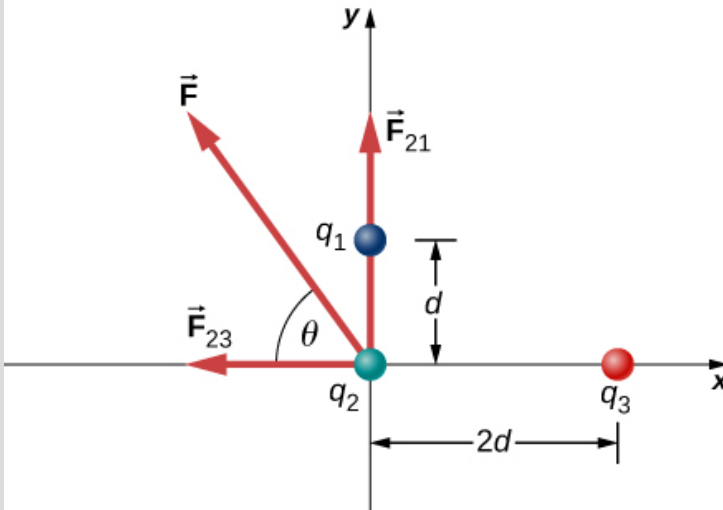
The eight source charges each apply a force on the single test charge Q . Each force can be calculated independently of the other seven forces. This is the essence of the superposition principle.

(Note that the force vector $\vec{\mathbf{F}}_i$ does not necessarily point in the same direction as the unit vector $\hat{\mathbf{r}}_i$; it may point in the opposite direction, $-\hat{\mathbf{r}}_i$. The signs of the source charge and test charge determine the direction of the force on the test charge.)

There is a complication, however. Just as the source charges each exert a force on the test charge, so too (by Newton's third law) does the test charge exert an equal and opposite force on each of the source charges. As a consequence, each source charge would change position. However, by [link](#), the force on the test charge is a function of position; thus, as the positions of the source charges change, the net force on the test charge necessarily changes, which changes the force, which again changes the positions. Thus, the entire mathematical analysis quickly becomes intractable. Later, we will learn techniques for handling this situation, but for now, we make the simplifying assumption that the source charges are fixed in place somehow, so that their positions are constant in time. (The test charge is allowed to move.) With this restriction in place, the analysis of charges is known as **electrostatics**, where “statics” refers to the constant (that is, static) positions of the source charges and the force is referred to as an **electrostatic force**.

Example:**The Net Force from Two Source Charges**

Three different, small charged objects are placed as shown in [\[link\]](#). The charges q_1 and q_3 are fixed in place; q_2 is free to move. Given $q_1 = 2e$, $q_2 = -3e$, and $q_3 = -5e$, and that $d = 2.0 \times 10^{-7} \text{ m}$, what is the net force on the middle charge q_2 ?



Source charges q_1 and q_3 each apply a force on q_2 .

Strategy

We use Coulomb's law again. The way the question is phrased indicates that q_2 is our test charge, so that q_1 and q_3 are source charges. The principle of superposition says that the force on q_2 from each of the other charges is unaffected by the presence of the other charge. Therefore, we write down the force on q_2 from each and add them together as vectors.

Solution

We have two source charges (q_1 and q_3), a test charge (q_2), distances (r_{21} and r_{23}), and we are asked to find a force. This calls for Coulomb's law and superposition of forces.

There are two forces:

Equation:

$$\vec{\mathbf{F}} = \vec{\mathbf{F}}_{21} + \vec{\mathbf{F}}_{23} = \frac{1}{4\pi\epsilon_0} \left[\frac{q_2 q_1}{r_{21}^2} \hat{\mathbf{j}} + \left(-\frac{q_2 q_3}{r_{23}^2} \hat{\mathbf{i}} \right) \right].$$

We can't add these forces directly because they don't point in the same direction: $\vec{\mathbf{F}}_{23}$ points only in the $-x$ -direction, while $\vec{\mathbf{F}}_{21}$ points only in the $+y$ -direction. The net force is obtained from applying the Pythagorean theorem to its x - and y -components:

Equation:

$$F = \sqrt{F_x^2 + F_y^2}$$

where

Equation:

$$\begin{aligned} F_x &= -F_{23} = -\frac{1}{4\pi\epsilon_0} \frac{q_2 q_3}{r_{23}^2} \\ &= -\left(8.99 \times 10^9 \frac{\text{N}\cdot\text{m}^2}{\text{C}^2}\right) \frac{(4.806 \times 10^{-19} \text{ C})(8.01 \times 10^{-19} \text{ C})}{(4.00 \times 10^{-7} \text{ m})^2} \\ &= -2.16 \times 10^{-14} \text{ N} \end{aligned}$$

and

Equation:

$$\begin{aligned} F_y &= F_{21} = \frac{1}{4\pi\epsilon_0} \frac{q_2 q_1}{r_{21}^2} \\ &= \left(8.99 \times 10^9 \frac{\text{N}\cdot\text{m}^2}{\text{C}^2}\right) \frac{(4.806 \times 10^{-19} \text{ C})(3.204 \times 10^{-19} \text{ C})}{(2.00 \times 10^{-7} \text{ m})^2} \\ &= 3.46 \times 10^{-14} \text{ N}. \end{aligned}$$

We find that

Equation:

$$F = \sqrt{F_x^2 + F_y^2} = 4.08 \times 10^{-14} \text{ N}$$

at an angle of

Equation:

$$\phi = \tan^{-1} \left(\frac{F_y}{F_x} \right) = \tan^{-1} \left(\frac{3.46 \times 10^{-14} \text{ N}}{-2.16 \times 10^{-14} \text{ N}} \right) = -58^\circ,$$

that is, 58° above the $-x$ -axis, as shown in the diagram.

Significance

Notice that when we substituted the numerical values of the charges, we did not include the negative sign of either q_2 or q_3 . Recall that negative signs on vector quantities indicate a reversal of direction of the vector in question. But for electric forces, the direction of the force is determined by the types (signs) of both interacting charges; we determine the force directions by considering whether the signs of the two charges are the same or are opposite. If you also include negative signs from negative charges when you substitute numbers, you run the risk of mathematically reversing the direction of the force you are calculating. Thus, the safest thing to do is to calculate just the magnitude of the force, using the absolute values of the charges, and determine the directions physically.

It's also worth noting that the only new concept in this example is how to calculate the electric forces; everything else (getting the net force from its components, breaking the

forces into their components, finding the direction of the net force) is the same as force problems you have done earlier.

Note:

Exercise:

Problem: Check Your Understanding What would be different if q_1 were negative?

Solution:

The net force would point 58° below the $-x$ -axis.

Summary

- Coulomb's law gives the magnitude of the force vector between point charges. It is

Equation:

$$\vec{\mathbf{F}}_{12}(r) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^2} \hat{\mathbf{r}}_{12}$$

where q_1 and q_2 are two point charges separated by a distance r . This Coulomb force is extremely basic, since most charges are due to point-like particles. It is responsible for all electrostatic effects and underlies most macroscopic forces.

Conceptual Questions

Exercise:

Problem:

Would defining the charge on an electron to be positive have any effect on Coulomb's law?

Exercise:

Problem:

An atomic nucleus contains positively charged protons and uncharged neutrons. Since nuclei do stay together, what must we conclude about the forces between these nuclear particles?

Solution:

The force holding the nucleus together must be greater than the electrostatic repulsive force on the protons.

Exercise:

Problem:

Is the force between two fixed charges influenced by the presence of other charges?

Problems

Exercise:

Problem:

Two point particles with charges $+3\ \mu\text{C}$ and $+5\ \mu\text{C}$ are held in place by 3-N forces on each charge in appropriate directions. (a) Draw a free-body diagram for each particle. (b) Find the distance between the charges.

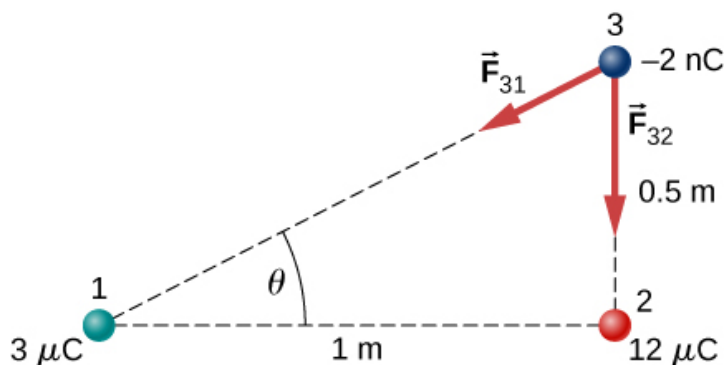
Exercise:

Problem:

Two charges $+3\ \mu\text{C}$ and $+12\ \mu\text{C}$ are fixed 1 m apart, with the second one to the right. Find the magnitude and direction of the net force on a -2-nC charge when placed at the following locations: (a) halfway between the two (b) half a meter to the left of the $+3\ \mu\text{C}$ charge (c) half a meter above the $+12\ \mu\text{C}$ charge in a direction perpendicular to the line joining the two fixed charges

Solution:

- a. charge 1 is $3\ \mu\text{C}$; charge 2 is $12\ \mu\text{C}$, $F_{31} = 2.16 \times 10^{-4}\ \text{N}$ to the left,
 $F_{32} = 8.63 \times 10^{-4}\ \text{N}$ to the right,
 $F_{\text{net}} = 6.47 \times 10^{-4}\ \text{N}$ to the right;
b. $F_{31} = 2.16 \times 10^{-4}\ \text{N}$ to the right,
 $F_{32} = 9.59 \times 10^{-5}\ \text{N}$ to the right,
 $F_{\text{net}} = 3.12 \times 10^{-4}\ \text{N}$ to the right,



;

$$\text{c. } \vec{F}_{31x} = -2.76 \times 10^{-5} \text{ N } \hat{i},$$

$$\vec{F}_{31y} = -1.38 \times 10^{-5} \text{ N } \hat{j},$$

$$\vec{F}_{32y} = -8.63 \times 10^{-4} \text{ N } \hat{j}$$

$$\vec{F}_{\text{net}} = -3.86 \times 10^{-5} \text{ N } \hat{i} - 8.83 \times 10^{-4} \text{ N } \hat{j}$$

Exercise:

Problem:

In a salt crystal, the distance between adjacent sodium and chloride ions is $2.82 \times 10^{-10} \text{ m}$. What is the force of attraction between the two singly charged ions?

Exercise:

Problem:

Protons in an atomic nucleus are typically 10^{-15} m apart. What is the electric force of repulsion between nuclear protons?

Solution:

$$F = 230.7 \text{ N}$$

Exercise:

Problem:

Suppose Earth and the Moon each carried a net negative charge $-Q$. Approximate both bodies as point masses and point charges.

(a) What value of Q is required to balance the gravitational attraction between Earth and the Moon?

(b) Does the distance between Earth and the Moon affect your answer? Explain.

(c) How many electrons would be needed to produce this charge?

Exercise:

Problem:

Point charges $q_1 = 50 \mu\text{C}$ and $q_2 = -25 \mu\text{C}$ are placed 1.0 m apart. What is the force on a third charge $q_3 = 20 \mu\text{C}$ placed midway between q_1 and q_2 ?

Solution:

$$F = 53.94 \text{ N}$$

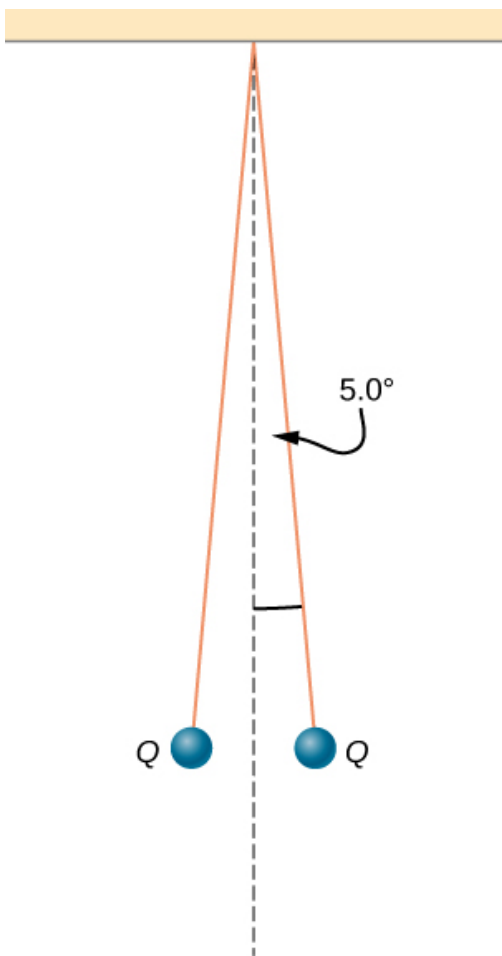
Exercise:

Problem:

Where must q_3 of the preceding problem be placed so that the net force on it is zero?

Exercise:**Problem:**

Two small balls, each of mass 5.0 g, are attached to silk threads 50 cm long, which are in turn tied to the same point on the ceiling, as shown below. When the balls are given the same charge Q , the threads hang at 5.0° to the vertical, as shown below. What is the magnitude of Q ? What are the signs of the two charges?



Solution:

The tension is $T = 0.049$ N. The horizontal component of the tension is 0.0043 N
 $d = 0.088$ m, $q = 6.1 \times 10^{-8}$ C.

The charges can be positive or negative, but both have to be the same sign.

Exercise:

Problem:

Point charges $Q_1 = 2.0 \mu\text{C}$ and $Q_2 = 4.0 \mu\text{C}$ are located at $\vec{r}_1 = (4.0\hat{i} - 2.0\hat{j} + 5.0\hat{k})\text{m}$ and $\vec{r}_2 = (8.0\hat{i} + 5.0\hat{j} - 9.0\hat{k})\text{m}$. What is the force of Q_2 on Q_1 ?

Exercise:**Problem:**

The net excess charge on two small spheres (small enough to be treated as point charges) is Q . Show that the force of repulsion between the spheres is greatest when each sphere has an excess charge $Q/2$. Assume that the distance between the spheres is so large compared with their radii that the spheres can be treated as point charges.

Solution:

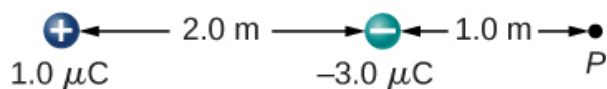
Let the charge on one of the spheres be nQ , where n is a fraction between 0 and 1. In the numerator of Coulomb's law, the term involving the charges is $nQ(1 - n)Q$. This is equal to $(n - n^2)Q^2$. Finding the maximum of this term gives $1 - 2n = 0 \Rightarrow n = \frac{1}{2}$

Exercise:**Problem:**

Two small, identical conducting spheres repel each other with a force of 0.050 N when they are 0.25 m apart. After a conducting wire is connected between the spheres and then removed, they repel each other with a force of 0.060 N. What is the original charge on each sphere?

Exercise:**Problem:**

A charge $q = 2.0 \mu\text{C}$ is placed at the point P shown below. What is the force on q ?

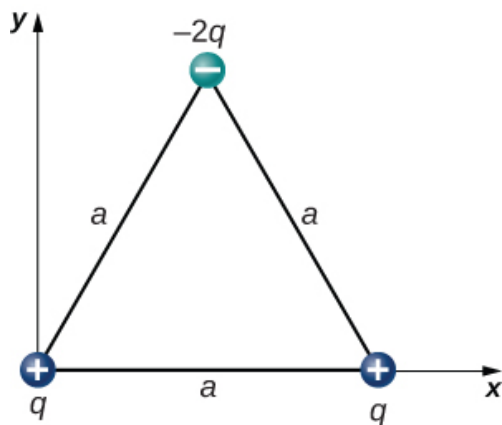
**Solution:**

Define right to be the positive direction and hence left is the negative direction, then $F = -0.05 \text{ N}$

Exercise:

Problem:

What is the net electric force on the charge located at the lower right-hand corner of the triangle shown here?

**Exercise:****Problem:**

Two fixed particles, each of charge $5.0 \times 10^{-6} \text{ C}$, are 24 cm apart. What force do they exert on a third particle of charge $-2.5 \times 10^{-6} \text{ C}$ that is 13 cm from each of them?

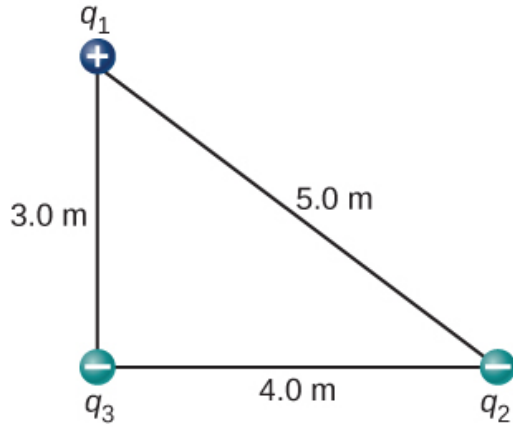
Solution:

The particles form triangle of sides 13, 13, and 24 cm. The x-components cancel, whereas there is a contribution to the y-component from both charges 24 cm apart. The y-axis passing through the third charge bisects the 24-cm line, creating two right triangles of sides 5, 12, and 13 cm.

$F_y = 2.56 \text{ N}$ in the negative y-direction since the force is attractive. The net force from both charges is $\vec{F}_{\text{net}} = -5.12 \text{ N}\hat{j}$.

Exercise:**Problem:**

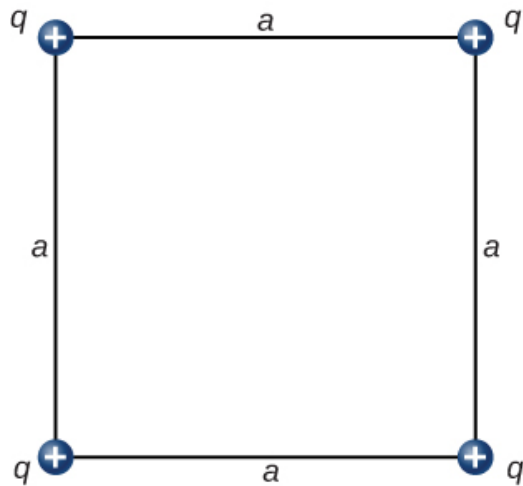
The charges $q_1 = 2.0 \times 10^{-7} \text{ C}$, $q_2 = -4.0 \times 10^{-7} \text{ C}$, and $q_3 = -1.0 \times 10^{-7} \text{ C}$ are placed at the corners of the triangle shown below. What is the force on q_1 ?



Exercise:

Problem:

What is the force on the charge q at the lower-right-hand corner of the square shown here?



Solution:

The diagonal is $\sqrt{2}a$ and the components of the force due to the diagonal charge has a factor $\cos \theta = \frac{1}{\sqrt{2}}$;

$$\vec{F}_{\text{net}} = \left[k \frac{q^2}{a^2} + k \frac{q^2}{2a^2} \frac{1}{\sqrt{2}} \right] \hat{i} - \left[k \frac{q^2}{a^2} + k \frac{q^2}{2a^2} \frac{1}{\sqrt{2}} \right] \hat{j}$$

Exercise:

Problem:

Point charges $q_1 = 10 \mu\text{C}$ and $q_2 = -30 \mu\text{C}$ are fixed at $r_1 = (3.0\hat{\mathbf{i}} - 4.0\hat{\mathbf{j}})\text{m}$ and $r_2 = (9.0\hat{\mathbf{i}} + 6.0\hat{\mathbf{j}})\text{m}$. What is the force of q_2 on q_1 ?

Glossary

Coulomb force

another term for the electrostatic force

Coulomb's law

mathematical equation calculating the electrostatic force vector between two charged particles

electrostatic force

amount and direction of attraction or repulsion between two charged bodies; the assumption is that the source charges have no acceleration

electrostatics

study of charged objects which are not in motion

permittivity of vacuum

also called the permittivity of free space, and constant describing the strength of the electric force in a vacuum

principle of superposition

useful fact that we can simply add up all of the forces due to charges acting on an object

Electric Field

By the end of this section, you will be able to:

- Explain the purpose of the electric field concept
- Describe the properties of the electric field
- Calculate the field of a collection of source charges of either sign

As we showed in the preceding section, the net electric force on a test charge is the vector sum of all the electric forces acting on it, from all of the various source charges, located at their various positions. But what if we use a different test charge, one with a different magnitude, or sign, or both? Or suppose we have a dozen different test charges we wish to try at the same location? We would have to calculate the sum of the forces from scratch. Fortunately, it is possible to define a quantity, called the **electric field**, which is independent of the test charge. It only depends on the configuration of the source charges, and once found, allows us to calculate the force on any test charge.

Defining a Field

Suppose we have N source charges $q_1, q_2, q_3, \dots, q_N$ located at positions $\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_N$, applying N electrostatic forces on a test charge Q . The net force on Q is (see [\[link\]](#))

Equation:

$$\begin{aligned}\vec{F} &= \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \dots + \vec{F}_N \\ &= \frac{1}{4\pi\epsilon_0} \left(\frac{Qq_1}{r_1^2} \hat{r}_1 + \frac{Qq_2}{r_2^2} \hat{r}_2 + \frac{Qq_3}{r_3^2} \hat{r}_3 + \dots + \frac{Qq_N}{r_N^2} \hat{r}_N \right) \\ &= Q \left[\frac{1}{4\pi\epsilon_0} \left(\frac{q_1}{r_1^2} \hat{r}_1 + \frac{q_2}{r_2^2} \hat{r}_2 + \frac{q_3}{r_3^2} \hat{r}_3 + \dots + \frac{q_N}{r_N^2} \hat{r}_N \right) \right].\end{aligned}$$

We can rewrite this as

Note:

Equation:

$$\vec{\mathbf{F}} = Q\vec{\mathbf{E}}$$

where

Equation:

$$\vec{\mathbf{E}} \equiv \frac{1}{4\pi\epsilon_0} \left(\frac{q_1}{r_1^2} \hat{\mathbf{r}}_1 + \frac{q_2}{r_2^2} \hat{\mathbf{r}}_2 + \frac{q_3}{r_3^2} \hat{\mathbf{r}}_3 + \cdots + \frac{q_N}{r_N^2} \hat{\mathbf{r}}_N \right)$$

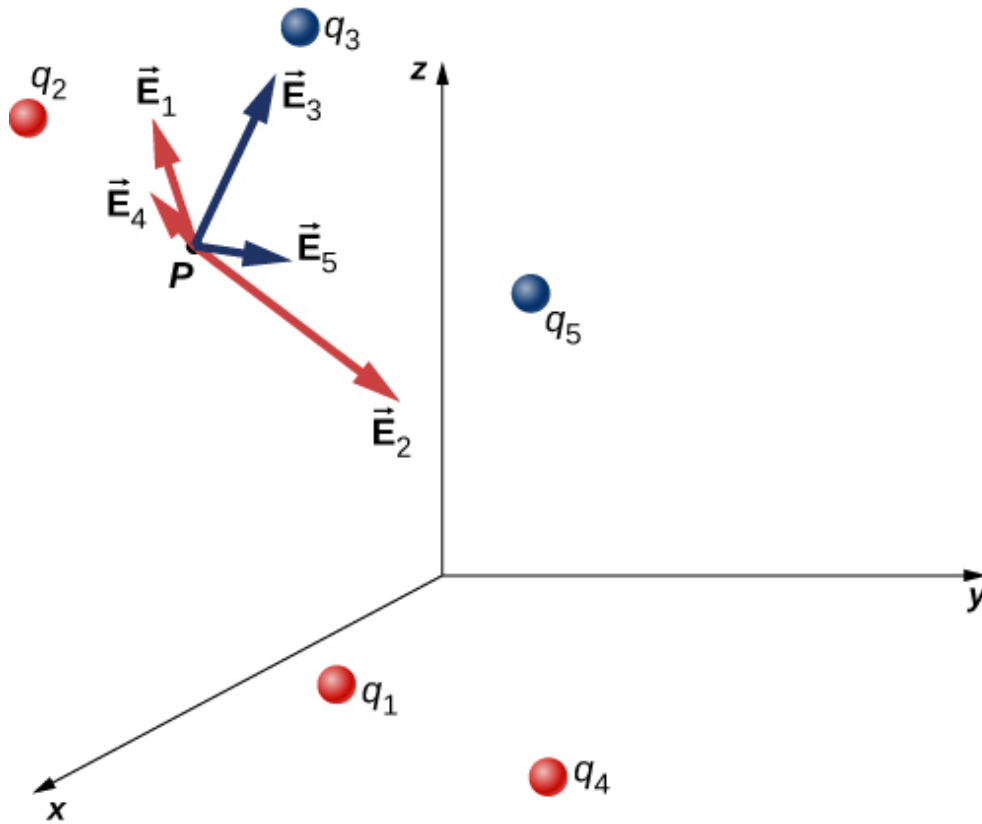
or, more compactly,

Note:

Equation:

$$\vec{\mathbf{E}}(P) \equiv \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N \frac{q_i}{r_i^2} \hat{\mathbf{r}}_i.$$

This expression is called the electric field at position $P = P(x, y, z)$ of the N source charges. Here, P is the location of the point in space where you are calculating the field and is relative to the positions $\vec{\mathbf{r}}_i$ of the source charges ([\[link\]](#)). Note that we have to impose a coordinate system to solve actual problems.



Each of these eight source charges creates its own electric field at every point in space; shown here are the field vectors at an arbitrary point P . Like the electric force, the net electric field obeys the superposition principle.

Notice that the calculation of the electric field makes no reference to the test charge. Thus, the physically useful approach is to calculate the electric field and then use it to calculate the force on some test charge later, if needed. Different test charges experience different forces [\[link\]](#), but it is the same electric field [\[link\]](#). That being said, recall that there is no fundamental difference between a test charge and a source charge; these are merely convenient labels for the system of interest. Any charge produces an electric field; however, just as Earth's orbit is not affected by Earth's own gravity, a charge is not subject to a force due to the electric field it generates. Charges are only subject to forces from the electric fields of other charges.

In this respect, the electric field \vec{E} of a point charge is similar to the gravitational field \vec{g} of Earth; once we have calculated the gravitational field at some point in space, we can use it any time we want to calculate the resulting force on any mass we choose to place at that point. In fact, this is exactly what we do when we say the gravitational field of Earth (near Earth's surface) has a value of 9.81 m/s^2 , and then we calculate the resulting force (i.e., weight) on different masses. Also, the general expression for calculating \vec{g} at arbitrary distances from the center of Earth (i.e., not just near Earth's surface) is very similar to the expression for \vec{E} : $\vec{g} = G \frac{M}{r^2} \hat{r}$, where G is a proportionality constant, playing the same role for \vec{g} as $\frac{1}{4\pi\epsilon_0}$ does for \vec{E} . The value of \vec{g} is calculated once and is then used in an endless number of problems.

To push the analogy further, notice the units of the electric field: From $F = QE$, the units of E are newtons per coulomb, N/C, that is, the electric field applies a force on each unit charge. Now notice the units of g : From $w = mg$, the units of g are newtons per kilogram, N/kg, that is, the gravitational field applies a force on each unit mass. We could say that the gravitational field of Earth, near Earth's surface, has a value of 9.81 N/kg .

The Meaning of “Field”

Recall from your studies of gravity that the word “field” in this context has a precise meaning. A field, in physics, is a physical quantity whose value depends on (is a function of) position, relative to the source of the field. In the case of the electric field, [\[link\]](#) shows that the value of \vec{E} (both the magnitude and the direction) depends on where in space the point P is located, measured from the locations \vec{r}_i of the source charges q_i .

In addition, since the electric field is a vector quantity, the electric field is referred to as a *vector field*. (The gravitational field is also a vector field.) In contrast, a field that has only a magnitude at every point is a *scalar field*. The temperature in a room is an example of a scalar field. It is a field because the temperature, in general, is different at different locations in the room, and it is a scalar field because temperature is a scalar quantity.

Also, as you did with the gravitational field of an object with mass, you should picture the electric field of a charge-bearing object (the source charge) as a continuous, immaterial substance that surrounds the source charge, filling all of space—in principle, to $\pm\infty$ in all directions. The field exists at every physical point in space. To put it another way, the electric charge on an object alters the space around the charged object in such a way that all other electrically charged objects in space experience an electric force as a result of being in that field. The electric field, then, is the mechanism by which the electric properties of the source charge are transmitted to and through the rest of the universe. (Again, the range of the electric force is infinite.)

We will see in subsequent chapters that the speed at which electrical phenomena travel is the same as the speed of light. There is a deep connection between the electric field and light.

Superposition

Yet another experimental fact about the field is that it obeys the superposition principle. In this context, that means that we can (in principle) calculate the total electric field of many source charges by calculating the electric field of only q_1 at position P , then calculate the field of q_2 at P , while—and this is the crucial idea—ignoring the field of, and indeed even the existence of, q_1 . We can repeat this process, calculating the field of each individual source charge, independently of the existence of any of the other charges. The total electric field, then, is the vector sum of all these fields. That, in essence, is what [\[link\]](#) says.

In the next section, we describe how to determine the shape of an electric field of a source charge distribution and how to sketch it.

The Direction of the Field

[\[link\]](#) enables us to determine the magnitude of the electric field, but we need the direction also. We use the convention that the direction of any electric field vector is the same as the direction of the electric force vector that the field would apply to a positive test charge placed in that field. Such a charge would be repelled by positive source charges (the force on it would point

away from the positive source charge) but attracted to negative charges (the force points toward the negative source).

Note:

Direction of the Electric Field

By convention, all electric fields \vec{E} point away from positive source charges and point toward negative source charges.

Note:

Add charges to the [Electric Field of Dreams](#) and see how they react to the electric field. Turn on a background electric field and adjust the direction and magnitude.

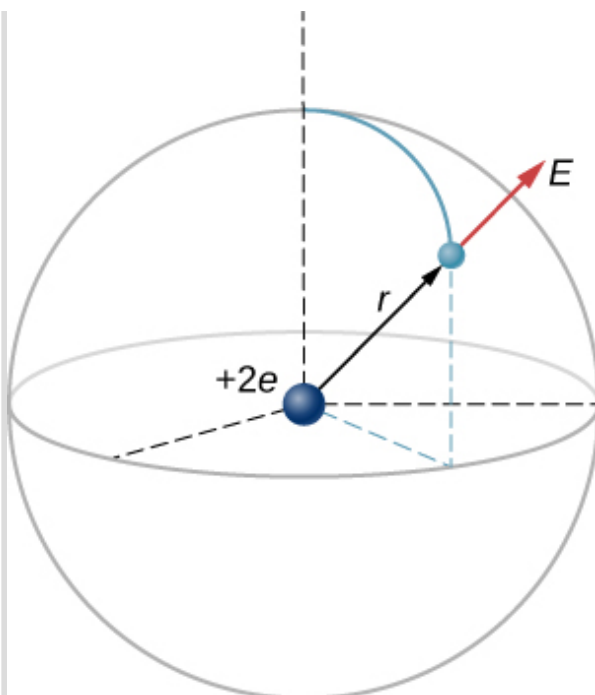
Example:

The E -field of an Atom

In an ionized helium atom, the most probable distance between the nucleus and the electron is $r = 26.5 \times 10^{-12}$ m. What is the electric field due to the nucleus at the location of the electron?

Strategy

Note that although the electron is mentioned, it is not used in any calculation. The problem asks for an electric field, not a force; hence, there is only one charge involved, and the problem specifically asks for the field due to the nucleus. Thus, the electron is a red herring; only its distance matters. Also, since the distance between the two protons in the nucleus is much, much smaller than the distance of the electron from the nucleus, we can treat the two protons as a single charge $+2e$ ([\[link\]](#)).



A schematic representation of a helium atom. Again, helium physically looks nothing like this, but this sort of diagram is helpful for calculating the electric field of the nucleus.

Solution

The electric field is calculated by

Equation:

$$\vec{\mathbf{E}} = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N \frac{q_i}{r_i^2} \hat{\mathbf{r}}_i.$$

Since there is only one source charge (the nucleus), this expression simplifies to

Equation:

$$\vec{\mathbf{E}} = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \hat{\mathbf{r}}.$$

Here $q = 2e = 2 (1.6 \times 10^{-19} \text{ C})$ (since there are two protons) and r is given; substituting gives

Equation:

$$\vec{\mathbf{E}} = \frac{1}{4\pi \left(8.85 \times 10^{-12} \frac{\text{C}^2}{\text{N}\cdot\text{m}^2}\right)} \frac{2 (1.6 \times 10^{-19} \text{ C})}{(26.5 \times 10^{-12} \text{ m})^2} \hat{\mathbf{r}} = 4.1 \times 10^{12} \frac{\text{N}}{\text{C}} \hat{\mathbf{r}}.$$

The direction of $\vec{\mathbf{E}}$ is radially away from the nucleus in all directions. Why? Because a positive test charge placed in this field would accelerate radially away from the nucleus (since it is also positively charged), and again, the convention is that the direction of the electric field vector is defined in terms of the direction of the force it would apply to positive test charges.

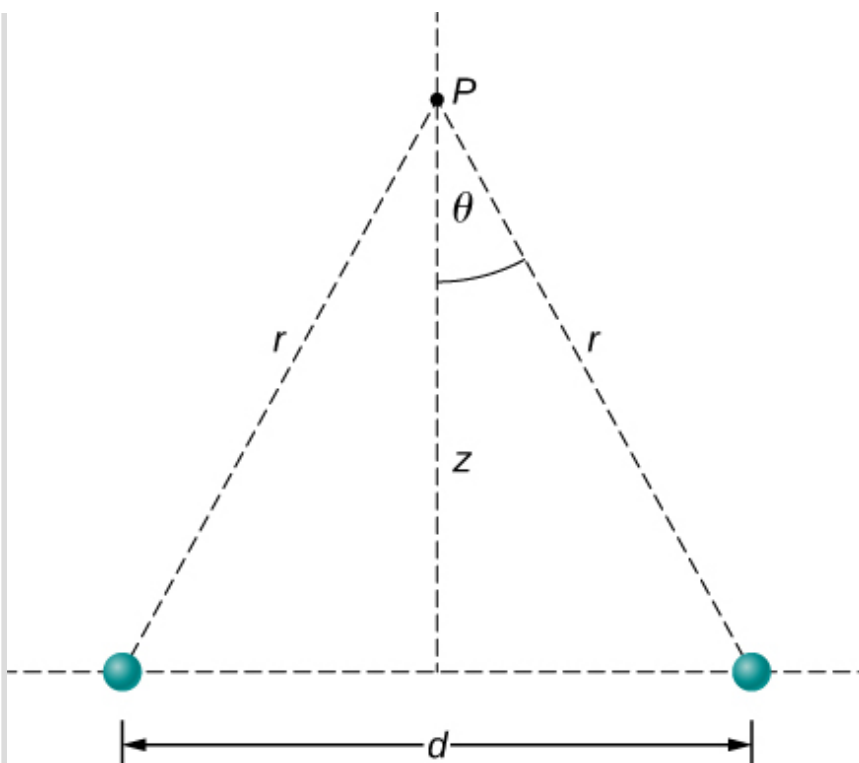
Example:

The E -Field above Two Equal Charges

(a) Find the electric field (magnitude and direction) a distance z above the midpoint between two equal charges $+q$ that are a distance d apart ([\[link\]](#)).

Check that your result is consistent with what you'd expect when $z \gg d$.

(b) The same as part (a), only this time make the right-hand charge $-q$ instead of $+q$.



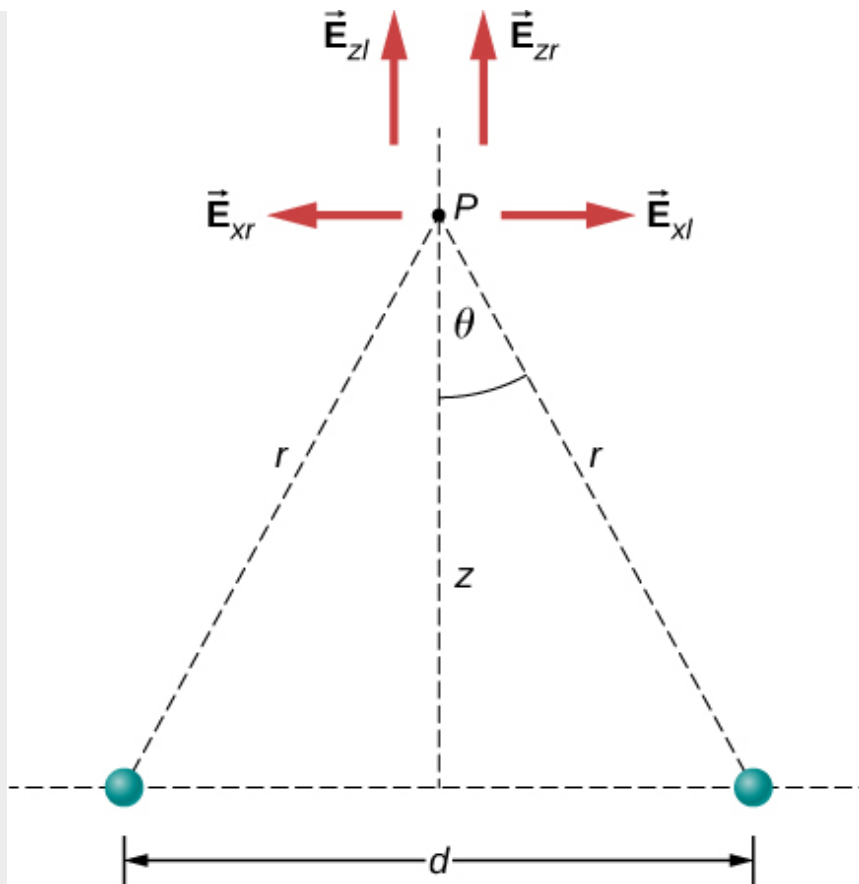
Finding the field of two identical source charges at the point P . Due to the symmetry, the net field at P is entirely vertical. (Notice that this is *not* true away from the midline between the charges.)

Strategy

We add the two fields as vectors, per [\[link\]](#). Notice that the system (and therefore the field) is symmetrical about the vertical axis; as a result, the horizontal components of the field vectors cancel. This simplifies the math. Also, we take care to express our final answer in terms of only quantities that are given in the original statement of the problem: q , z , d , and constants (π , ϵ_0).

Solution

- a. By symmetry, the horizontal (x)-components of \vec{E} cancel ([\[link\]](#));
- $$E_x = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \sin \theta - \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \sin \theta = 0.$$



Note that the horizontal components of the electric fields from the two charges cancel each other out, while the vertical components add together.

The vertical (z)-component is given by

Equation:

$$E_z = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \cos \theta + \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \cos \theta = \frac{1}{4\pi\epsilon_0} \frac{2q}{r^2} \cos \theta.$$

Since none of the other components survive, this is the entire electric field, and it points in the $\hat{\mathbf{k}}$ direction. Notice that this calculation uses the principle of **superposition**; we calculate the fields of the two charges independently and then add them together.

What we want to do now is replace the quantities in this expression that we don't know (such as r), or can't easily measure (such as $\cos \theta$) with quantities that we do know, or can measure. In this case, by geometry,

Equation:

$$r^2 = z^2 + \left(\frac{d}{2}\right)^2$$

and

Equation:

$$\cos \theta = \frac{z}{r} = \frac{z}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^{1/2}}.$$

Thus, substituting,

Equation:

$$\vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \frac{2q}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]} \frac{z}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^{1/2}} \hat{\mathbf{k}}.$$

Simplifying, the desired answer is

Equation:

$$\vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \frac{2qz}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^{3/2}} \hat{\mathbf{k}}.$$

- b. If the source charges are equal and opposite, the vertical components cancel because $E_z = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \cos \theta - \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \cos \theta = 0$

and we get, for the horizontal component of $\vec{\mathbf{E}}$,

Equation:

$$\begin{aligned}
\vec{\mathbf{E}}(z) &= \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \sin \theta \hat{\mathbf{i}} - \frac{1}{4\pi\epsilon_0} \frac{-q}{r^2} \sin \theta \hat{\mathbf{i}} \\
&= \frac{1}{4\pi\epsilon_0} \frac{2q}{r^2} \sin \theta \hat{\mathbf{i}} \\
&= \frac{1}{4\pi\epsilon_0} \frac{2q}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]} \frac{\left(\frac{d}{2}\right)}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^{1/2}} \hat{\mathbf{i}}.
\end{aligned}$$

This becomes

Equation:

$$\vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \frac{qd}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^{3/2}} \hat{\mathbf{i}}.$$

Significance

It is a very common and very useful technique in physics to check whether your answer is reasonable by evaluating it at extreme cases. In this example, we should evaluate the field expressions for the cases $d = 0$, $z \gg d$, and $z \rightarrow \infty$, and confirm that the resulting expressions match our physical expectations. Let's do so:

Let's start with [\[link\]](#), the field of two identical charges. From far away (i.e., $z \gg d$), the two source charges should “merge” and we should then “see” the field of just one charge, of size $2q$. So, let $z \gg d$; then we can neglect d^2 in [\[link\]](#) to obtain

Equation:

$$\begin{aligned}
\lim_{d \rightarrow 0} \vec{\mathbf{E}} &= \frac{1}{4\pi\epsilon_0} \frac{2qz}{[z^2]^{3/2}} \hat{\mathbf{k}} \\
&= \frac{1}{4\pi\epsilon_0} \frac{2qz}{z^3} \hat{\mathbf{k}} \\
&= \frac{1}{4\pi\epsilon_0} \frac{(2q)}{z^2} \hat{\mathbf{k}},
\end{aligned}$$

which is the correct expression for a field at a distance z away from a charge $2q$.

Next, we consider the field of equal and opposite charges, [\[link\]](#). It can be shown (via a Taylor expansion) that for $d \ll z \ll \infty$, this becomes

Equation:

$$\vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \frac{qd}{z^3} \hat{\mathbf{i}},$$

which is the field of a dipole, a system that we will study in more detail later. (Note that the units of $\vec{\mathbf{E}}$ are still correct in this expression, since the units of d in the numerator cancel the unit of the “extra” z in the denominator.) If z is very large ($z \rightarrow \infty$), then $E \rightarrow 0$, as it should; the two charges “merge” and so cancel out.

Note:

Exercise:

Problem:

Check Your Understanding What is the electric field due to a single point particle?

Solution:

$$\vec{\mathbf{E}} = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \hat{\mathbf{r}}$$

Note:

Try this [simulation of electric field hockey](#) to get the charge in the goal by placing other charges on the field.

Summary

- The electric field is an alteration of space caused by the presence of an electric charge. The electric field mediates the electric force between a source charge and a test charge.
- The electric field, like the electric force, obeys the superposition principle

- The field is a vector; by definition, it points away from positive charges and toward negative charges.

Conceptual Questions

Exercise:

Problem:

When measuring an electric field, could we use a negative rather than a positive test charge?

Solution:

Either sign of the test charge could be used, but the convention is to use a positive test charge.

Exercise:

Problem:

During fair weather, the electric field due to the net charge on Earth points downward. Is Earth charged positively or negatively?

Exercise:

Problem:

If the electric field at a point on the line between two charges is zero, what do you know about the charges?

Solution:

The charges are of the same sign.

Exercise:

Problem:

Two charges lie along the x -axis. Is it true that the net electric field always vanishes at some point (other than infinity) along the x -axis?

Problems

Exercise:

Problem:

A particle of charge $2.0 \times 10^{-8} \text{ C}$ experiences an upward force of magnitude $4.0 \times 10^{-6} \text{ N}$ when it is placed in a particular point in an electric field. (a) What is the electric field at that point? (b) If a charge $q = -1.0 \times 10^{-8} \text{ C}$ is placed there, what is the force on it?

Solution:

- a. $E = 2.0 \times 10^2 \frac{\text{N}}{\text{C}}$ up;
- b. $F = 2.0 \times 10^{-6} \text{ N}$ down

Exercise:

Problem:

On a typical clear day, the atmospheric electric field points downward and has a magnitude of approximately 100 N/C . Compare the gravitational and electric forces on a small dust particle of mass $2.0 \times 10^{-15} \text{ g}$ that carries a single electron charge. What is the acceleration (both magnitude and direction) of the dust particle?

Exercise:

Problem:

Consider an electron that is 10^{-10} m from an alpha particle ($q = 3.2 \times 10^{-19} \text{ C}$). (a) What is the electric field due to the alpha particle at the location of the electron? (b) What is the electric field due to the electron at the location of the alpha particle? (c) What is the electric force on the alpha particle? On the electron?

Solution:

- a. $E = 2.88 \times 10^{11} \text{ N/C}$;
- b. $E = 1.44 \times 10^{11} \text{ N/C}$;

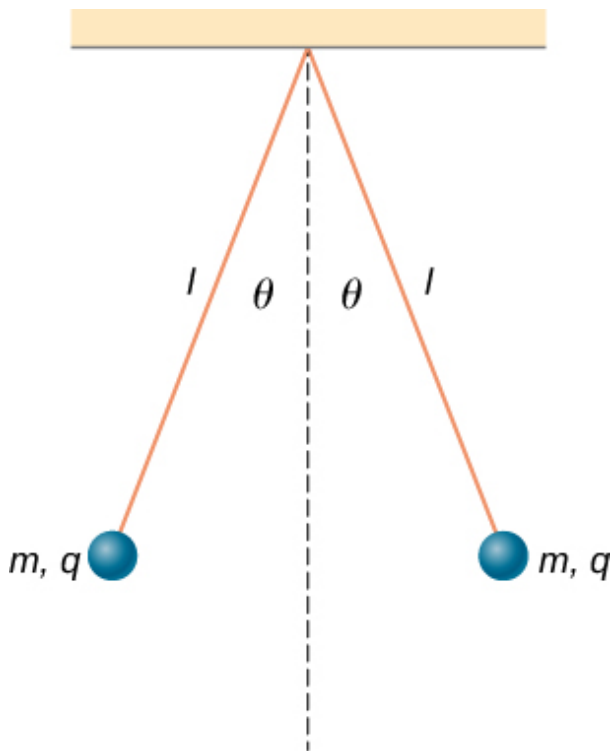
c. $F = 4.61 \times 10^{-8}$ N on alpha particle;
 $F = 4.61 \times 10^{-8}$ N on electron

Exercise:

Problem:

Each the balls shown below carries a charge q and has a mass m . The length of each thread is l , and at equilibrium, the balls are separated by an angle 2θ . How does θ vary with q and l ? Show that θ satisfies

$$\sin(\theta)^2 \tan(\theta) = \frac{q^2}{16\pi\epsilon_0 g l^2 m}.$$



Exercise:

Problem:

What is the electric field at a point where the force on a -2.0×10^{-6} C charge is $(4.0\hat{\mathbf{i}} - 6.0\hat{\mathbf{j}}) \times 10^{-6}$ N?

Solution:

$$\mathbf{E} = (-2.0\hat{\mathbf{i}} + 3.0\hat{\mathbf{j}}) \text{ N}$$

Exercise:

Problem:

A proton is suspended in the air by an electric field at the surface of Earth. What is the strength of this electric field?

Exercise:

Problem:

The electric field in a particular thundercloud is $2.0 \times 10^5 \text{ N/C}$. What is the acceleration of an electron in this field?

Solution:

$$F = 3.204 \times 10^{-14} \text{ N},$$
$$a = 3.517 \times 10^{16} \text{ m/s}^2$$

Exercise:

Problem:

A small piece of cork whose mass is 2.0 g is given a charge of $5.0 \times 10^{-7} \text{ C}$. What electric field is needed to place the cork in equilibrium under the combined electric and gravitational forces?

Exercise:

Problem:

If the electric field is 100 N/C at a distance of 50 cm from a point charge q , what is the value of q ?

Solution:

$$q = 2.78 \times 10^{-9} \text{ C}$$

Exercise:

Problem:

What is the electric field of a proton at the first Bohr orbit for hydrogen ($r = 5.29 \times 10^{-11} \text{ m}$)? What is the force on the electron in that orbit?

Exercise:**Problem:**

(a) What is the electric field of an oxygen nucleus at a point that is 10^{-10} m from the nucleus? (b) What is the force this electric field exerts on a second oxygen nucleus placed at that point?

Solution:

- a. $E = 1.15 \times 10^{12} \text{ N/C}$;
- b. $F = 1.47 \times 10^{-6} \text{ N}$

Exercise:**Problem:**

Two point charges, $q_1 = 2.0 \times 10^{-7} \text{ C}$ and $q_2 = -6.0 \times 10^{-8} \text{ C}$, are held 25.0 cm apart. (a) What is the electric field at a point 5.0 cm from the negative charge and along the line between the two charges? (b) What is the force on an electron placed at that point?

Exercise:**Problem:**

Point charges $q_1 = 50 \mu\text{C}$ and $q_2 = -25 \mu\text{C}$ are placed 1.0 m apart. (a) What is the electric field at a point midway between them? (b) What is the force on a charge $q_3 = 20 \mu\text{C}$ situated there?

Solution:

If the q_2 is to the right of q_1 , the electric field vector from both charges point to the right. a. $E = 2.70 \times 10^6 \text{ N/C}$;
b. $F = 54.0 \text{ N}$

Exercise:**Problem:**

Can you arrange the two point charges $q_1 = -2.0 \times 10^{-6} \text{ C}$ and $q_2 = 4.0 \times 10^{-6} \text{ C}$ along the x -axis so that $E = 0$ at the origin?

Exercise:**Problem:**

Point charges $q_1 = q_2 = 4.0 \times 10^{-6} \text{ C}$ are fixed on the x -axis at $x = -3.0 \text{ m}$ and $x = 3.0 \text{ m}$. What charge q must be placed at the origin so that the electric field vanishes at $x = 0, y = 3.0 \text{ m}$?

Solution:

There is 45° right triangle geometry. The x -components of the electric field at $y = 3 \text{ m}$ cancel. The y -components give

$$E(y = 3 \text{ m}) = 2.83 \times 10^3 \text{ N/C}.$$

At the origin we have a negative charge of magnitude $q = -2.83 \times 10^{-6} \text{ C}$.

Glossary**electric field**

physical phenomenon created by a charge; it “transmits” a force between two charges

superposition

concept that states that the net electric field of multiple source charges is the vector sum of the field of each source charge calculated individually

Calculating Electric Fields of Charge Distributions

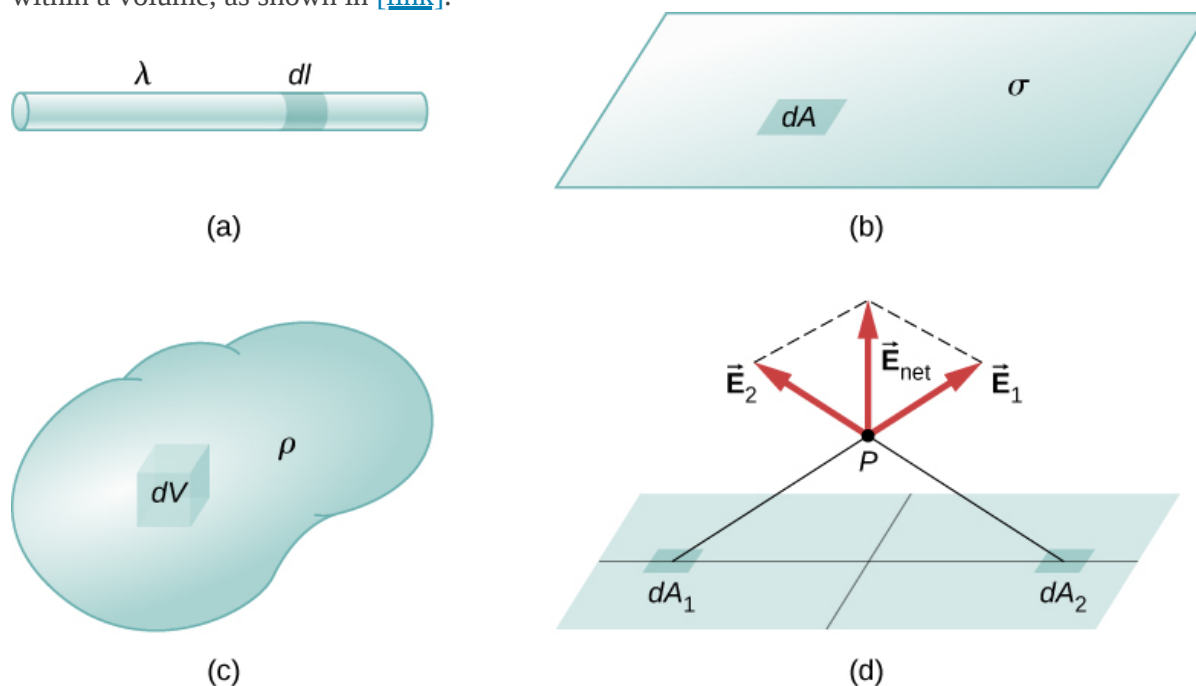
By the end of this section, you will be able to:

- Explain what a continuous source charge distribution is and how it is related to the concept of quantization of charge
- Describe line charges, surface charges, and volume charges
- Calculate the field of a continuous source charge distribution of either sign

The charge distributions we have seen so far have been discrete: made up of individual point particles. This is in contrast with a **continuous charge distribution**, which has at least one nonzero dimension. If a charge distribution is continuous rather than discrete, we can generalize the definition of the electric field. We simply divide the charge into infinitesimal pieces and treat each piece as a point charge.

Note that because charge is quantized, there is no such thing as a “truly” continuous charge distribution. However, in most practical cases, the total charge creating the field involves such a huge number of discrete charges that we can safely ignore the discrete nature of the charge and consider it to be continuous. This is exactly the kind of approximation we make when we deal with a bucket of water as a continuous fluid, rather than a collection of H_2O molecules.

Our first step is to define a charge density for a charge distribution along a line, across a surface, or within a volume, as shown in [\[link\]](#).



The configuration of charge differential elements for a (a) line charge, (b) sheet of charge, and (c) a volume of charge. Also note that (d) some of the components of the total electric field cancel out, with the remainder resulting in a net electric field.

Definitions of charge density:

- $\lambda \equiv$ charge per unit length (**linear charge density**); units are coulombs per meter (C/m)
- $\sigma \equiv$ charge per unit area (**surface charge density**); units are coulombs per square meter (C/m²)
- $\rho \equiv$ charge per unit volume (**volume charge density**); units are coulombs per cubic meter (C/m³)

Then, for a line charge, a surface charge, and a volume charge, the summation in [\[link\]](#) becomes an integral and q_i is replaced by $dq = \lambda dl$, σdA , or ρdV , respectively:

Equation:

$$\text{Point charges:} \quad \vec{\mathbf{E}}(P) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N \left(\frac{q_i}{r^2} \right) \hat{\mathbf{r}}$$

Equation:

$$\text{Line charge:} \quad \vec{\mathbf{E}}(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{line}} \left(\frac{\lambda dl}{r^2} \right) \hat{\mathbf{r}}$$

Equation:

$$\text{Surface charge:} \quad \vec{\mathbf{E}}(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{surface}} \left(\frac{\sigma dA}{r^2} \right) \hat{\mathbf{r}}$$

Equation:

$$\text{Volume charge:} \quad \vec{\mathbf{E}}(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{volume}} \left(\frac{\rho dV}{r^2} \right) \hat{\mathbf{r}}$$

The integrals are generalizations of the expression for the field of a point charge. They implicitly include and assume the principle of superposition. The “trick” to using them is almost always in coming up with correct expressions for dl , dA , or dV , as the case may be, expressed in terms of r , and also expressing the charge density function appropriately. It may be constant; it might be dependent on location.

Note carefully the meaning of r in these equations: It is the distance from the charge element (q_i , λdl , σdA , ρdV) to the location of interest, $P(x, y, z)$ (the point in space where you want to determine the field). However, don’t confuse this with the meaning of $\hat{\mathbf{r}}$; we are using it and the vector notation $\vec{\mathbf{E}}$ to write three integrals at once. That is, [\[link\]](#) is actually

Equation:

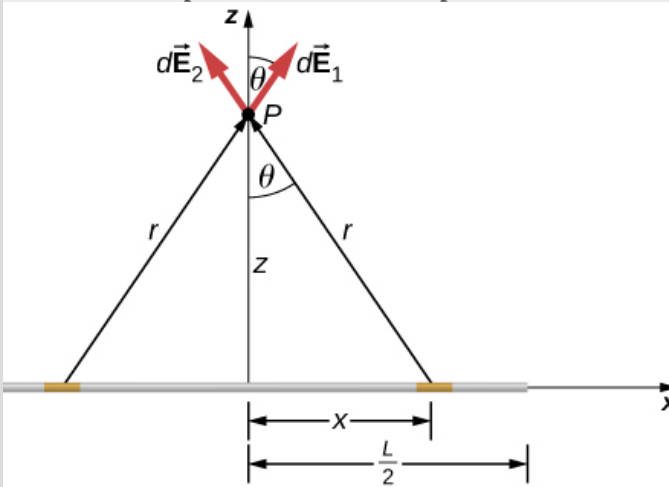
$$E_x(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{line}} \left(\frac{\lambda dl}{r^2} \right)_x, \quad E_y(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{line}} \left(\frac{\lambda dl}{r^2} \right)_y, \quad E_z(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{line}} \left(\frac{\lambda dl}{r^2} \right)_z.$$

Example:**Electric Field of a Line Segment**

Find the electric field a distance z above the midpoint of a straight line segment of length L that carries a uniform line charge density λ .

Strategy

Since this is a continuous charge distribution, we conceptually break the wire segment into differential pieces of length dl , each of which carries a differential amount of charge $dq = \lambda dl$. Then, we calculate the differential field created by two symmetrically placed pieces of the wire, using the symmetry of the setup to simplify the calculation ([link](#)). Finally, we integrate this differential field expression over the length of the wire (half of it, actually, as we explain below) to obtain the complete electric field expression.



A uniformly charged segment of wire. The electric field at point P can be found by applying the superposition principle to symmetrically placed charge elements and integrating.

Solution

Before we jump into it, what do we expect the field to “look like” from far away? Since it is a finite line segment, from far away, it should look like a point charge. We will check the expression we get to see if it meets this expectation.

The electric field for a line charge is given by the general expression

Equation:

$$\vec{\mathbf{E}}(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{line}} \frac{\lambda dl}{r^2} \hat{\mathbf{r}}.$$

The symmetry of the situation (our choice of the two identical differential pieces of charge) implies the horizontal (x)-components of the field cancel, so that the net field points in the z -direction. Let’s check this formally.

The total field $\vec{\mathbf{E}}(P)$ is the vector sum of the fields from each of the two charge elements (call them $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$, for now):

Equation:

$$\vec{\mathbf{E}}(P) = \vec{\mathbf{E}}_1 + \vec{\mathbf{E}}_2 = E_{1x}\hat{\mathbf{i}} + E_{1z}\hat{\mathbf{k}} + E_{2x}(-\hat{\mathbf{i}}) + E_{2z}\hat{\mathbf{k}}.$$

Because the two charge elements are identical and are the same distance away from the point P where we want to calculate the field, $E_{1x} = E_{2x}$, so those components cancel. This leaves

Equation:

$$\vec{\mathbf{E}}(P) = E_{1z}\hat{\mathbf{k}} + E_{2z}\hat{\mathbf{k}} = E_1 \cos \theta \hat{\mathbf{k}} + E_2 \cos \theta \hat{\mathbf{k}}.$$

These components are also equal, so we have

Equation:

$$\begin{aligned}\vec{\mathbf{E}}(P) &= \frac{1}{4\pi\epsilon_0} \int \frac{\lambda dl}{r^2} \cos \theta \hat{\mathbf{k}} + \frac{1}{4\pi\epsilon_0} \int \frac{\lambda dl}{r^2} \cos \theta \hat{\mathbf{k}} \\ &= \frac{1}{4\pi\epsilon_0} \int_0^{L/2} \frac{2\lambda dx}{r^2} \cos \theta \hat{\mathbf{k}}\end{aligned}$$

where our differential line element dl is dx , in this example, since we are integrating along a line of charge that lies on the x -axis. (The limits of integration are 0 to $\frac{L}{2}$, not $-\frac{L}{2}$ to $+\frac{L}{2}$, because we have constructed the net field from two differential pieces of charge dq . If we integrated along the entire length, we would pick up an erroneous factor of 2.)

In principle, this is complete. However, to actually calculate this integral, we need to eliminate all the variables that are not given. In this case, both r and θ change as we integrate outward to the end of the line charge, so those are the variables to get rid of. We can do that the same way we did for the two point charges: by noticing that

Equation:

$$r = (z^2 + x^2)^{1/2}$$

and

Equation:

$$\cos \theta = \frac{z}{r} = \frac{z}{(z^2 + x^2)^{1/2}}.$$

Substituting, we obtain

Equation:

$$\begin{aligned}\vec{\mathbf{E}}(P) &= \frac{1}{4\pi\epsilon_0} \int_0^{L/2} \frac{2\lambda dx}{(z^2 + x^2)} \frac{z}{(z^2 + x^2)^{1/2}} \hat{\mathbf{k}} \\ &= \frac{1}{4\pi\epsilon_0} \int_0^{L/2} \frac{2\lambda z}{(z^2 + x^2)^{3/2}} dx \hat{\mathbf{k}} \\ &= \frac{2\lambda z}{4\pi\epsilon_0} \left[\frac{x}{z^2 \sqrt{z^2 + x^2}} \right]_0^{L/2} \hat{\mathbf{k}}\end{aligned}$$

which simplifies to

Equation:

$$\vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \frac{\lambda L}{z\sqrt{z^2 + \frac{L^2}{4}}} \hat{\mathbf{k}}.$$

Significance

Notice, once again, the use of symmetry to simplify the problem. This is a very common strategy for calculating electric fields. The fields of nonsymmetrical charge distributions have to be handled with multiple integrals and may need to be calculated numerically by a computer.

Note:

Exercise:

Problem:

Check Your Understanding How would the strategy used above change to calculate the electric field at a point a distance z above one end of the finite line segment?

Solution:

We will no longer be able to take advantage of symmetry. Instead, we will need to calculate each of the two components of the electric field with their own integral.

Example:

Electric Field of an Infinite Line of Charge

Find the electric field a distance z above the midpoint of an infinite line of charge that carries a uniform line charge density λ .

Strategy

This is exactly like the preceding example, except the limits of integration will be $-\infty$ to $+\infty$.

Solution

Again, the horizontal components cancel out, so we wind up with

Equation:

$$\vec{\mathbf{E}}(P) = \frac{1}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{\lambda dx}{r^2} \cos \theta \hat{\mathbf{k}}$$

where our differential line element dl is dx , in this example, since we are integrating along a line of charge that lies on the x -axis. Again,

Equation:

$$\cos \theta = \frac{z}{r} = \frac{z}{(z^2 + x^2)^{1/2}}.$$

Substituting, we obtain

Equation:

$$\begin{aligned}
 \vec{\mathbf{E}}(P) &= \frac{1}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{\lambda dx}{(z^2 + x^2)} \frac{z}{(z^2 + x^2)^{1/2}} \hat{\mathbf{k}} \\
 &= \frac{1}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{\lambda z}{(z^2 + x^2)^{3/2}} dx \hat{\mathbf{k}} \\
 &= \frac{\lambda z}{4\pi\epsilon_0} \left[\frac{x}{z^2 \sqrt{z^2 + x^2}} \right] \Big|_{-\infty}^{\infty} \hat{\mathbf{k}},
 \end{aligned}$$

which simplifies to

Equation:

$$\vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \frac{2\lambda}{z} \hat{\mathbf{k}}.$$

Significance

Our strategy for working with continuous charge distributions also gives useful results for charges with infinite dimension.

In the case of a finite line of charge, note that for $z \gg L$, z^2 dominates the L in the denominator, so that [\[link\]](#) simplifies to

Equation:

$$\vec{\mathbf{E}} \approx \frac{1}{4\pi\epsilon_0} \frac{\lambda L}{z^2} \hat{\mathbf{k}}.$$

If you recall that $\lambda L = q$, the total charge on the wire, we have retrieved the expression for the field of a point charge, as expected.

In the limit $L \rightarrow \infty$, on the other hand, we get the field of an **infinite straight wire**, which is a straight wire whose length is much, much greater than either of its other dimensions, and also much, much greater than the distance at which the field is to be calculated:

Note:

Equation:

$$\vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \frac{2\lambda}{z} \hat{\mathbf{k}}.$$

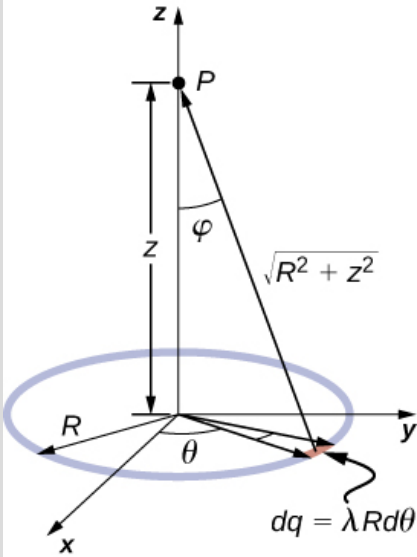
An interesting artifact of this infinite limit is that we have lost the usual $1/r^2$ dependence that we are used to. This will become even more intriguing in the case of an infinite plane.

Example:**Electric Field due to a Ring of Charge**

A ring has a uniform charge density λ , with units of coulomb per unit meter of arc. Find the electric field at a point on the axis passing through the center of the ring.

Strategy

We use the same procedure as for the charged wire. The difference here is that the charge is distributed on a circle. We divide the circle into infinitesimal elements shaped as arcs on the circle and use polar coordinates shown in [\[link\]](#).



The system and variable for calculating the electric field due to a ring of charge.

Solution

The electric field for a line charge is given by the general expression

Equation:

$$\vec{\mathbf{E}}(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{line}} \frac{\lambda dl}{r^2} \hat{\mathbf{r}}.$$

A general element of the arc between θ and $\theta + d\theta$ is of length $Rd\theta$ and therefore contains a charge equal to $\lambda Rd\theta$. The element is at a distance of $r = \sqrt{z^2 + R^2}$ from P , the angle is $\cos \phi = \frac{z}{\sqrt{z^2 + R^2}}$, and therefore the electric field is

Equation:

$$\begin{aligned}
 \vec{E}(P) &= \frac{1}{4\pi\epsilon_0} \int_{\text{line}} \frac{\lambda dl}{r^2} \hat{r} = \frac{1}{4\pi\epsilon_0} \int_0^{2\pi} \frac{\lambda R d\theta}{z^2 + R^2} \frac{z}{\sqrt{z^2 + R^2}} \hat{z} \\
 &= \frac{1}{4\pi\epsilon_0} \frac{\lambda R z}{(z^2 + R^2)^{3/2}} \hat{z} \int_0^{2\pi} d\theta = \frac{1}{4\pi\epsilon_0} \frac{2\pi \lambda R z}{(z^2 + R^2)^{3/2}} \hat{z} \\
 &= \frac{1}{4\pi\epsilon_0} \frac{q_{\text{tot}} z}{(z^2 + R^2)^{3/2}} \hat{z}.
 \end{aligned}$$

Significance

As usual, symmetry simplified this problem, in this particular case resulting in a trivial integral. Also, when we take the limit of $z \gg R$, we find that

Equation:

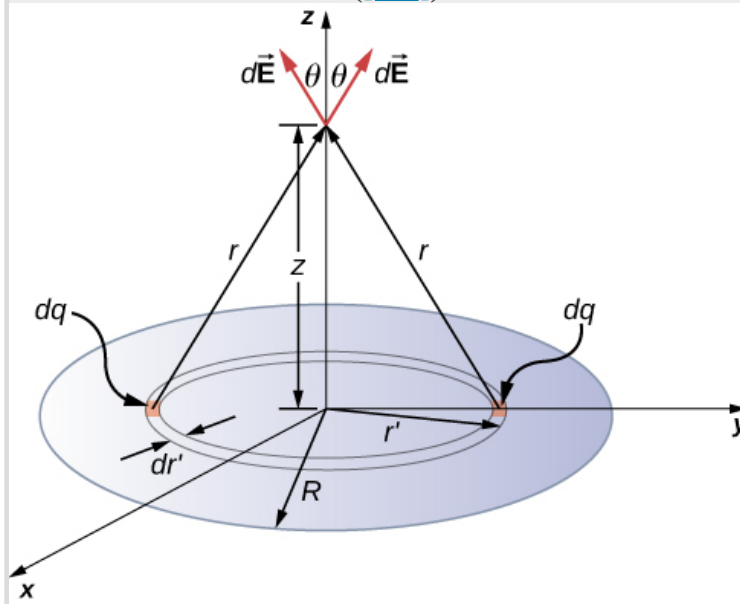
$$\vec{E} \approx \frac{1}{4\pi\epsilon_0} \frac{q_{\text{tot}}}{z^2} \hat{z},$$

as we expect.

Example:

The Field of a Disk

Find the electric field of a circular thin disk of radius R and uniform charge density at a distance z above the center of the disk ([link](#))



A uniformly charged disk. As in the line charge example, the field above the center of this disk can be calculated by taking advantage of the symmetry of the charge distribution.

Strategy

The electric field for a surface charge is given by

Equation:

$$\vec{\mathbf{E}}(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{surface}} \frac{\sigma dA}{r^2} \hat{\mathbf{r}}.$$

To solve surface charge problems, we break the surface into symmetrical differential “stripes” that match the shape of the surface; here, we’ll use rings, as shown in the figure. Again, by symmetry, the horizontal components cancel and the field is entirely in the vertical ($\hat{\mathbf{k}}$) direction. The vertical component of the electric field is extracted by multiplying by $\cos \theta$, so

Equation:

$$\vec{\mathbf{E}}(P) = \frac{1}{4\pi\epsilon_0} \int_{\text{surface}} \frac{\sigma dA}{r^2} \cos \theta \hat{\mathbf{k}}.$$

As before, we need to rewrite the unknown factors in the integrand in terms of the given quantities. In this case,

Equation:

$$\begin{aligned} dA &= 2\pi r' dr' \\ r^2 &= r'^2 + z^2 \\ \cos \theta &= \frac{z}{(r'^2 + z^2)^{1/2}}. \end{aligned}$$

(Please take note of the two different “ r ’s” here; r is the distance from the differential ring of charge to the point P where we wish to determine the field, whereas r' is the distance from the center of the disk to the differential ring of charge.) Also, we already performed the polar angle integral in writing down dA .

Solution

Substituting all this in, we get

Equation:

$$\begin{aligned} \vec{\mathbf{E}}(P) &= \vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \int_0^R \frac{\sigma (2\pi r' dr') z}{(r'^2 + z^2)^{3/2}} \hat{\mathbf{k}} \\ &= \frac{1}{4\pi\epsilon_0} (2\pi\sigma z) \left(\frac{1}{z} - \frac{1}{\sqrt{R^2 + z^2}} \right) \hat{\mathbf{k}} \end{aligned}$$

or, more simply,

Equation:

$$\vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \left(2\pi\sigma - \frac{2\pi\sigma z}{\sqrt{R^2 + z^2}} \right) \hat{\mathbf{k}}.$$

Significance

Again, it can be shown (via a Taylor expansion) that when $z \gg R$, this reduces to

Equation:

$$\vec{\mathbf{E}}(z) \approx \frac{1}{4\pi\epsilon_0} \frac{\sigma\pi R^2}{z^2} \hat{\mathbf{k}},$$

which is the expression for a point charge $Q = \sigma\pi R^2$.

Note:

Exercise:

Problem:

Check Your Understanding How would the above limit change with a uniformly charged rectangle instead of a disk?

Solution:

The point charge would be $Q = \sigma ab$ where a and b are the sides of the rectangle but otherwise identical.

As $R \rightarrow \infty$, [\[link\]](#) reduces to the field of an **infinite plane**, which is a flat sheet whose area is much, much greater than its thickness, and also much, much greater than the distance at which the field is to be calculated:

Note:

Equation:

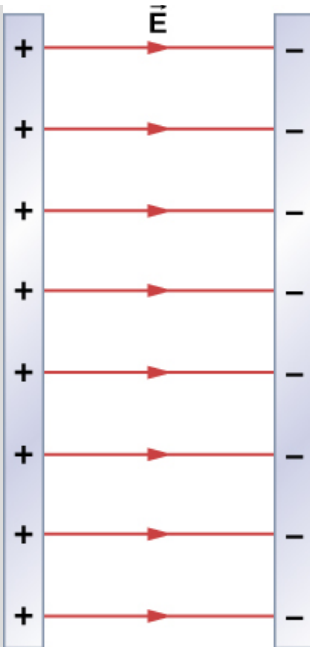
$$\vec{\mathbf{E}} = \frac{\sigma}{2\epsilon_0} \hat{\mathbf{k}}.$$

Note that this field is constant. This surprising result is, again, an artifact of our limit, although one that we will make use of repeatedly in the future. To understand why this happens, imagine being placed above an infinite plane of constant charge. Does the plane look any different if you vary your altitude? No—you still see the plane going off to infinity, no matter how far you are from it. It is important to note that [\[link\]](#) is because we are above the plane. If we were below, the field would point in the $-\hat{\mathbf{k}}$ direction.

Example:

The Field of Two Infinite Planes

Find the electric field everywhere resulting from two infinite planes with equal but opposite charge densities ([\[link\]](#)).



Two charged infinite planes. Note the direction of the electric field.

Strategy

We already know the electric field resulting from a single infinite plane, so we may use the principle of superposition to find the field from two.

Solution

The electric field points away from the positively charged plane and toward the negatively charged plane. Since the σ are equal and opposite, this means that in the region outside of the two planes, the electric fields cancel each other out to zero.

However, in the region between the planes, the electric fields add, and we get

Equation:

$$\vec{\mathbf{E}} = \frac{\sigma}{\epsilon_0} \hat{\mathbf{i}}$$

for the electric field. The $\hat{\mathbf{i}}$ is because in the figure, the field is pointing in the $+x$ -direction.

Significance

Systems that may be approximated as two infinite planes of this sort provide a useful means of creating uniform electric fields.

Note:

Exercise:

Problem:

Check Your Understanding What would the electric field look like in a system with two parallel positively charged planes with equal charge densities?

Solution:

The electric field would be zero in between, and have magnitude $\frac{\sigma}{\epsilon_0}$ everywhere else.

Summary

- A very large number of charges can be treated as a continuous charge distribution, where the calculation of the field requires integration. Common cases are:
 - one-dimensional (like a wire); uses a line charge density λ
 - two-dimensional (metal plate); uses surface charge density σ
 - three-dimensional (metal sphere); uses volume charge density ρ
- The “source charge” is a differential amount of charge dq . Calculating dq depends on the type of source charge distribution:

Equation:

$$dq = \lambda dl; \quad dq = \sigma dA; \quad dq = \rho dV.$$

- Symmetry of the charge distribution is usually key.
- Important special cases are the field of an “infinite” wire and the field of an “infinite” plane.

Conceptual Questions

Exercise:**Problem:**

Give a plausible argument as to why the electric field outside an infinite charged sheet is constant.

Solution:

At infinity, we would expect the field to go to zero, but because the sheet is infinite in extent, this is not the case. Everywhere you are, you see an infinite plane in all directions.

Exercise:**Problem:**

Compare the electric fields of an infinite sheet of charge, an infinite, charged conducting plate, and infinite, oppositely charged parallel plates.

Exercise:

Problem:

Describe the electric fields of an infinite charged plate and of two infinite, charged parallel plates in terms of the electric field of an infinite sheet of charge.

Solution:

The infinite charged plate would have $E = \frac{\sigma}{2\epsilon_0}$ everywhere. The field would point toward the plate if it were negatively charged and point away from the plate if it were positively charged. The electric field of the parallel plates would be zero between them if they had the same charge, and E would be $E = \frac{\sigma}{\epsilon_0}$ everywhere else. If the charges were opposite, the situation is reversed, zero outside the plates and $E = \frac{\sigma}{\epsilon_0}$ between them.

Exercise:**Problem:**

A negative charge is placed at the center of a ring of uniform positive charge. What is the motion (if any) of the charge? What if the charge were placed at a point on the axis of the ring other than the center?

Problems**Exercise:****Problem:**

A thin conducting plate 1.0 m on the side is given a charge of -2.0×10^{-6} C. An electron is placed 1.0 cm above the center of the plate. What is the acceleration of the electron?

Exercise:**Problem:**

Calculate the magnitude and direction of the electric field 2.0 m from a long wire that is charged uniformly at $\lambda = 4.0 \times 10^{-6}$ C/m.

Solution:

$$\vec{E}(z) = 3.6 \times 10^4 \text{ N/C} \hat{k}$$

Exercise:**Problem:**

Two thin conducting plates, each 25.0 cm on a side, are situated parallel to one another and 5.0 mm apart. If 10^{11} electrons are moved from one plate to the other, what is the electric field between the plates?

Exercise:

Problem:

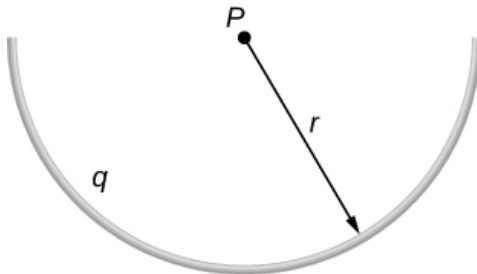
The charge per unit length on the thin rod shown below is λ . What is the electric field at the point P ? (*Hint: Solve this problem by first considering the electric field $d\vec{E}$ at P due to a small segment dx of the rod, which contains charge $dq = \lambda dx$. Then find the net field by integrating $d\vec{E}$ over the length of the rod.*)

**Solution:**

$$dE = \frac{1}{4\pi\epsilon_0} \frac{\lambda dx}{(x+a)^2}, \quad E = \frac{\lambda}{4\pi\epsilon_0} \left[\frac{1}{l+a} - \frac{1}{a} \right]$$

Exercise:**Problem:**

The charge per unit length on the thin semicircular wire shown below is λ . What is the electric field at the point P ?

**Exercise:****Problem:**

Two thin parallel conducting plates are placed 2.0 cm apart. Each plate is 2.0 cm on a side; one plate carries a net charge of $8.0 \mu\text{C}$, and the other plate carries a net charge of $-8.0 \mu\text{C}$. What is the charge density on the inside surface of each plate? What is the electric field between the plates?

Solution:

$$\sigma = 0.02 \text{ C/m}^2 \quad E = 2.26 \times 10^9 \text{ N/C}$$

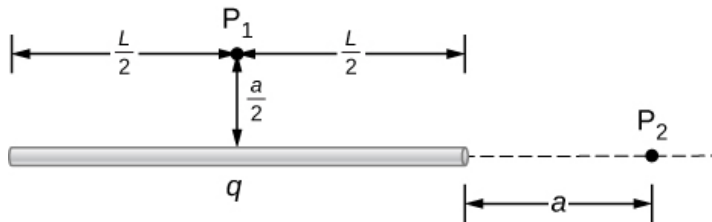
Exercise:

Problem:

A thin conducting plate 2.0 m on a side is given a total charge of $-10.0 \mu\text{C}$. (a) What is the electric field 1.0 cm above the plate? (b) What is the force on an electron at this point? (c) Repeat these calculations for a point 2.0 cm above the plate. (d) When the electron moves from 1.0 to 2.0 cm above the plate, how much work is done on it by the electric field?

Exercise:**Problem:**

A total charge q is distributed uniformly along a thin, straight rod of length L (see below). What is the electric field at P_1 ? At P_2 ?

**Solution:**

$$\text{At } P_1: \vec{E}(y) = \frac{1}{4\pi\epsilon_0} \frac{\lambda L}{y\sqrt{y^2 + \frac{L^2}{4}}} \hat{\mathbf{j}} \Rightarrow \frac{1}{4\pi\epsilon_0} \frac{q}{\frac{a}{2}\sqrt{(\frac{a}{2})^2 + \frac{L^2}{4}}} \hat{\mathbf{j}} = \frac{1}{\pi\epsilon_0} \frac{q}{a\sqrt{a^2 + L^2}} \hat{\mathbf{j}}$$

At P_2 : Put the origin at the end of L .

$$dE = \frac{1}{4\pi\epsilon_0} \frac{\lambda dx}{(x+a)^2}, \quad \vec{E} = -\frac{q}{4\pi\epsilon_0 l} \left[\frac{1}{l+a} - \frac{1}{a} \right] \hat{\mathbf{i}}$$

Exercise:**Problem:**

Charge is distributed along the entire x -axis with uniform density λ . How much work does the electric field of this charge distribution do on an electron that moves along the y -axis from $y = a$ to $y = b$?

Exercise:**Problem:**

Charge is distributed along the entire x -axis with uniform density λ_x and along the entire y -axis with uniform density λ_y . Calculate the resulting electric field at (a) $\vec{r} = a\hat{\mathbf{i}} + b\hat{\mathbf{j}}$ and (b) $\vec{r} = c\hat{\mathbf{k}}$.

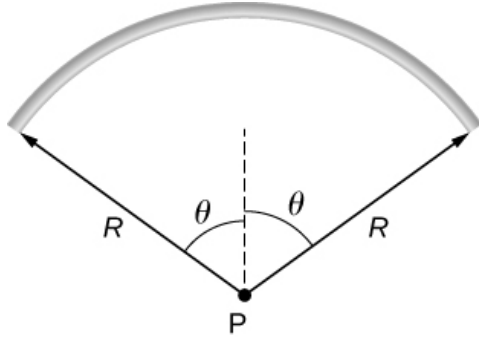
Solution:

$$\text{a. } \vec{E}(\vec{r}) = \frac{1}{4\pi\epsilon_0} \frac{2\lambda_x}{b} \hat{\mathbf{i}} + \frac{1}{4\pi\epsilon_0} \frac{2\lambda_y}{a} \hat{\mathbf{j}}; \text{ b. } \frac{1}{4\pi\epsilon_0} \frac{2(\lambda_x + \lambda_y)}{c} \hat{\mathbf{k}}$$

Exercise:

Problem:

A rod bent into the arc of a circle subtends an angle 2θ at the center P of the circle (see below). If the rod is charged uniformly with a total charge Q , what is the electric field at P ?

**Exercise:****Problem:**

A proton moves in the electric field $\vec{E} = 200\hat{i}$ N/C. (a) What are the force on and the acceleration of the proton? (b) Do the same calculation for an electron moving in this field.

Solution:

- a. $\vec{F} = 3.2 \times 10^{-17} \text{ N}\hat{i}$,
 $\vec{a} = 1.92 \times 10^{10} \text{ m/s}^2\hat{i}$;
 b. $\vec{F} = -3.2 \times 10^{-17} \text{ N}\hat{i}$,
 $\vec{a} = -3.51 \times 10^{13} \text{ m/s}^2\hat{i}$

Exercise:**Problem:**

An electron and a proton, each starting from rest, are accelerated by the same uniform electric field of 200 N/C. Determine the distance and time for each particle to acquire a kinetic energy of 3.2×10^{-16} J.

Exercise:**Problem:**

A spherical water droplet of radius $25 \mu\text{m}$ carries an excess 250 electrons. What vertical electric field is needed to balance the gravitational force on the droplet at the surface of the earth?

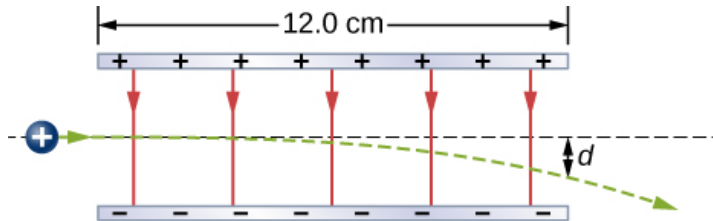
Solution:

$$m = 6.5 \times 10^{-11} \text{ kg},$$

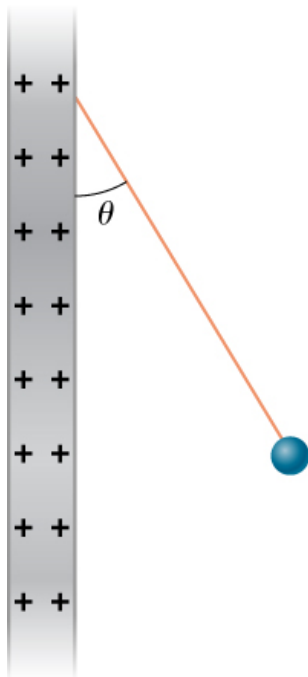
$$E = 1.6 \times 10^7 \text{ N/C}$$

Exercise:**Problem:**

A proton enters the uniform electric field produced by the two charged plates shown below. The magnitude of the electric field is $4.0 \times 10^5 \text{ N/C}$, and the speed of the proton when it enters is $1.5 \times 10^7 \text{ m/s}$. What distance d has the proton been deflected downward when it leaves the plates?

**Exercise:****Problem:**

Shown below is a small sphere of mass 0.25 g that carries a charge of $9.0 \times 10^{-10} \text{ C}$. The sphere is attached to one end of a very thin silk string 5.0 cm long. The other end of the string is attached to a large vertical conducting plate that has a charge density of $30 \times 10^{-6} \text{ C/m}^2$. What is the angle that the string makes with the vertical?



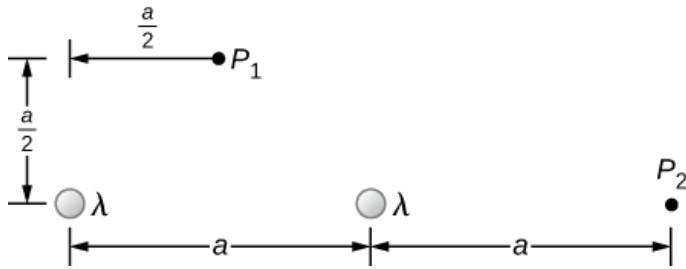
Solution:

$E = 1.70 \times 10^6 \text{ N/C}$,
 $F = 1.53 \times 10^{-3} \text{ N}$ $T \cos \theta = mg$ $T \sin \theta = qE$,
 $\tan \theta = 0.62 \Rightarrow \theta = 32.0^\circ$,
 This is independent of the length of the string.

Exercise:

Problem:

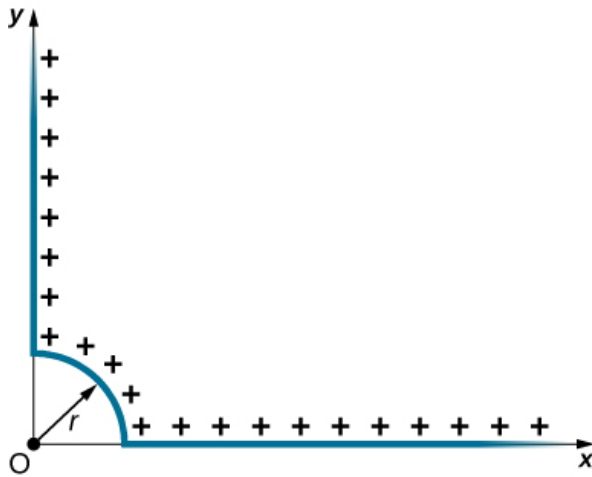
Two infinite rods, each carrying a uniform charge density λ , are parallel to one another and perpendicular to the plane of the page. (See below.) What is the electrical field at P_1 ? At P_2 ?



Exercise:

Problem:

Positive charge is distributed with a uniform density λ along the positive x -axis from r to ∞ , along the positive y -axis from r to ∞ , and along a 90° arc of a circle of radius r , as shown below. What is the electric field at O ?



Solution:

$$\text{circular arc } dE_x(-\hat{\mathbf{i}}) = \frac{1}{4\pi\epsilon_0} \frac{\lambda ds}{r^2} \cos \theta(-\hat{\mathbf{i}}),$$

$$\vec{E}_x = \frac{\lambda}{4\pi\epsilon_0 r} (-\hat{\mathbf{i}}),$$

$$dE_y(-\hat{\mathbf{j}}) = \frac{1}{4\pi\epsilon_0} \frac{\lambda ds}{r^2} \sin \theta(-\hat{\mathbf{j}}),$$

$$\vec{E}_y = \frac{\lambda}{4\pi\epsilon_0 r}(-\hat{j});$$

$$y\text{-axis: } \vec{E}_x = \frac{\lambda}{4\pi\epsilon_0 r}(-\hat{i});$$

$$x\text{-axis: } \vec{E}_y = \frac{\lambda}{4\pi\epsilon_0 r}(-\hat{j}),$$

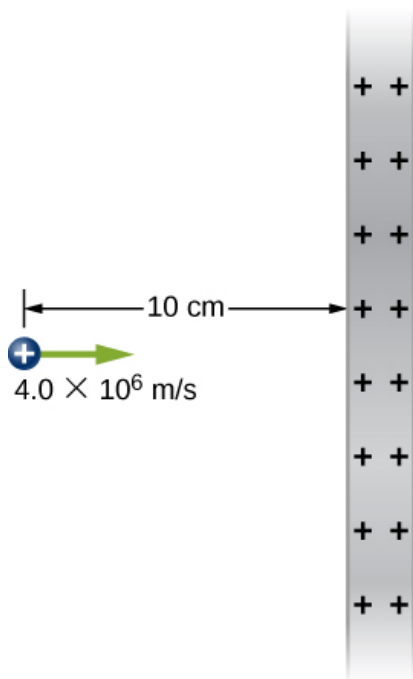
$$\vec{E} = \frac{\lambda}{2\pi\epsilon_0 r}(-\hat{i}) + \frac{\lambda}{2\pi\epsilon_0 r}(-\hat{j})$$

Exercise:

Problem:

From a distance of 10 cm, a proton is projected with a speed of $v = 4.0 \times 10^6$ m/s directly at a large, positively charged plate whose charge density is $\sigma = 2.0 \times 10^{-5}$ C/m². (See below.)

(a) Does the proton reach the plate? (b) If not, how far from the plate does it turn around?



Exercise:

Problem:

A particle of mass m and charge $-q$ moves along a straight line away from a fixed particle of charge Q . When the distance between the two particles is r_0 , $-q$ is moving with a speed v_0 . (a) Use the work-energy theorem to calculate the maximum separation of the charges. (b) What do you have to assume about v_0 to make this calculation? (c) What is the minimum value of v_0 such that $-q$ escapes from Q ?

Solution:

$$a. W = \frac{1}{2}m(v^2 - v_0^2), \frac{Qq}{4\pi\epsilon_0} \left(\frac{1}{r} - \frac{1}{r_0} \right) = \frac{1}{2}m(v^2 - v_0^2) \Rightarrow r_0 - r = \frac{4\pi\epsilon_0}{Qq} \frac{1}{2}mr_0m(v^2 - v_0^2)$$

; b. $r_0 - r$ is negative; therefore, $v_0 > v$,

$$r \rightarrow \infty, \text{ and } v \rightarrow 0: \frac{Qq}{4\pi\epsilon_0} \left(-\frac{1}{r_0} \right) = -\frac{1}{2}mv_0^2 \Rightarrow v_0 = \sqrt{\frac{Qq}{2\pi\epsilon_0 mr_0}}$$

Glossary

continuous charge distribution

total source charge composed of so large a number of elementary charges that it must be treated as continuous, rather than discrete

infinite plane

flat sheet in which the dimensions making up the area are much, much greater than its thickness, and also much, much greater than the distance at which the field is to be calculated; its field is constant

infinite straight wire

straight wire whose length is much, much greater than either of its other dimensions, and also much, much greater than the distance at which the field is to be calculated

linear charge density

amount of charge in an element of a charge distribution that is essentially one-dimensional (the width and height are much, much smaller than its length); its units are C/m

surface charge density

amount of charge in an element of a two-dimensional charge distribution (the thickness is small); its units are C/m²

volume charge density

amount of charge in an element of a three-dimensional charge distribution; its units are C/m³

Electric Field Lines

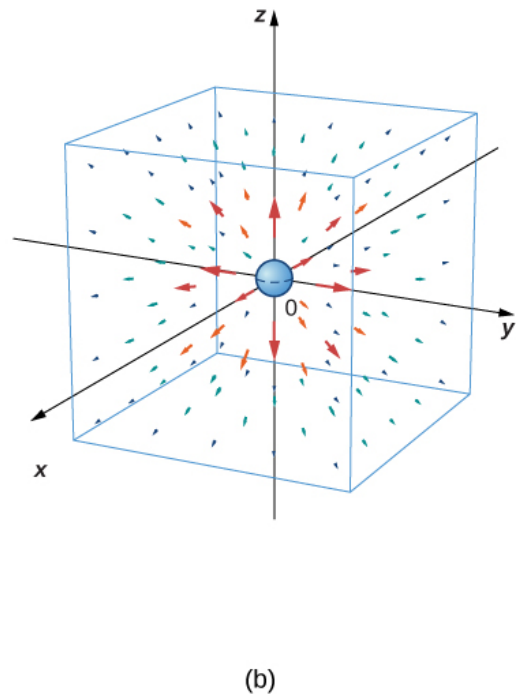
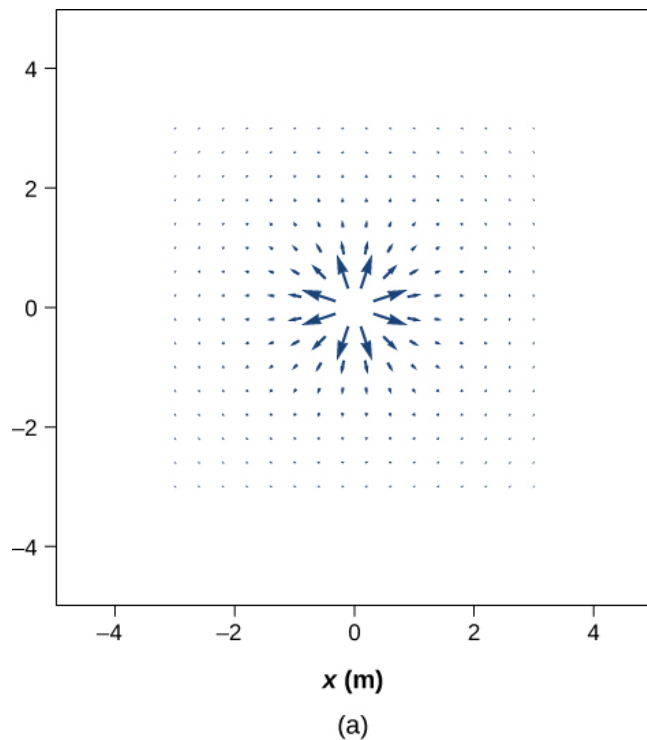
By the end of this section, you will be able to:

- Explain the purpose of an electric field diagram
- Describe the relationship between a vector diagram and a field line diagram
- Explain the rules for creating a field diagram and why these rules make physical sense
- Sketch the field of an arbitrary source charge

Now that we have some experience calculating electric fields, let's try to gain some insight into the geometry of electric fields. As mentioned earlier, our model is that the charge on an object (the source charge) alters space in the region around it in such a way that when another charged object (the test charge) is placed in that region of space, that test charge experiences an electric force. The concept of electric **field lines**, and of electric field line diagrams, enables us to visualize the way in which the space is altered, allowing us to visualize the field. The purpose of this section is to enable you to create sketches of this geometry, so we will list the specific steps and rules involved in creating an accurate and useful sketch of an electric field.

It is important to remember that electric fields are three-dimensional. Although in this book we include some pseudo-three-dimensional images, several of the diagrams that you'll see (both here, and in subsequent chapters) will be two-dimensional projections, or cross-sections. Always keep in mind that in fact, you're looking at a three-dimensional phenomenon.

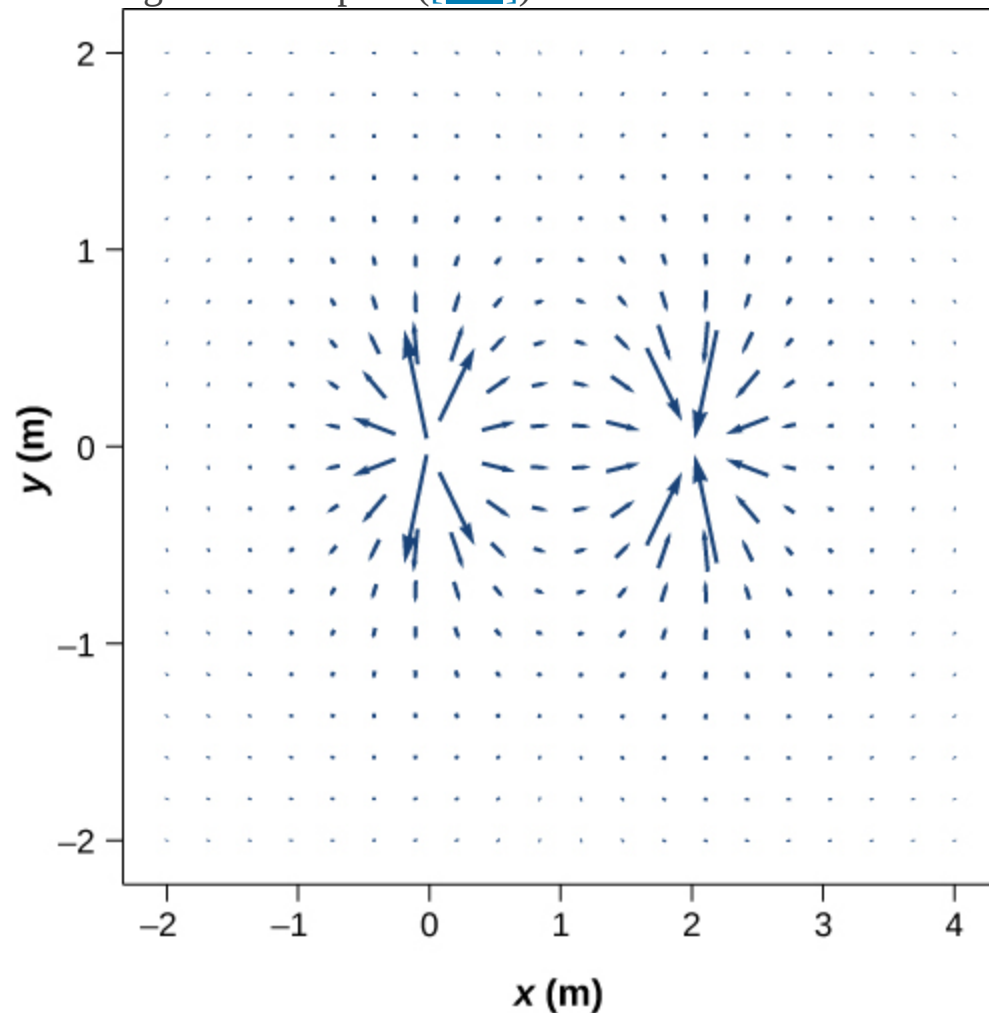
Our starting point is the physical fact that the electric field of the source charge causes a test charge in that field to experience a force. By definition, electric field vectors point in the same direction as the electric force that a (hypothetical) positive test charge would experience, if placed in the field ([link](#))



The electric field of a positive point charge. A large number of field vectors are shown. Like all vector arrows, the length of each vector is proportional to the magnitude of the field at each point. (a) Field in two dimensions; (b) field in three dimensions.

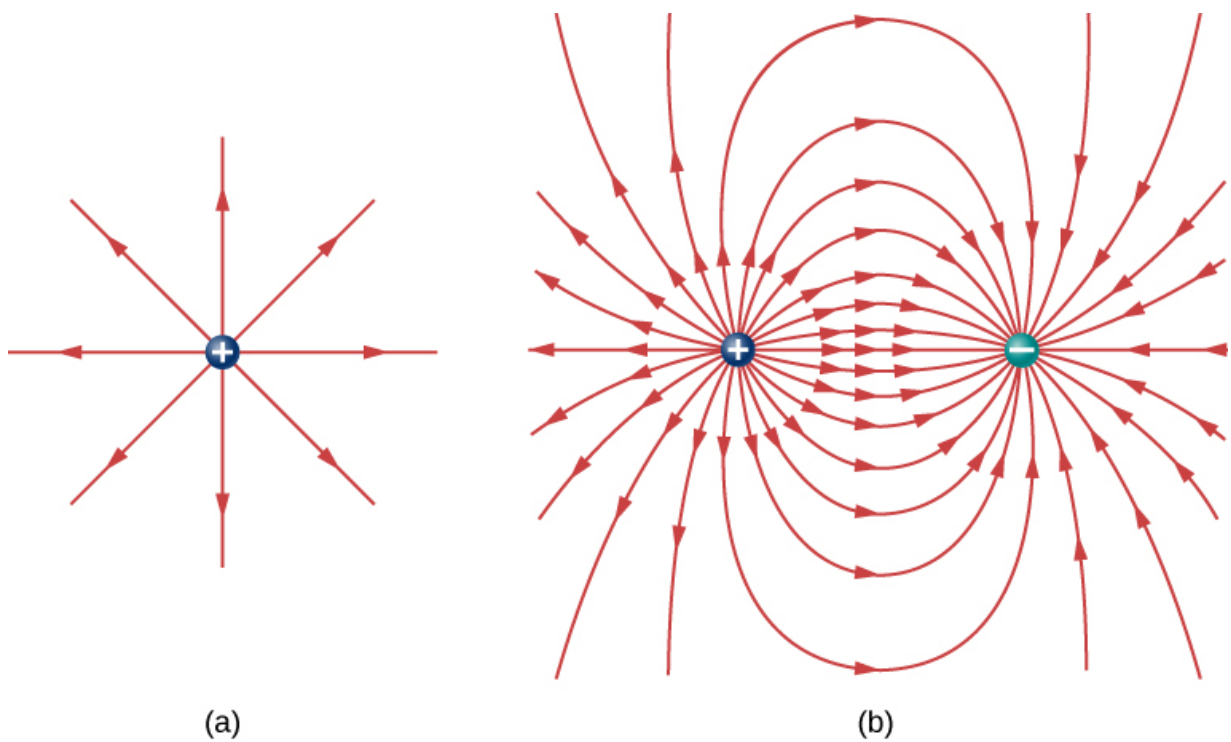
We've plotted many field vectors in the figure, which are distributed uniformly around the source charge. Since the electric field is a vector, the arrows that we draw correspond at every point in space to both the magnitude and the direction of the field at that point. As always, the length of the arrow that we draw corresponds to the magnitude of the field vector at that point. For a point source charge, the length decreases by the square of the distance from the source charge. In addition, the direction of the field vector is radially away from the source charge, because the direction of the electric field is defined by the direction of the force that a positive test charge would experience in that field. (Again, keep in mind that the actual field is three-dimensional; there are also field lines pointing out of and into the page.)

This diagram is correct, but it becomes less useful as the source charge distribution becomes more complicated. For example, consider the vector field diagram of a dipole ([\[link\]](#)).



The vector field of a dipole. Even with just two identical charges, the vector field diagram becomes difficult to understand.

There is a more useful way to present the same information. Rather than drawing a large number of increasingly smaller vector arrows, we instead connect all of them together, forming continuous lines and curves, as shown in [\[link\]](#).

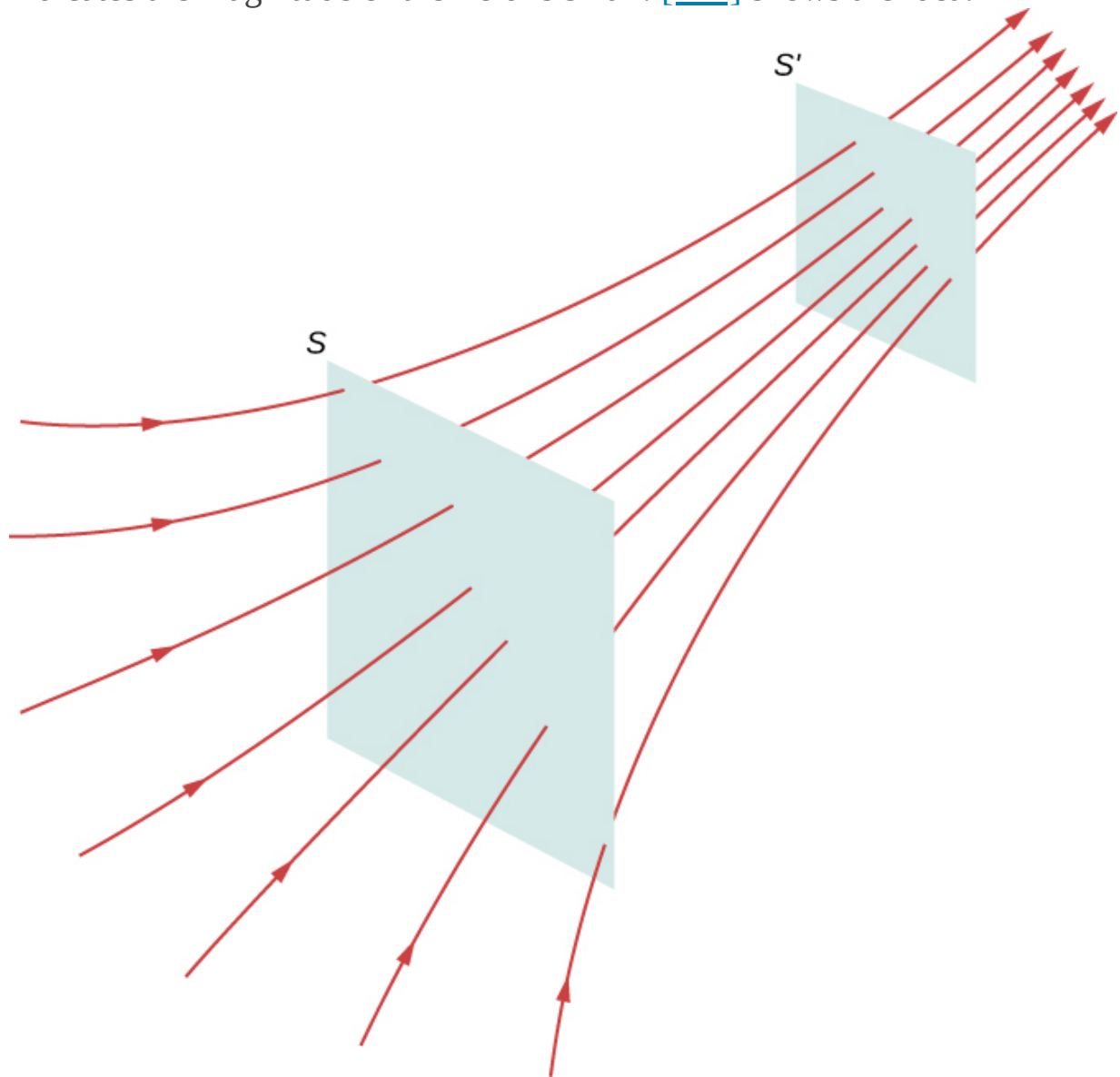


(a) The electric field line diagram of a positive point charge. (b) The field line diagram of a dipole. In both diagrams, the magnitude of the field is indicated by the field line density. The field *vectors* (not shown here) are everywhere tangent to the field lines.

Although it may not be obvious at first glance, these field diagrams convey the same information about the electric field as do the vector diagrams. First, the direction of the field at every point is simply the direction of the field vector at that same point. In other words, at any point in space, the field vector at each point is tangent to the field line at that same point. The arrowhead placed on a field line indicates its direction.

As for the magnitude of the field, that is indicated by the **field line density**—that is, the number of field lines per unit area passing through a small cross-sectional area perpendicular to the electric field. This field line density is drawn to be proportional to the magnitude of the field at that cross-section. As a result, if the field lines are close together (that is, the field line density is greater), this indicates that the magnitude of the field is

large at that point. If the field lines are far apart at the cross-section, this indicates the magnitude of the field is small. [\[link\]](#) shows the idea.



Electric field lines passing through imaginary areas. Since the number of lines passing through each area is the same, but the areas themselves are different, the field line density is different. This indicates different magnitudes of the electric field at these points.

In [\[link\]](#), the same number of field lines passes through both surfaces (S and S'), but the surface S is larger than surface S' . Therefore, the density of field lines (number of lines per unit area) is larger at the location of S' , indicating that the electric field is stronger at the location of S' than at S . The rules for creating an electric field diagram are as follows.

Note:

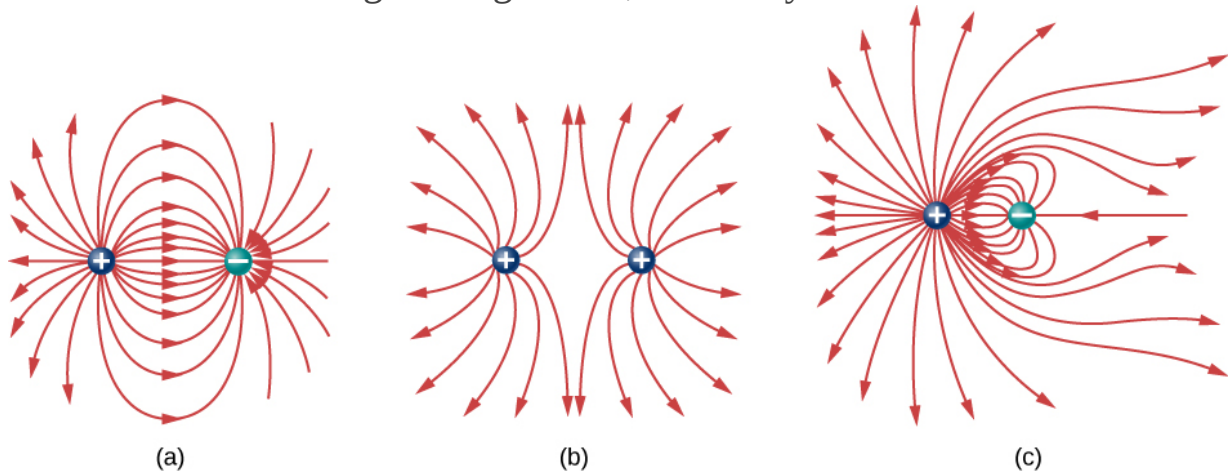
Drawing Electric Field Lines

1. Electric field lines either originate on positive charges or come in from infinity, and either terminate on negative charges or extend out to infinity.
2. The number of field lines originating or terminating at a charge is proportional to the magnitude of that charge. A charge of $2q$ will have twice as many lines as a charge of q .
3. At every point in space, the field vector at that point is tangent to the field line at that same point.
4. The field line density at any point in space is proportional to (and therefore is representative of) the magnitude of the field at that point in space.
5. Field lines can never cross. Since a field line represents the direction of the field at a given point, if two field lines crossed at some point, that would imply that the electric field was pointing in two different directions at a single point. This in turn would suggest that the (net) force on a test charge placed at that point would point in two different directions. Since this is obviously impossible, it follows that field lines must never cross.

Always keep in mind that field lines serve only as a convenient way to visualize the electric field; they are not physical entities. Although the direction and relative intensity of the electric field can be deduced from a set of field lines, the lines can also be misleading. For example, the field lines drawn to represent the electric field in a region must, by necessity, be

discrete. However, the actual electric field in that region exists at every point in space.

Field lines for three groups of discrete charges are shown in [\[link\]](#). Since the charges in parts (a) and (b) have the same magnitude, the same number of field lines are shown starting from or terminating on each charge. In (c), however, we draw three times as many field lines leaving the $+3q$ charge as entering the $-q$. The field lines that do not terminate at $-q$ emanate outward from the charge configuration, to infinity.



Three typical electric field diagrams. (a) A dipole. (b) Two identical charges. (c) Two charges with opposite signs and different magnitudes. Can you tell from the diagram which charge has the larger magnitude?

The ability to construct an accurate electric field diagram is an important, useful skill; it makes it much easier to estimate, predict, and therefore calculate the electric field of a source charge. The best way to develop this skill is with software that allows you to place source charges and then will draw the net field upon request. We strongly urge you to search the Internet for a program. Once you've found one you like, run several simulations to get the essential ideas of field diagram construction. Then practice drawing field diagrams, and checking your predictions with the computer-drawn diagrams.

Note:

One example of a [field-line drawing program](#) is from the PhET “Charges and Fields” simulation.

Summary

- Electric field diagrams assist in visualizing the field of a source charge.
- The magnitude of the field is proportional to the field line density.
- Field vectors are everywhere tangent to field lines.

Conceptual Questions

Exercise:**Problem:**

If a point charge is released from rest in a uniform electric field, will it follow a field line? Will it do so if the electric field is not uniform?

Solution:

yes; no

Exercise:**Problem:**

Under what conditions, if any, will the trajectory of a charged particle not follow a field line?

Exercise:**Problem:**

How would you experimentally distinguish an electric field from a gravitational field?

Solution:

At the surface of Earth, the gravitational field is always directed in toward Earth's center. An electric field could move a charged particle in a different direction than toward the center of Earth. This would indicate an electric field is present.

Exercise:

Problem:

A representation of an electric field shows 10 field lines perpendicular to a square plate. How many field lines should pass perpendicularly through the plate to depict a field with twice the magnitude?

Exercise:

Problem:

What is the ratio of the number of electric field lines leaving a charge $10q$ and a charge q ?

Solution:

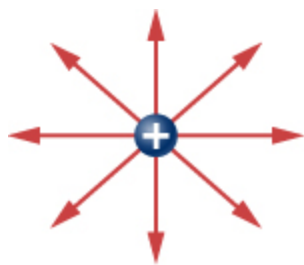
10

Problems

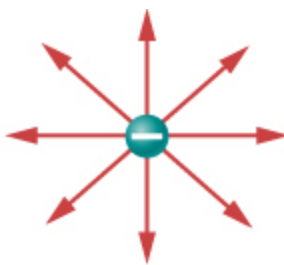
Exercise:

Problem:

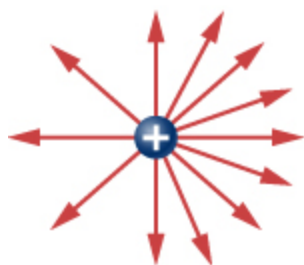
Which of the following electric field lines are incorrect for point charges? Explain why.



(a)



(b)



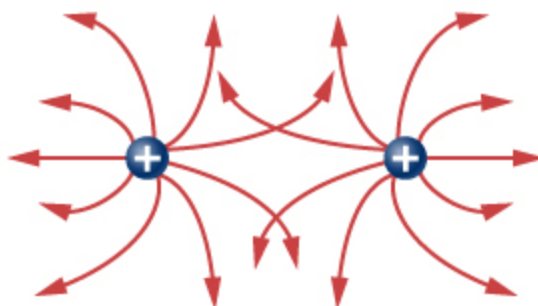
(c)



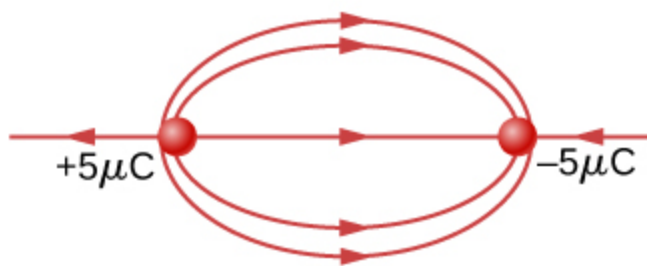
(d)



(e)



(f)



(g)

Exercise:**Problem:**

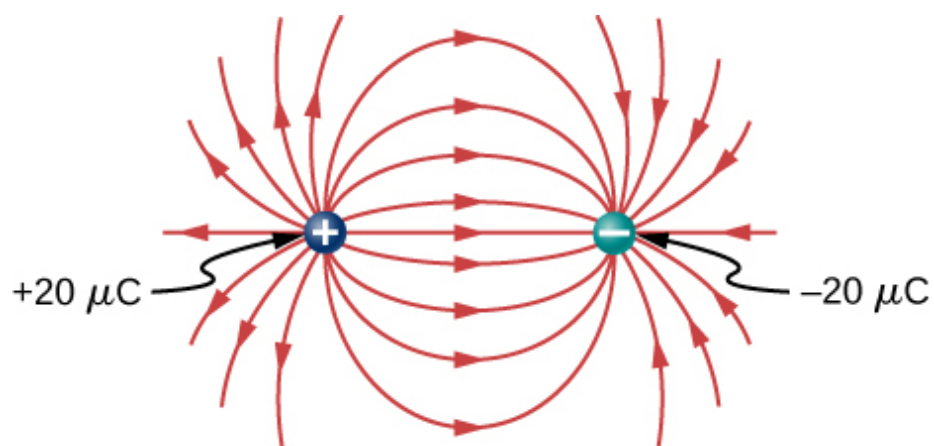
In this exercise, you will practice drawing electric field lines. Make sure you represent both the magnitude and direction of the electric field adequately. Note that the number of lines into or out of charges is proportional to the charges.

(a) Draw the electric field lines map for two charges $+20\ \mu\text{C}$ and $-20\ \mu\text{C}$ situated 5 cm from each other.

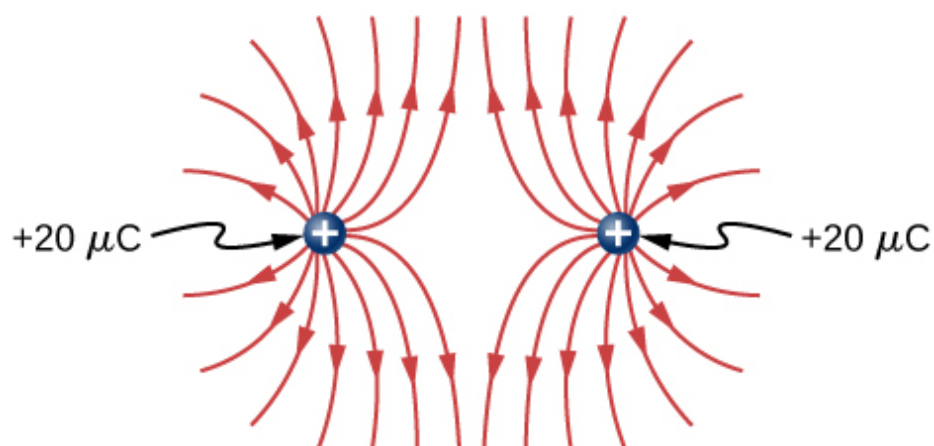
(b) Draw the electric field lines map for two charges $+20\ \mu\text{C}$ and $+20\ \mu\text{C}$ situated 5 cm from each other.

(c) Draw the electric field lines map for two charges $+20\ \mu\text{C}$ and $-30\ \mu\text{C}$ situated 5 cm from each other.

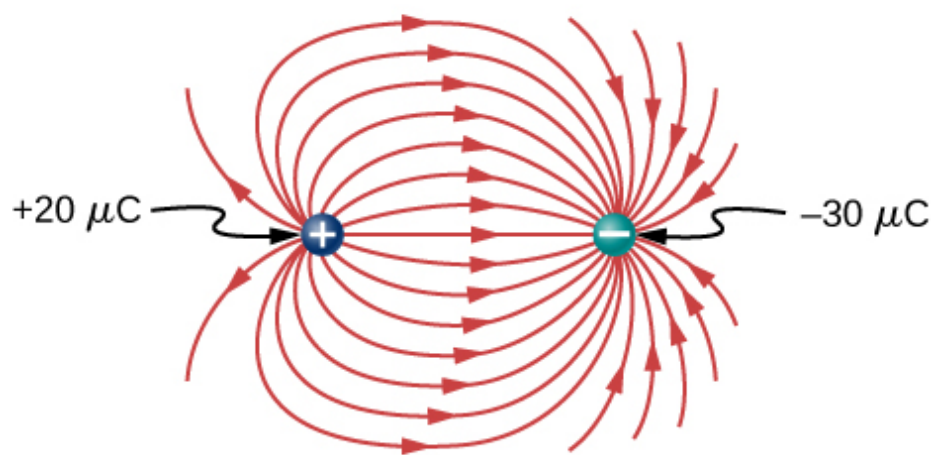
Solution:



(a)



(b)



(c)


Exercise:**Problem:**

Draw the electric field for a system of three particles of charges $+1\ \mu\text{C}$, $+2\ \mu\text{C}$, and $-3\ \mu\text{C}$ fixed at the corners of an equilateral triangle of side 2 cm.

Exercise:**Problem:**

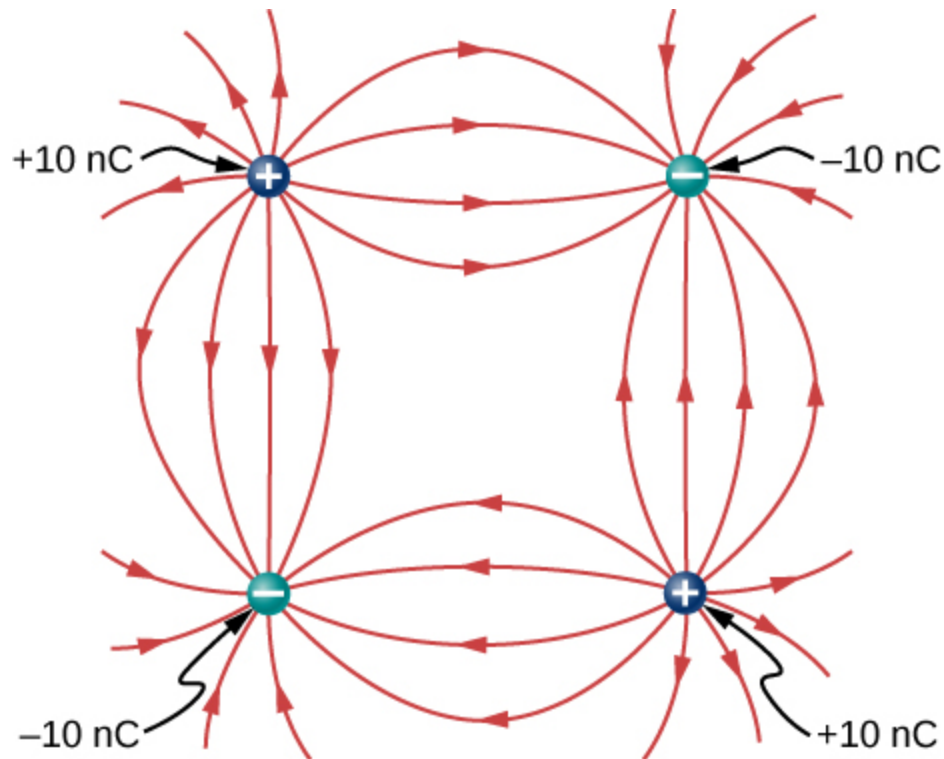
Two charges of equal magnitude but opposite sign make up an electric dipole. A quadrupole consists of two electric dipoles that are placed anti-parallel at two edges of a square as shown.

$+10\ \text{nC}$   $-10\ \text{nC}$

$-10\ \text{nC}$   $+10\ \text{nC}$

Draw the electric field of the charge distribution.

Solution:



Exercise:

Problem:

Suppose the electric field of an isolated point charge decreased with distance as $1/r^{2+\delta}$ rather than as $1/r^2$. Show that it is then impossible to draw continuous field lines so that their number per unit area is proportional to E .

Glossary

field line

smooth, usually curved line that indicates the direction of the electric field

field line density

number of field lines per square meter passing through an imaginary area; its purpose is to indicate the field strength at different points in space

Electric Dipoles

By the end of this section, you will be able to:

- Describe a permanent dipole
- Describe an induced dipole
- Define and calculate an electric dipole moment
- Explain the physical meaning of the dipole moment

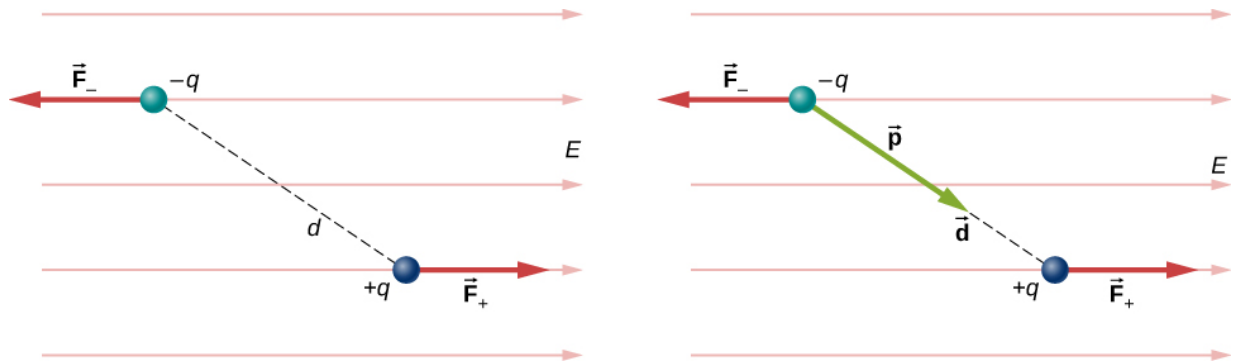
Earlier we discussed, and calculated, the electric field of a dipole: two equal and opposite charges that are “close” to each other. (In this context, “close” means that the distance d between the two charges is much, much less than the distance of the field point P , the location where you are calculating the field.) Let’s now consider what happens to a dipole when it is placed in an external field $\vec{\mathbf{E}}$. We assume that the dipole is a **permanent dipole**; it exists without the field, and does not break apart in the external field.

Rotation of a Dipole due to an Electric Field

For now, we deal with only the simplest case: The external field is uniform in space. Suppose we have the situation depicted in [\[link\]](#), where we denote the distance between the charges as the vector $\vec{\mathbf{d}}$, pointing from the negative charge to the positive charge. The forces on the two charges are equal and opposite, so there is no net force on the dipole. However, there is a torque:

Equation:

$$\begin{aligned}\vec{\tau} &= \left(\frac{\vec{\mathbf{d}}}{2} \times \vec{\mathbf{F}}_+ \right) + \left(-\frac{\vec{\mathbf{d}}}{2} \times \vec{\mathbf{F}}_- \right) \\ &= \left[\left(\frac{\vec{\mathbf{d}}}{2} \right) \times \left(+q\vec{\mathbf{E}} \right) + \left(-\frac{\vec{\mathbf{d}}}{2} \right) \times \left(-q\vec{\mathbf{E}} \right) \right] \\ &= q\vec{\mathbf{d}} \times \vec{\mathbf{E}}.\end{aligned}$$



A dipole in an external electric field. (a) The net force on the dipole is zero, but the net torque is not. As a result, the dipole rotates, becoming aligned with the external field. (b) The dipole moment is a convenient way to characterize this effect. The \vec{d} points in the same direction as \vec{p} .

The quantity $q\vec{d}$ (the magnitude of each charge multiplied by the vector distance between them) is a property of the dipole; its value, as you can see, determines the torque that the dipole experiences in the external field. It is useful, therefore, to define this product as the so-called **dipole moment** of the dipole:

Note:

Equation:

$$\vec{p} \equiv q\vec{d}.$$

We can therefore write

Note:

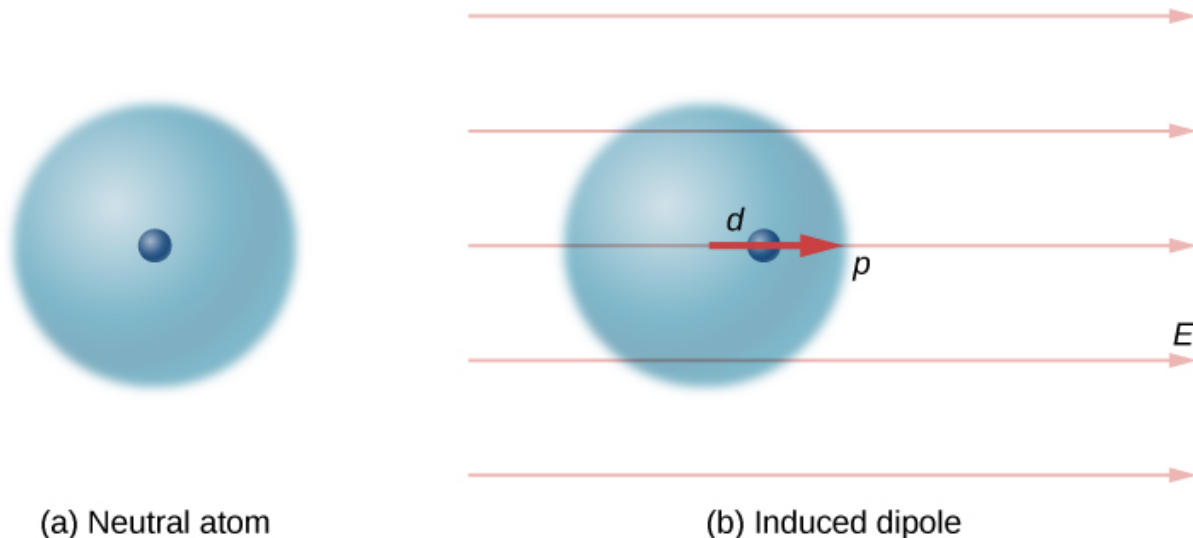
Equation:

$$\vec{\tau} = \vec{p} \times \vec{E}.$$

Recall that a torque changes the angular velocity of an object, the dipole, in this case. In this situation, the effect is to rotate the dipole (that is, align the direction of \vec{p}) so that it is parallel to the direction of the external field.

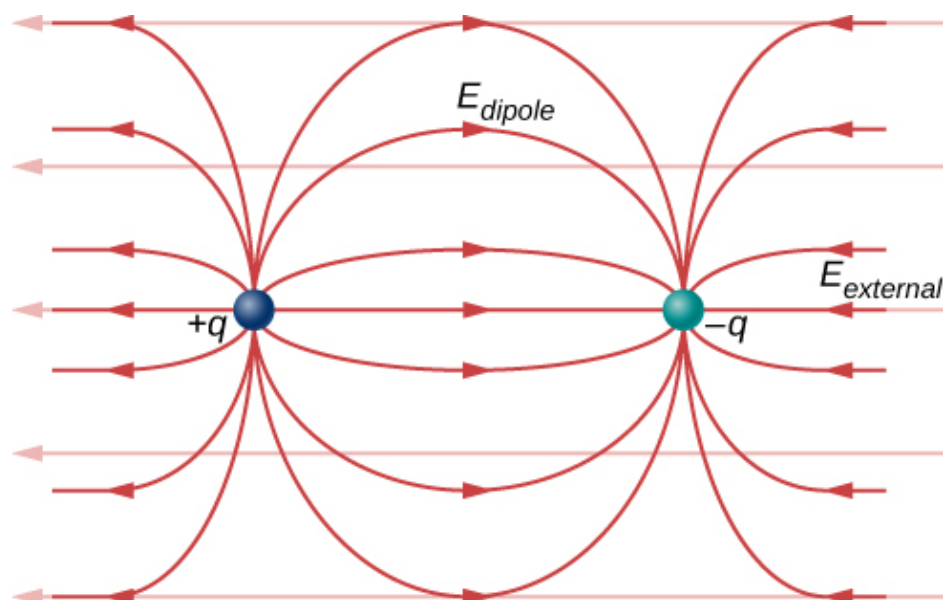
Induced Dipoles

Neutral atoms are, by definition, electrically neutral; they have equal amounts of positive and negative charge. Furthermore, since they are spherically symmetrical, they do not have a “built-in” dipole moment the way most asymmetrical molecules do. They obtain one, however, when placed in an external electric field, because the external field causes oppositely directed forces on the positive nucleus of the atom versus the negative electrons that surround the nucleus. The result is a new charge distribution of the atom, and therefore, an **induced dipole** moment ([\[link\]](#)).



A dipole is induced in a neutral atom by an external electric field. The induced dipole moment is aligned with the external field.

An important fact here is that, just as for a rotated polar molecule, the result is that the dipole moment ends up aligned parallel to the external electric field. Generally, the magnitude of an induced dipole is much smaller than that of an inherent dipole. For both kinds of dipoles, notice that once the alignment of the dipole (rotated or induced) is complete, the net effect is to decrease the total electric field $\vec{\mathbf{E}}_{\text{total}} = \vec{\mathbf{E}}_{\text{external}} + \vec{\mathbf{E}}_{\text{dipole}}$ in the regions inside the dipole charges ([\[link\]](#)). By “inside” we mean in between the charges. This effect is crucial for capacitors, as you will see in [Capacitance](#).



The net electric field is the vector sum of the field of the dipole plus the external field.

Recall that we found the electric field of a dipole in [\[link\]](#). If we rewrite it in terms of the dipole moment we get:

Equation:

$$\vec{\mathbf{E}}(z) = \frac{-1}{4\pi\epsilon_0} \frac{\vec{\mathbf{p}}}{z^3}.$$

The form of this field is shown in [\[link\]](#). Notice that along the plane perpendicular to the axis of the dipole and midway between the charges, the direction of the electric field is opposite that of the dipole and gets weaker the further from the axis one goes. Similarly, on the axis of the dipole (but outside it), the field points in the same direction as the dipole, again getting weaker the further one gets from the charges.

Summary

- If a permanent dipole is placed in an external electric field, it results in a torque that aligns it with the external field.
- If a nonpolar atom (or molecule) is placed in an external field, it gains an induced dipole that is aligned with the external field.
- The net field is the vector sum of the external field plus the field of the dipole (physical or induced).
- The strength of the polarization is described by the dipole moment of the dipole, $\vec{p} = q\vec{d}$.

Key Equations

Coulomb's law	$\vec{F}_{12}(r) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^2} \hat{r}_{12}$
Superposition of electric forces	$\vec{F}(r) = \frac{1}{4\pi\epsilon_0} Q \sum_{i=1}^N \frac{q_i}{r_i^2} \hat{r}_i$
Electric force due to an electric field	$\vec{F} = Q\vec{E}$
Electric field at point P	$\vec{E}(P) \equiv \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N \frac{q_i}{r_i^2} \hat{r}_i$
Field of an infinite wire	

	$\vec{\mathbf{E}}(z) = \frac{1}{4\pi\epsilon_0} \frac{2\lambda}{z} \hat{\mathbf{k}}$
Field of an infinite plane	$\vec{\mathbf{E}} = \frac{\sigma}{2\epsilon_0} \hat{\mathbf{k}}$
Dipole moment	$\vec{\mathbf{p}} \equiv q\vec{\mathbf{d}}$
Torque on dipole in external E-field	$\vec{\tau} = \vec{\mathbf{p}} \times \vec{\mathbf{E}}$

Conceptual Questions

Exercise:

Problem:

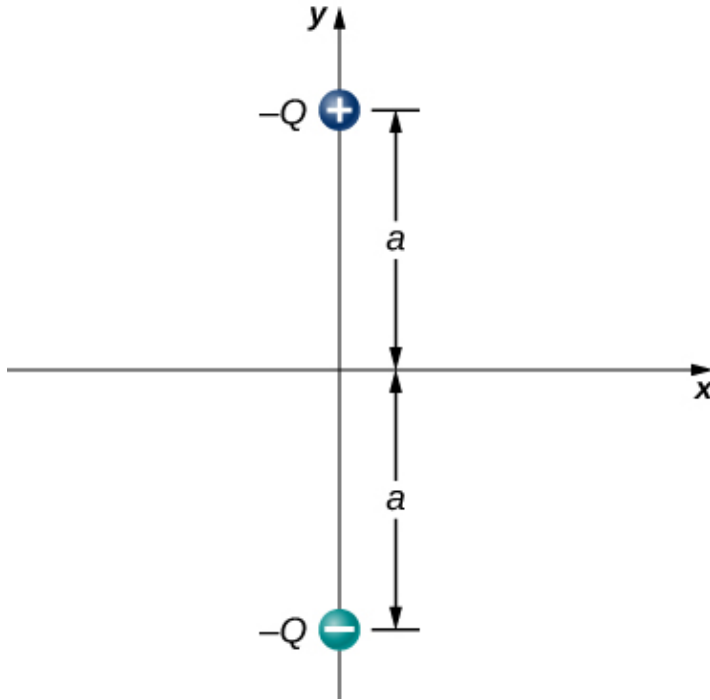
What are the stable orientation(s) for a dipole in an external electric field?
What happens if the dipole is slightly perturbed from these orientations?

Problems

Exercise:

Problem:

Consider the equal and opposite charges shown below. (a) Show that at all points on the x -axis for which $|x| \gg a$, $E \approx Qa/2\pi\epsilon_0 x^3$. (b) Show that at all points on the y -axis for which $|y| \gg a$, $E \approx Qa/\pi\epsilon_0 y^3$.



Solution:

$$\begin{aligned}
 E_x &= 0, \\
 E_y &= \frac{1}{4\pi\epsilon_0} \left[\frac{2q}{(x^2+a^2)} \frac{a}{\sqrt{(x^2+a^2)}} \right] \\
 \Rightarrow x \gg a &\Rightarrow \frac{1}{2\pi\epsilon_0} \frac{qa}{x^3}, \\
 E_y &= \frac{q}{4\pi\epsilon_0} \left[\frac{2ya+2ya}{(y-a)^2(y+a)^2} \right] \\
 \Rightarrow y \gg a &\Rightarrow \frac{1}{\pi\epsilon_0} \frac{qa}{y^3}
 \end{aligned}$$

Exercise:

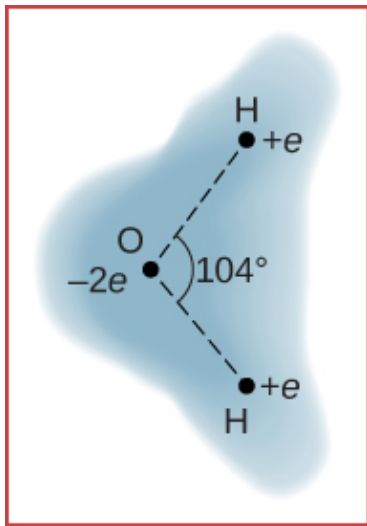
Problem:

(a) What is the dipole moment of the configuration shown above? If $Q = 4.0 \mu\text{C}$, (b) what is the torque on this dipole with an electric field of $4.0 \times 10^5 \text{ N/C}\hat{\mathbf{i}}$? (c) What is the torque on this dipole with an electric field of $-4.0 \times 10^5 \text{ N/C}\hat{\mathbf{i}}$? (d) What is the torque on this dipole with an electric field of $\pm 4.0 \times 10^5 \text{ N/C}\hat{\mathbf{j}}$?

Exercise:

Problem:

A water molecule consists of two hydrogen atoms bonded with one oxygen atom. The bond angle between the two hydrogen atoms is 104° (see below). Calculate the net dipole moment of a hypothetical water molecule where the charge at the oxygen molecule is $-2e$ and at each hydrogen atom is $+e$. The net dipole moment of the molecule is the vector sum of the individual dipole moment between the two O-Hs. The separation O-H is 0.9578 angstroms.



Solution:

The net dipole moment of the molecule is the vector sum of the individual dipole moments between the two O-H. The separation O-H is 0.9578 angstroms:

$$\vec{p} = 1.889 \times 10^{-29} \text{ Cm } \hat{i}$$

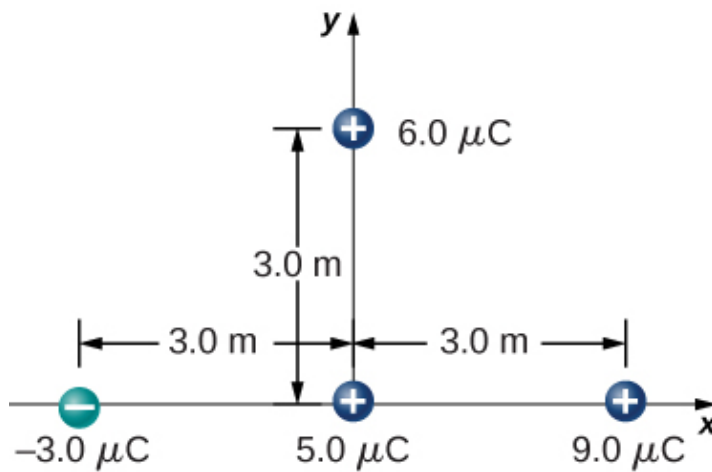
Additional Problems**Exercise:**

Problem:

Point charges $q_1 = 2.0 \mu\text{C}$ and $q_2 = 4.0 \mu\text{C}$ are located at $r_1 = (4.0\hat{i} - 2.0\hat{j} + 2.0\hat{k})\text{m}$ and $r_2 = (8.0\hat{i} + 5.0\hat{j} - 9.0\hat{k})\text{m}$. What is the force of q_2 on q_1 ?

Exercise:

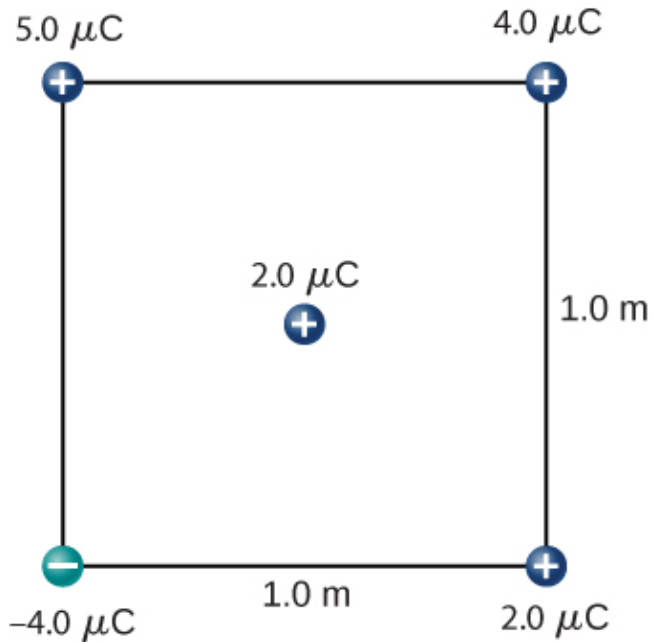
Problem: What is the force on the $5.0\text{-}\mu\text{C}$ charge shown below?

**Solution:**

$$\vec{F}_{\text{net}} = \left[-8.99 \times 10^9 \frac{3.0 \times 10^{-6}(5.0 \times 10^{-6})}{(3.0 \text{ m})^2} - 8.99 \times 10^9 \frac{9.0 \times 10^{-6}(5.0 \times 10^{-6})}{(3.0 \text{ m})^2} \right] \hat{i} - 8.99 \times 10^9 \frac{6.0 \times 10^{-6}(5.0 \times 10^{-6})}{(3.0 \text{ m})^2} \hat{j} = -0.06 \text{ N} \hat{i} - 0.03 \text{ N} \hat{j}$$

Exercise:**Problem:**

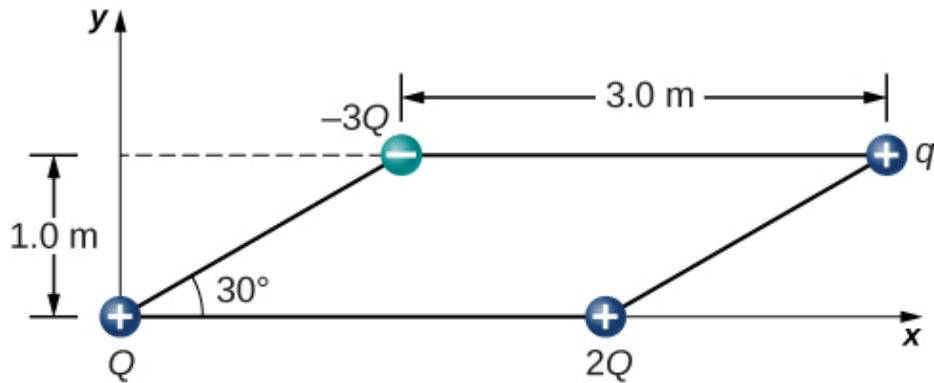
What is the force on the $2.0\text{-}\mu\text{C}$ charge placed at the center of the square shown below?



Exercise:

Problem:

Four charged particles are positioned at the corners of a parallelogram as shown below. If $q = 5.0 \mu\text{C}$ and $Q = 8.0 \mu\text{C}$, what is the net force on q ?



Solution:

Charges Q and q form a right triangle of sides 1 m and $3 + \sqrt{3} \text{ m}$.

Charges $2Q$ and q form a right triangle of sides 1 m and $\sqrt{3} \text{ m}$.

$$F_x = 0.049 \text{ N},$$

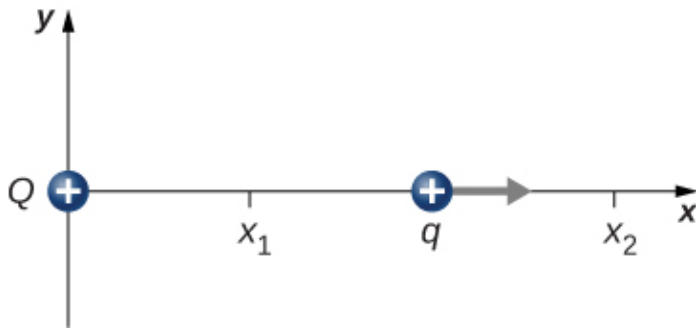
$$F_y = 0.093 \text{ N},$$

$$\vec{F}_{\text{net}} = 0.036 \text{ N } \hat{i} + 0.09 \text{ N } \hat{j}$$

Exercise:

Problem:

A charge Q is fixed at the origin and a second charge q moves along the x -axis, as shown below. How much work is done on q by the electric force when q moves from x_1 to x_2 ?



Exercise:

Problem:

A charge $q = -2.0 \mu\text{C}$ is released from rest when it is 2.0 m from a fixed charge $Q = 6.0 \mu\text{C}$. What is the kinetic energy of q when it is 1.0 m from Q ?

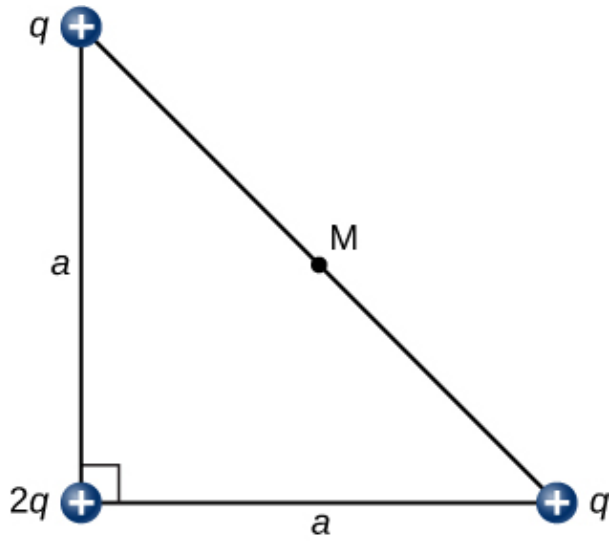
Solution:

$$W = 0.054 \text{ J}$$

Exercise:

Problem:

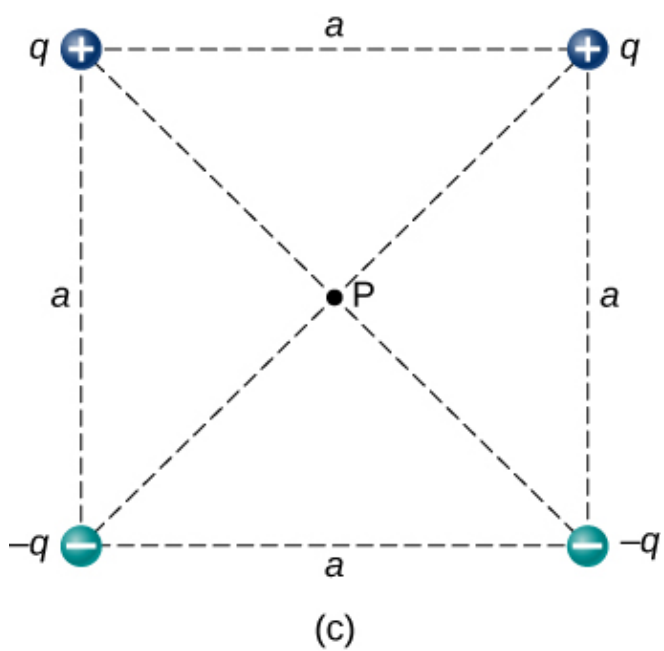
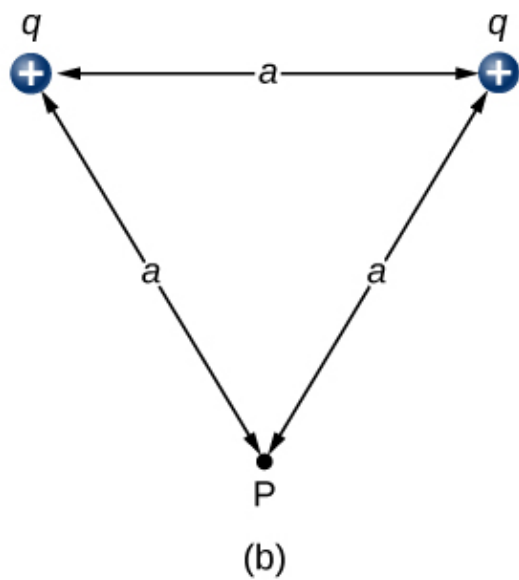
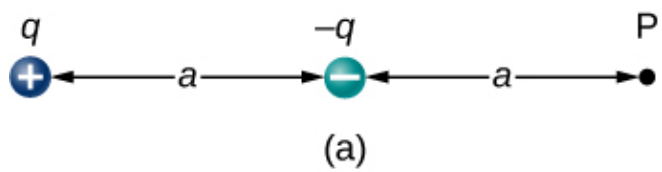
What is the electric field at the midpoint M of the hypotenuse of the triangle shown below?



Exercise:

Problem:

Find the electric field at P for the charge configurations shown below.



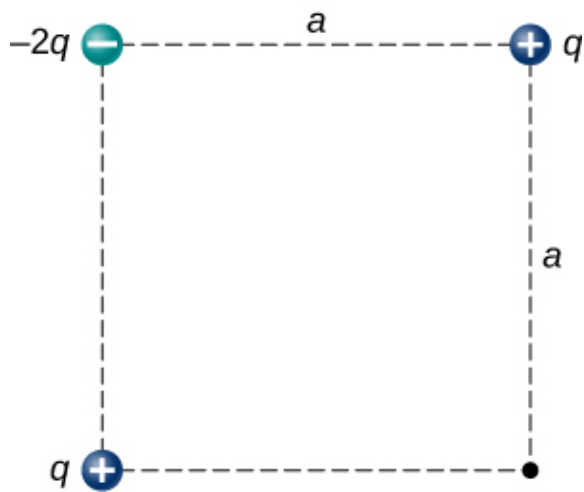
Solution:

a. $\vec{E} = \frac{1}{4\pi\epsilon_0} \left(\frac{q}{(2a)^2} - \frac{q}{a^2} \right) \hat{i}$; b. $\vec{E} = \frac{\sqrt{3}}{4\pi\epsilon_0} \frac{q}{a^2} (-\hat{j})$; c.
 $\vec{E} = \frac{2}{\pi\epsilon_0} \frac{q}{a^2} \frac{1}{\sqrt{2}} (-\hat{j})$

Exercise:

Problem:

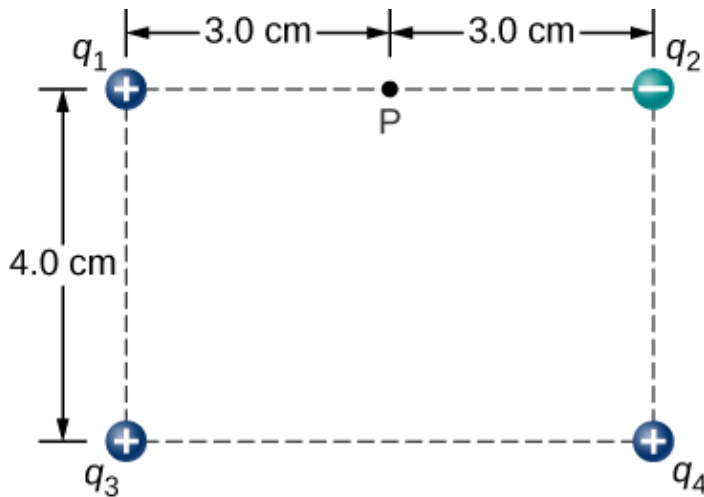
(a) What is the electric field at the lower-right-hand corner of the square shown below? (b) What is the force on a charge q placed at that point?



Exercise:

Problem:

Point charges are placed at the four corners of a rectangle as shown below:
 $q_1 = 2.0 \times 10^{-6} \text{ C}$, $q_2 = -2.0 \times 10^{-6} \text{ C}$, $q_3 = 4.0 \times 10^{-6} \text{ C}$, and
 $q_4 = 1.0 \times 10^{-6} \text{ C}$. What is the electric field at P ?



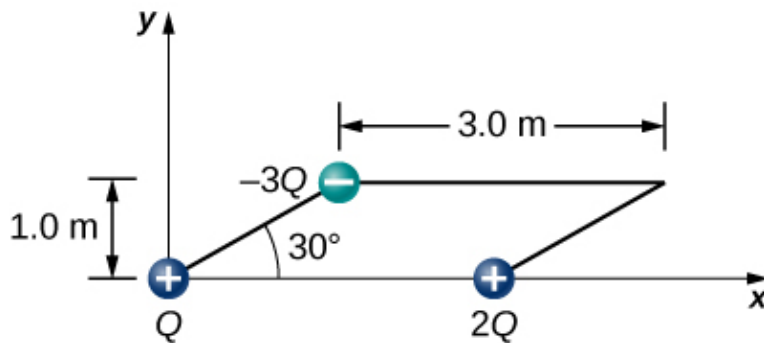
Solution:

$$\vec{E}_{\text{net}} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 + \vec{E}_4 = (4.65\hat{i} + 1.44\hat{j}) \times 10^7 \text{ N/C}$$

Exercise:

Problem:

Three charges are positioned at the corners of a parallelogram as shown below. (a) If $Q = 8.0 \mu\text{C}$, what is the electric field at the unoccupied corner? (b) What is the force on a $5.0\text{-}\mu\text{C}$ charge placed at this corner?



Exercise:

Problem:

A positive charge q is released from rest at the origin of a rectangular coordinate system and moves under the influence of the electric field $\vec{E} = E_0 (1 + x/a)\hat{i}$. What is the kinetic energy of q when it passes through $x = 3a$?

Solution:

$$F = qE_0 (1 + x/a) \quad W = \frac{1}{2}m(v^2 - v_0^2),$$

$$\frac{1}{2}mv^2 = qE_0\left(\frac{15a}{2}\right) \text{ J}$$

Exercise:**Problem:**

A particle of charge $-q$ and mass m is placed at the center of a uniformly charged ring of total charge Q and radius R . The particle is displaced a small distance along the axis perpendicular to the plane of the ring and released. Assuming that the particle is constrained to move along the axis, show that the particle oscillates in simple harmonic motion with a frequency $f = \frac{1}{2\pi} \sqrt{\frac{qQ}{4\pi\epsilon_0 m R^3}}$.

Exercise:**Problem:**

Charge is distributed uniformly along the entire y -axis with a density λ_y and along the positive x -axis from $x = a$ to $x = b$ with a density λ_x . What is the force between the two distributions?

Solution:

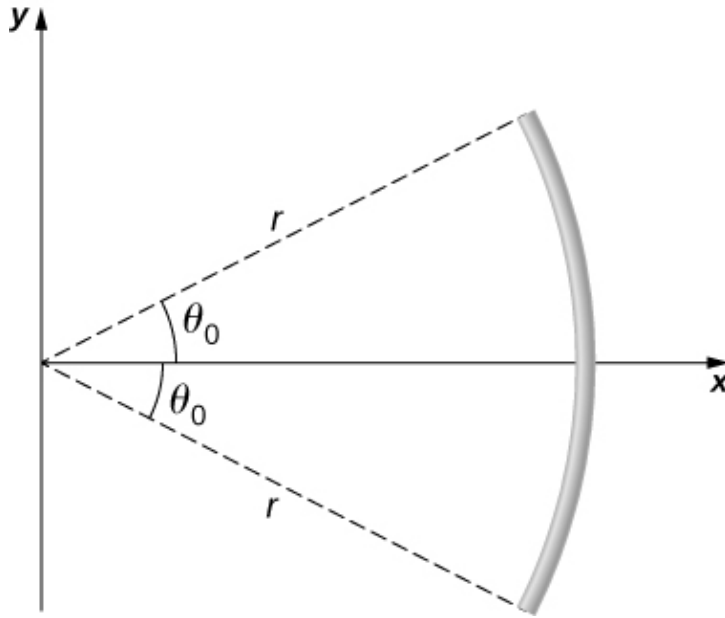
Electric field of wire at x : $\vec{E}(x) = \frac{1}{4\pi\epsilon_0} \frac{2\lambda_y}{x} \hat{i}$,

$$dF = \frac{\lambda_y \lambda_x}{2\pi\epsilon_0} (\ln b - \ln a)$$

Exercise:

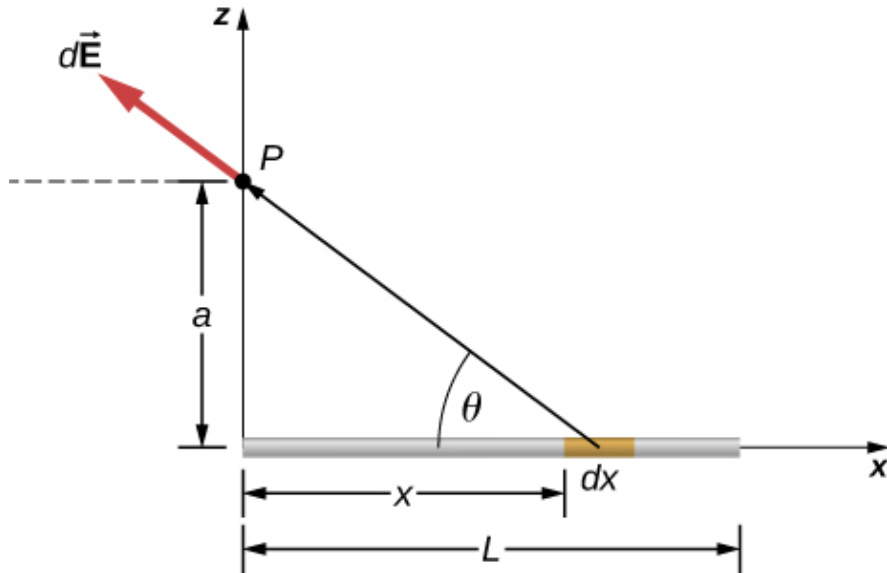
Problem:

The circular arc shown below carries a charge per unit length $\lambda = \lambda_0 \cos \theta$, where θ is measured from the x -axis. What is the electric field at the origin?

**Exercise:****Problem:**

Calculate the electric field due to a uniformly charged rod of length L , aligned with the x -axis with one end at the origin; at a point P on the z -axis.

Solution:



$$dE_x = \frac{1}{4\pi\epsilon_0} \frac{\lambda dx}{(x^2+a^2)^{3/2}} \frac{x}{\sqrt{x^2+a^2}},$$

$$\vec{E}_x = \frac{\lambda}{4\pi\epsilon_0} \left[\frac{1}{\sqrt{L^2+a^2}} - \frac{1}{a} \right] \hat{\mathbf{i}},$$

$$dE_z = \frac{1}{4\pi\epsilon_0} \frac{\lambda dx}{(x^2+a^2)^{3/2}} \frac{a}{\sqrt{x^2+a^2}},$$

$$\vec{E}_z = \frac{\lambda}{4\pi\epsilon_0 a} \frac{L}{\sqrt{L^2+a^2}} \hat{\mathbf{k}},$$

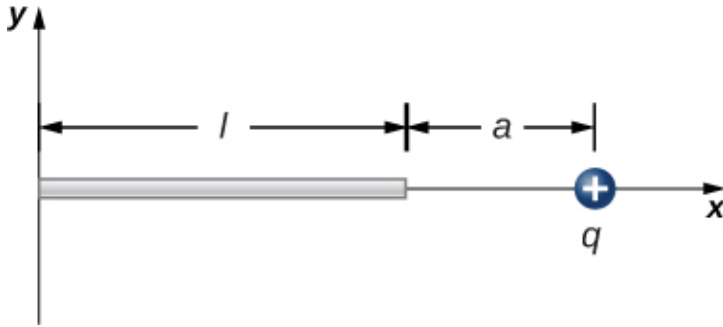
Substituting \$z\$ for \$a\$, we have:

$$\vec{E}(z) = \frac{\lambda}{4\pi\epsilon_0} \left[\frac{1}{\sqrt{L^2+z^2}} - \frac{1}{z} \right] \hat{\mathbf{i}} + \frac{\lambda}{4\pi\epsilon_0 z} \frac{L}{\sqrt{L^2+z^2}} \hat{\mathbf{k}}$$

Exercise:

Problem:

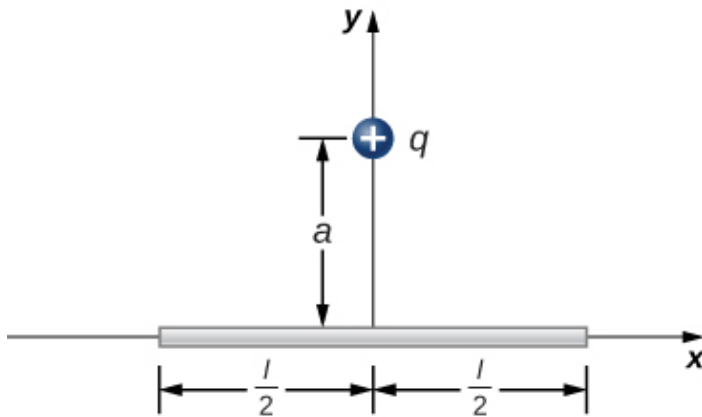
The charge per unit length on the thin rod shown below is \$\lambda\$. What is the electric force on the point charge \$q\$? Solve this problem by first considering the electric force \$d\vec{F}\$ on \$q\$ due to a small segment \$dx\$ of the rod, which contains charge \$\lambda dx\$. Then, find the net force by integrating \$d\vec{F}\$ over the length of the rod.



Exercise:

Problem:

The charge per unit length on the thin rod shown here is λ . What is the electric force on the point charge q ? (See the preceding problem.)



Solution:

There is a net force only in the y -direction. Let θ be the angle the vector from dx to q makes with the x -axis. The components along the x -axis cancel due to symmetry, leaving the y -component of the force.

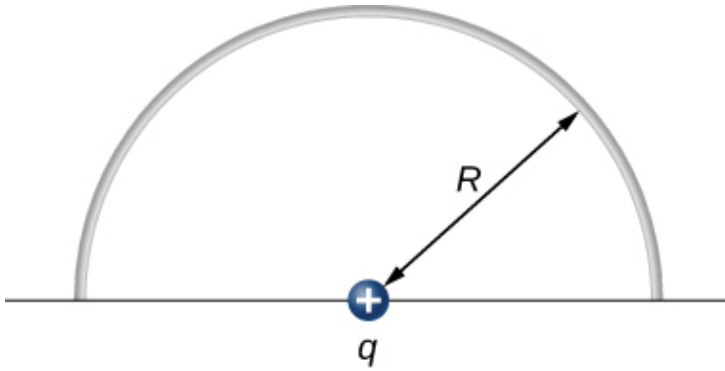
$$dF_y = \frac{1}{4\pi\epsilon_0} \frac{aq\lambda dx}{(x^2 + a^2)^{3/2}},$$

$$F_y = \frac{1}{2\pi\epsilon_0} \frac{q\lambda}{a} \left[\frac{l/2}{((l/2)^2 + a^2)^{1/2}} \right]$$

Exercise:

Problem:

The charge per unit length on the thin semicircular wire shown below is λ . What is the electric force on the point charge q ? (See the preceding problems.)

**Glossary**

dipole moment

property of a dipole; it characterizes the combination of distance between the opposite charges, and the magnitude of the charges

induced dipole

typically an atom, or a spherically symmetric molecule; a dipole created due to opposite forces displacing the positive and negative charges

permanent dipole

typically a molecule; a dipole created by the arrangement of the charged particles from which the dipole is created

Introduction

class="introduction"

This chapter introduces the concept of flux, which relates a physical quantity and the area through which it is flowing.

Although we introduce this concept with the electric field, the concept may be used for many other quantities, such as fluid flow. (credit: modification of work by “Alessandro”/Flickr)



Flux is a general and broadly applicable concept in physics. However, in this chapter, we concentrate on the flux of the electric field. This allows us to introduce Gauss's law, which is particularly useful for finding the electric fields of charge distributions exhibiting spatial symmetry. The main topics discussed here are

1. **Electric flux.** We define electric flux for both open and closed surfaces.
2. **Gauss's law.** We derive Gauss's law for an arbitrary charge distribution and examine the role of electric flux in Gauss's law.
3. **Calculating electric fields with Gauss's law.** The main focus of this chapter is to explain how to use Gauss's law to find the electric fields of spatially symmetrical charge distributions. We discuss the importance of choosing a Gaussian surface and provide examples involving the applications of Gauss's law.
4. **Electric fields in conductors.** Gauss's law provides useful insight into the absence of electric fields in conducting materials.

So far, we have found that the electrostatic field begins and ends at point charges and that the field of a point charge varies inversely with the square of the distance from that charge. These characteristics of the electrostatic field lead to an important mathematical relationship known as Gauss's law. This law is named in honor of the extraordinary German mathematician and scientist Karl Friedrich Gauss ([\[link\]](#)). Gauss's law gives us an elegantly simple way of finding the electric field, and, as you will see, it can be much easier to use than the integration method described in the previous chapter. However, there is a catch—Gauss's law has a limitation in that, while always true, it can be readily applied only for charge distributions with certain symmetries.



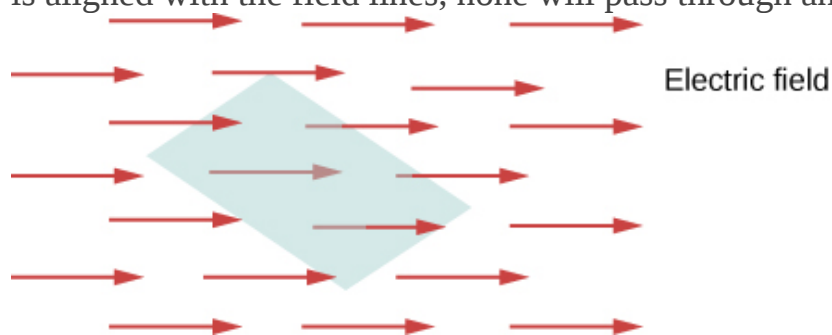
Karl Friedrich Gauss (1777–1855) was a legendary mathematician of the nineteenth century. Although his major contributions were to the field of mathematics, he also did important work in physics and astronomy.

Electric Flux

By the end of this section, you will be able to:

- Define the concept of flux
- Describe electric flux
- Calculate electric flux for a given situation

The concept of **flux** describes how much of something goes through a given area. More formally, it is the dot product of a vector field (in this chapter, the electric field) with an area. You may conceptualize the flux of an electric field as a measure of the number of electric field lines passing through an area ([\[link\]](#)). The larger the area, the more field lines go through it and, hence, the greater the flux; similarly, the stronger the electric field is (represented by a greater density of lines), the greater the flux. On the other hand, if the area rotated so that the plane is aligned with the field lines, none will pass through and there will be no flux.



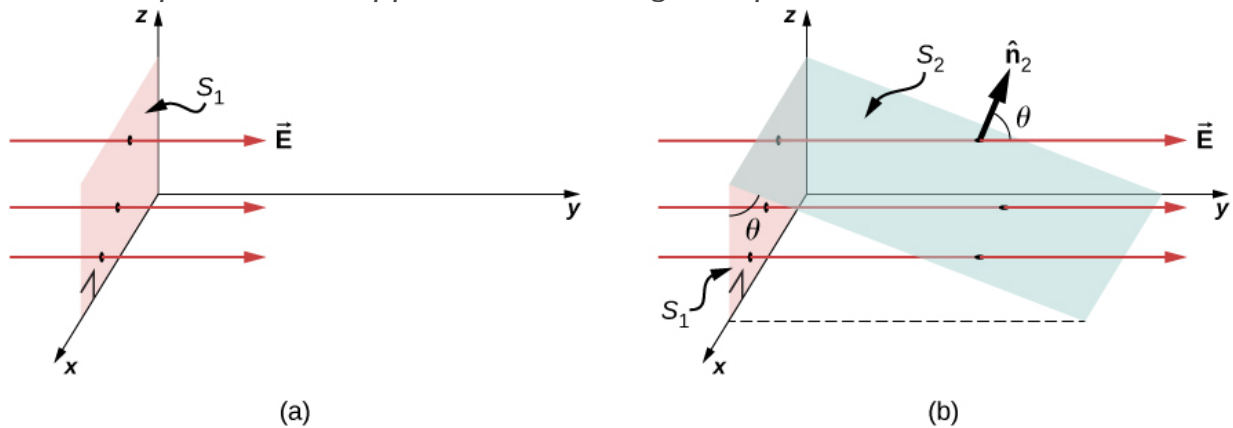
The flux of an electric field through the shaded area captures information about the “number” of electric field lines passing through the area. The numerical value of the electric flux depends on the magnitudes of the electric field and the area, as well as the relative orientation of the area with respect to the direction of the electric field.

A macroscopic analogy that might help you imagine this is to put a hula hoop in a flowing river. As you change the angle of the hoop relative to the direction of the current, more or less of the flow will go through the hoop. Similarly, the amount of flow through the hoop depends on the strength of the current and the size of the hoop. Again, flux is a general concept; we can also use it to describe the amount

of sunlight hitting a solar panel or the amount of energy a telescope receives from a distant star, for example.

To quantify this idea, [\[link\]](#)(a) shows a planar surface S_1 of area A_1 that is perpendicular to the uniform electric field $\vec{E} = E\hat{y}$. If N field lines pass through S_1 , then we know from the definition of electric field lines ([Electric Charges and Fields](#)) that $N/A_1 \propto E$, or $N \propto EA_1$.

The quantity EA_1 is the **electric flux** through S_1 . We represent the electric flux through an open surface like S_1 by the symbol Φ . Electric flux is a scalar quantity and has an SI unit of newton-meters squared per coulomb ($\text{N} \cdot \text{m}^2/\text{C}$). Notice that $N \propto EA_1$ may also be written as $N \propto \Phi$, demonstrating that *electric flux is a measure of the number of field lines crossing a surface*.



(a) A planar surface S_1 of area A_1 is perpendicular to the electric field $E\hat{j}$. N field lines cross surface S_1 . (b) A surface S_2 of area A_2 whose projection onto the xz -plane is S_1 . The same number of field lines cross each surface.

Now consider a planar surface that is not perpendicular to the field. How would we represent the electric flux? [\[link\]](#)(b) shows a surface S_2 of area A_2 that is inclined at an angle θ to the xz -plane and whose projection in that plane is S_1 (area A_1). The areas are related by $A_2 \cos \theta = A_1$. Because the same number of field lines crosses both S_1 and S_2 , the fluxes through both surfaces must be the same. The flux through S_2 is therefore $\Phi = EA_1 = EA_2 \cos \theta$. Designating \hat{n}_2 as a unit vector normal to S_2 (see [\[link\]](#)(b)), we obtain

Equation:

$$\Phi = \vec{\mathbf{E}} \cdot \hat{\mathbf{n}}_2 A_2.$$

Note:

Check out this [video](#) to observe what happens to the flux as the area changes in size and angle, or the electric field changes in strength.

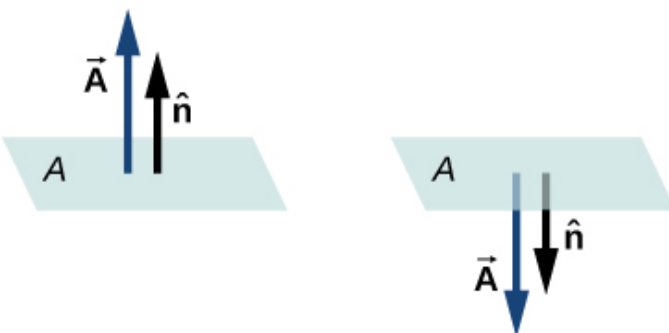
Area Vector

For discussing the flux of a vector field, it is helpful to introduce an area vector $\vec{\mathbf{A}}$. This allows us to write the last equation in a more compact form. What should the magnitude of the area vector be? What should the direction of the area vector be? What are the implications of how you answer the previous question?

The **area vector** of a flat surface of area A has the following magnitude and direction:

- Magnitude is equal to area (A)
- Direction is along the normal to the surface ($\hat{\mathbf{n}}$); that is, perpendicular to the surface.

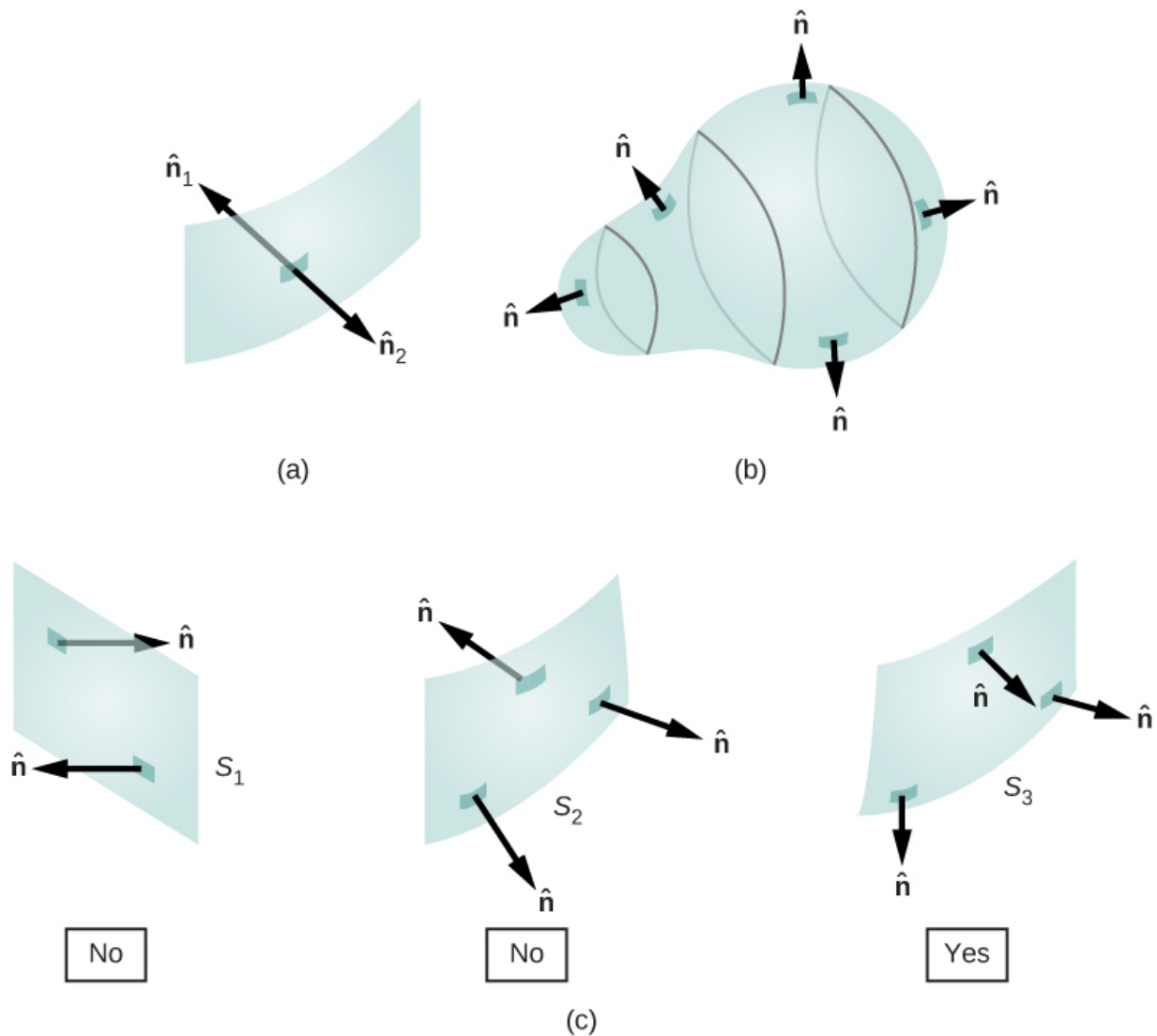
Since the normal to a flat surface can point in either direction from the surface, the direction of the area vector of an open surface needs to be chosen, as shown in [\[link\]](#).



The direction of the area vector of an

open surface needs to be chosen; it could be either of the two cases displayed here. The area vector of a part of a closed surface is defined to point from the inside of the closed space to the outside. This rule gives a unique direction.

Since $\hat{\mathbf{n}}$ is a unit normal to a surface, it has two possible directions at every point on that surface ([link](#)(a)). For an open surface, we can use either direction, as long as we are consistent over the entire surface. Part (c) of the figure shows several cases.



(a) Two potential normal vectors arise at every point on a surface. (b) The outward normal is used to calculate the flux through a closed surface. (c) Only S_3 has been given a consistent set of normal vectors that allows us to define the flux through the surface.

However, if a surface is closed, then the surface encloses a volume. In that case, the direction of the normal vector at any point on the surface points from the inside to the outside. On a *closed surface* such as that of [link](#)(b), \hat{n} is chosen to be the *outward normal* at every point, to be consistent with the sign convention for electric charge.

Electric Flux

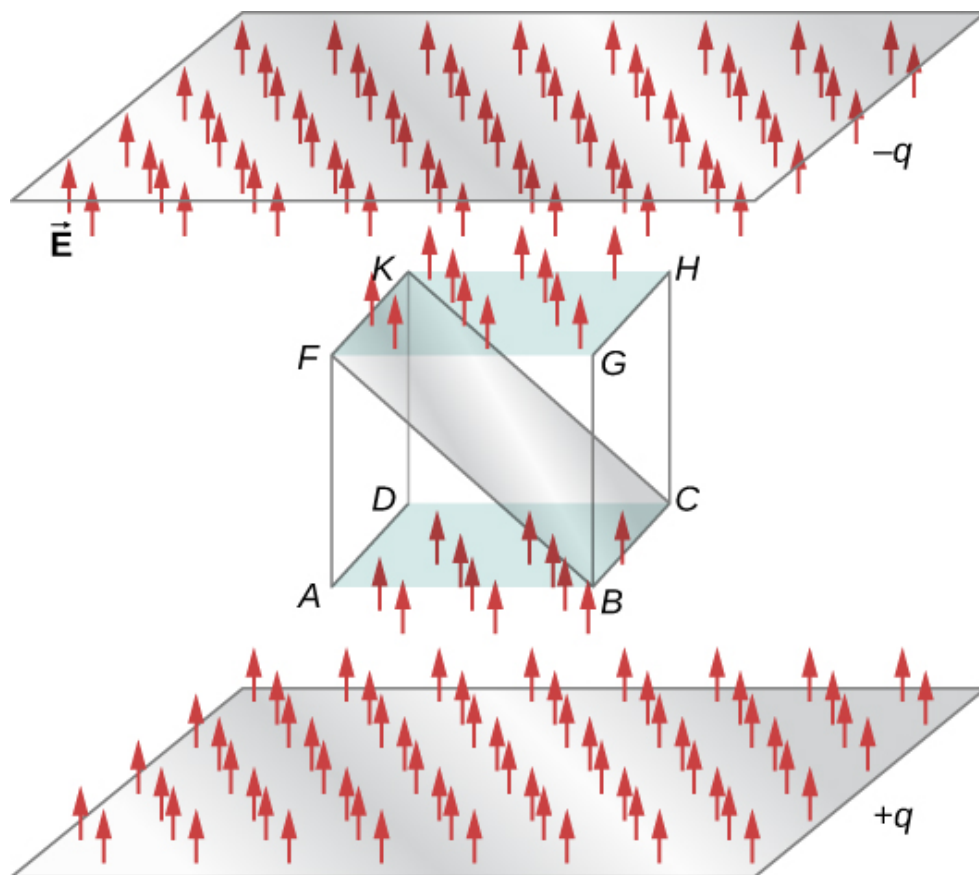
Now that we have defined the area vector of a surface, we can define the electric flux of a uniform electric field through a flat area as the scalar product of the electric field and the area vector, as defined in [Products of Vectors](#):

Note:

Equation:

$$\Phi = \vec{\mathbf{E}} \cdot \vec{\mathbf{A}} \text{ (uniform } \vec{\mathbf{E}}, \text{ flat surface).}$$

[\[link\]](#) shows the electric field of an oppositely charged, parallel-plate system and an imaginary box between the plates. The electric field between the plates is uniform and points from the positive plate toward the negative plate. A calculation of the flux of this field through various faces of the box shows that the net flux through the box is zero. Why does the flux cancel out here?



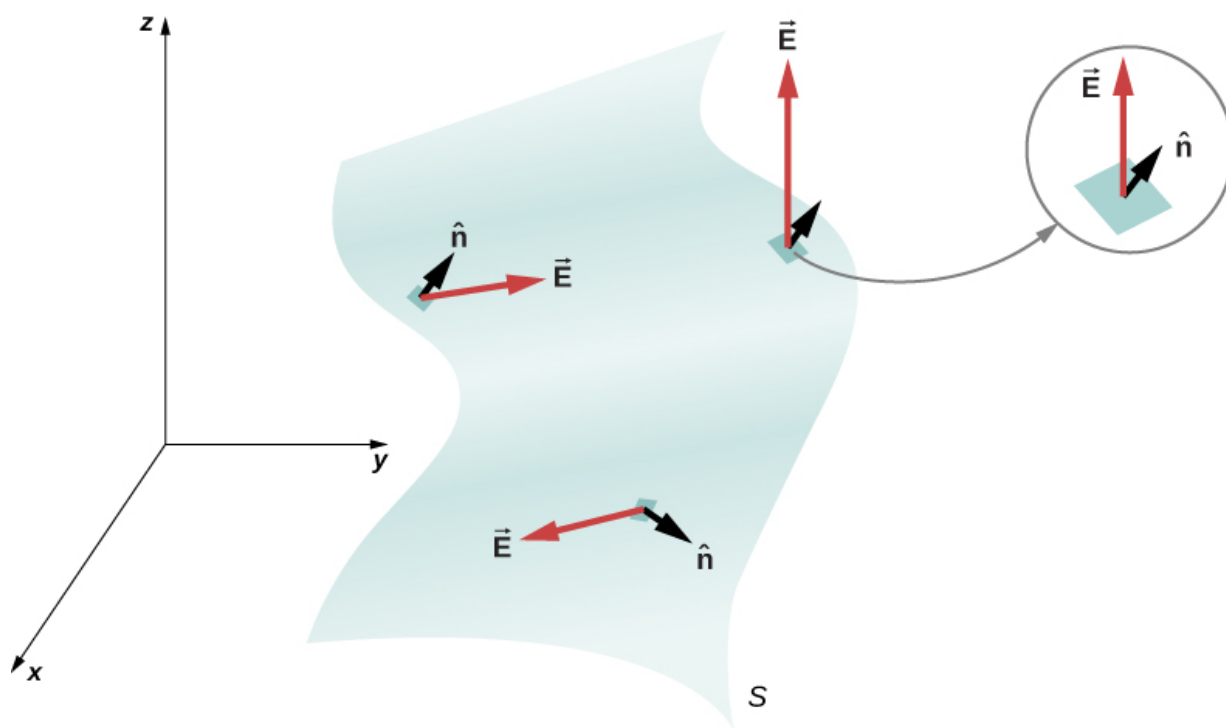
Electric flux through a cube, placed between two charged plates. Electric flux through the bottom face ($ABCD$) is

negative, because \vec{E} is in the opposite direction to the normal to the surface. The electric flux through the top face ($EFGH$) is positive, because the electric field and the normal are in the same direction. The electric flux through the other faces is zero, since the electric field is perpendicular to the normal vectors of those faces. The net electric flux through the cube is the sum of fluxes through the six faces. Here, the net flux through the cube is equal to zero. The magnitude of the flux through rectangle $BCKF$ is equal to the magnitudes of the flux through both the top and bottom faces.

The reason is that the sources of the electric field are outside the box. Therefore, if any electric field line enters the volume of the box, it must also exit somewhere on the surface because there is no charge inside for the lines to land on. Therefore,

quite generally, electric flux through a closed surface is zero if there are no sources of electric field, whether positive or negative charges, inside the enclosed volume. In general, when field lines leave (or “flow out of”) a closed surface, Φ is positive; when they enter (or “flow into”) the surface, Φ is negative.

Any smooth, non-flat surface can be replaced by a collection of tiny, approximately flat surfaces, as shown in [\[link\]](#). If we divide a surface S into small patches, then we notice that, as the patches become smaller, they can be approximated by flat surfaces. This is similar to the way we treat the surface of Earth as locally flat, even though we know that globally, it is approximately spherical.



A surface is divided into patches to find the flux.

To keep track of the patches, we can number them from 1 through N . Now, we define the area vector for each patch as the area of the patch pointed in the direction of the normal. Let us denote the area vector for the i th patch by $\delta \vec{\mathbf{A}}_i$. (We have used the symbol δ to remind us that the area is of an arbitrarily small

patch.) With sufficiently small patches, we may approximate the electric field over any given patch as uniform. Let us denote the average electric field at the location of the i th patch by $\vec{\mathbf{E}}_i$.

Equation:

$$\vec{\mathbf{E}}_i = \text{average electric field over the } i\text{th patch.}$$

Therefore, we can write the electric flux Φ_i through the area of the i th patch as

Equation:

$$\Phi_i = \vec{\mathbf{E}}_i \cdot \delta\vec{\mathbf{A}}_i \text{ (} i\text{th patch).}$$

The flux through each of the individual patches can be constructed in this manner and then added to give us an estimate of the net flux through the entire surface S , which we denote simply as Φ .

Equation:

$$\Phi = \sum_{i=1}^N \Phi_i = \sum_{i=1}^N \vec{\mathbf{E}}_i \cdot \delta\vec{\mathbf{A}}_i \text{ (} N \text{ patch estimate).}$$

This estimate of the flux gets better as we decrease the size of the patches. However, when you use smaller patches, you need more of them to cover the same surface. In the limit of infinitesimally small patches, they may be considered to have area dA and unit normal $\hat{\mathbf{n}}$. Since the elements are infinitesimal, they may be assumed to be planar, and $\vec{\mathbf{E}}_i$ may be taken as constant over any element. Then the flux $d\Phi$ through an area dA is given by $d\Phi = \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA$. It is positive when the angle between $\vec{\mathbf{E}}_i$ and $\hat{\mathbf{n}}$ is less than 90° and negative when the angle is greater than 90° . The net flux is the sum of the infinitesimal flux elements over the entire surface. With infinitesimally small patches, you need infinitely many patches, and the limit of the sum becomes a surface integral. With \int_S representing the integral over S ,

Note:

Equation:

$$\Phi = \int_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = \int_S \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} \text{ (open surface).}$$

In practical terms, surface integrals are computed by taking the antiderivatives of both dimensions defining the area, with the edges of the surface in question being the bounds of the integral.

To distinguish between the flux through an open surface like that of [\[link\]](#) and the flux through a closed surface (one that completely bounds some volume), we represent flux through a closed surface by

Note:

Equation:

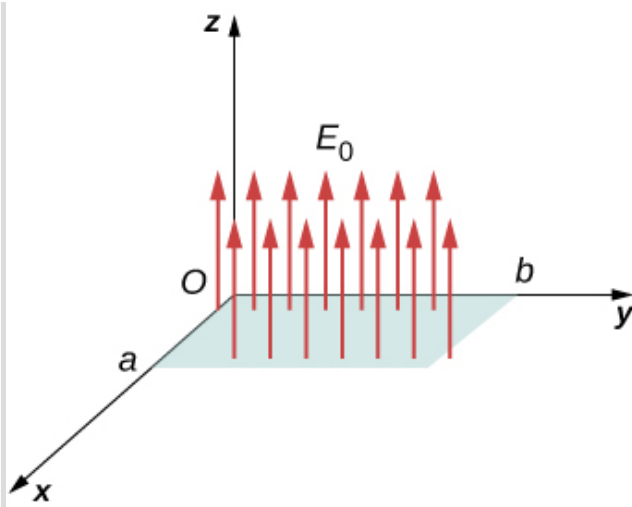
$$\Phi = \oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = \oint_S \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} \text{ (closed surface)}$$

where the circle through the integral symbol simply means that the surface is closed, and we are integrating over the entire thing. If you only integrate over a portion of a closed surface, that means you are treating a subset of it as an open surface.

Example:

Flux of a Uniform Electric Field

A constant electric field of magnitude E_0 points in the direction of the positive z -axis ([\[link\]](#)). What is the electric flux through a rectangle with sides a and b in the (a) xy -plane and in the (b) xz -plane?



Calculating the flux of E_0 through a rectangular surface.

Strategy

Apply the definition of flux: $\Phi = \vec{\mathbf{E}} \cdot \vec{\mathbf{A}}$ (uniform $\vec{\mathbf{E}}$), where the definition of dot product is crucial.

Solution

- In this case, $\Phi = \vec{\mathbf{E}}_0 \cdot \vec{\mathbf{A}} = E_0 A = E_0 ab$.
- Here, the direction of the area vector is either along the positive y-axis or toward the negative y-axis. Therefore, the scalar product of the electric field with the area vector is zero, giving zero flux.

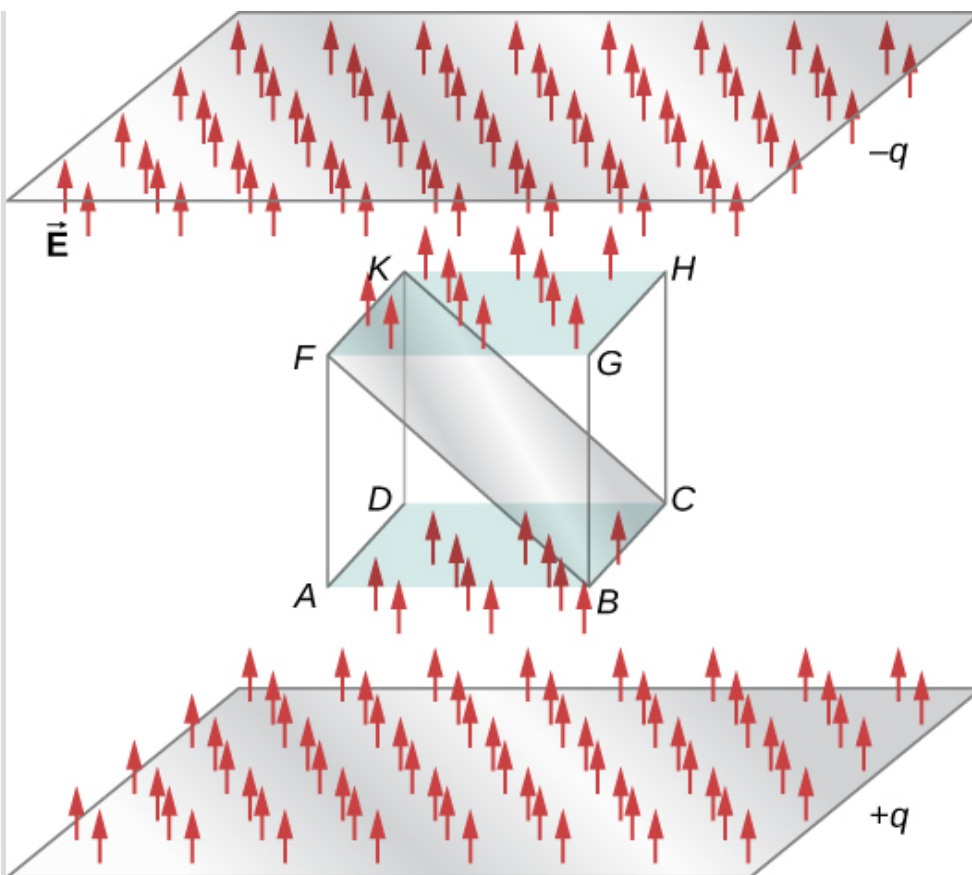
Significance

The relative directions of the electric field and area can cause the flux through the area to be zero.

Example:

Flux of a Uniform Electric Field through a Closed Surface

A constant electric field of magnitude E_0 points in the direction of the positive z-axis ([link](#)). What is the net electric flux through a cube?



Calculating the flux of E_0 through a closed cubic surface.

Strategy

Apply the definition of flux: $\Phi = \vec{E} \cdot \vec{A}$ (uniform \vec{E}), noting that a closed surface eliminates the ambiguity in the direction of the area vector.

Solution

Through the top face of the cube, $\Phi = \vec{E}_0 \cdot \vec{A} = E_0 A$.

Through the bottom face of the cube, $\Phi = \vec{E}_0 \cdot \vec{A} = -E_0 A$, because the area vector here points downward.

Along the other four sides, the direction of the area vector is perpendicular to the direction of the electric field. Therefore, the scalar product of the electric field with the area vector is zero, giving zero flux.

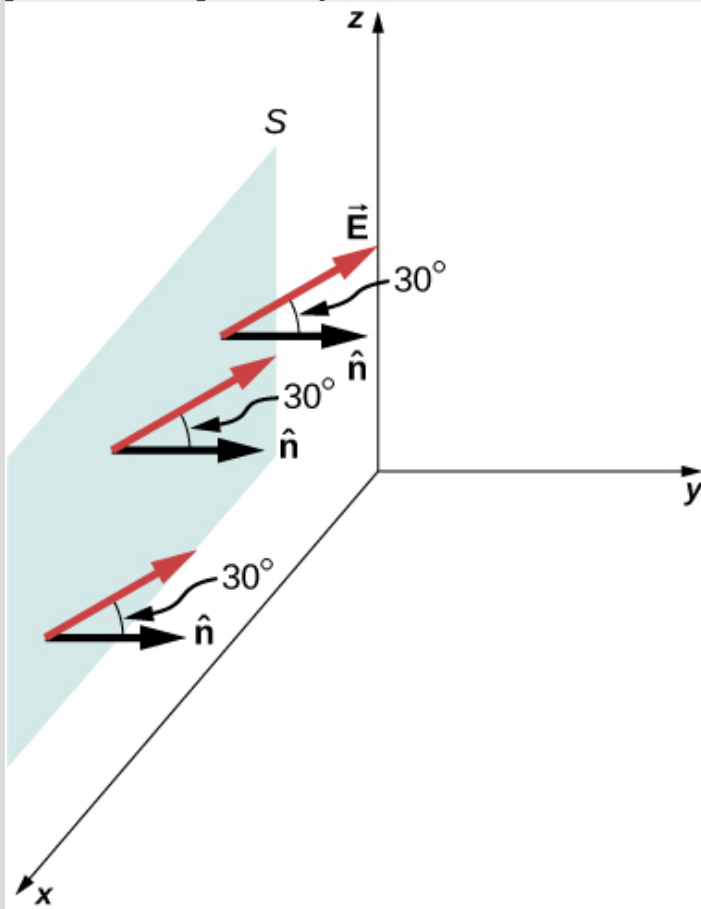
The net flux is $\Phi_{\text{net}} = E_0 A - E_0 A + 0 + 0 + 0 + 0 = 0$.

Significance

The net flux of a uniform electric field through a closed surface is zero.

Example:**Electric Flux through a Plane, Integral Method**

A uniform electric field \vec{E} of magnitude 10 N/C is directed parallel to the yz -plane at 30° above the xy -plane, as shown in [\[link\]](#). What is the electric flux through the plane surface of area 6.0 m^2 located in the xz -plane? Assume that \hat{n} points in the positive y -direction.



The electric field produces a net electric flux through the surface S .

Strategy

Apply $\Phi = \int_S \vec{E} \cdot \hat{n} \, dA$, where the direction and magnitude of the electric field are constant.

Solution

The angle between the uniform electric field $\vec{\mathbf{E}}$ and the unit normal $\hat{\mathbf{n}}$ to the planar surface is 30° . Since both the direction and magnitude are constant, E comes outside the integral. All that is left is a surface integral over dA , which is A . Therefore, using the open-surface equation, we find that the electric flux through the surface is

Equation:

$$\begin{aligned}\Phi &= \int_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = EA \cos \theta \\ &= (10 \text{ N/C})(6.0 \text{ m}^2)(\cos 30^\circ) = 52 \text{ N} \cdot \text{m}^2/\text{C}.\end{aligned}$$

Significance

Again, the relative directions of the field and the area matter, and the general equation with the integral will simplify to the simple dot product of area and electric field.

Note:

Exercise:

Problem:

Check Your Understanding What angle should there be between the electric field and the surface shown in [\[link\]](#) in the previous example so that no electric flux passes through the surface?

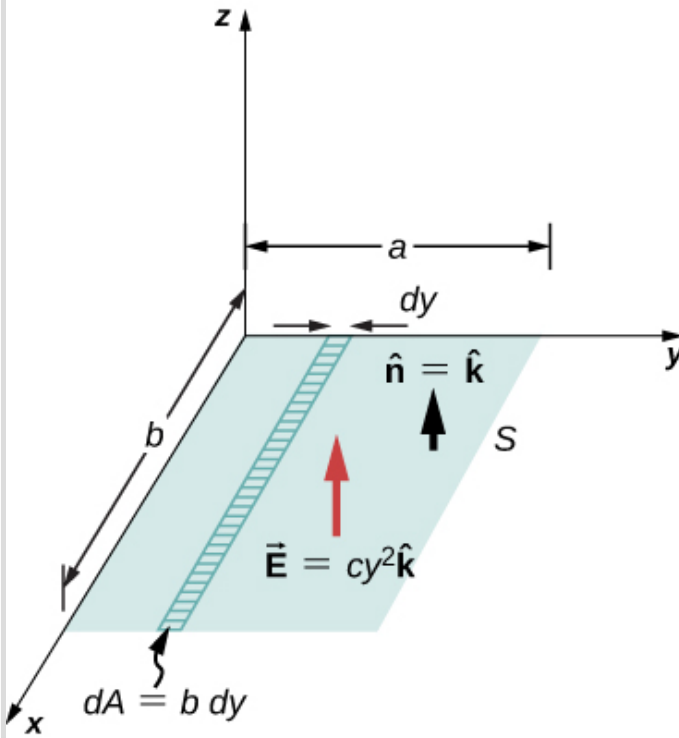
Solution:

Place it so that its unit normal is perpendicular to $\vec{\mathbf{E}}$.

Example:

Inhomogeneous Electric Field

What is the total flux of the electric field $\vec{\mathbf{E}} = cy^2\hat{\mathbf{k}}$ through the rectangular surface shown in [\[link\]](#)?



Since the electric field is not constant over the surface, an integration is necessary to determine the flux.

Strategy

Apply $\Phi = \int_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA$. We assume that the unit normal $\hat{\mathbf{n}}$ to the given surface points in the positive z -direction, so $\hat{\mathbf{n}} = \hat{\mathbf{k}}$. Since the electric field is not uniform over the surface, it is necessary to divide the surface into infinitesimal strips along which $\vec{\mathbf{E}}$ is essentially constant. As shown in [\[link\]](#), these strips are parallel to the x -axis, and each strip has an area $dA = b dy$.

Solution

From the open surface integral, we find that the net flux through the rectangular surface is

Equation:

$$\begin{aligned}\Phi &= \int_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} \, dA = \int_0^a (cy^2 \hat{\mathbf{k}}) \cdot \hat{\mathbf{k}} (b \, dy) \\ &= cb \int_0^a y^2 \, dy = \frac{1}{3} a^3 bc.\end{aligned}$$

Significance

For a non-constant electric field, the integral method is required.

Note:

Exercise:

Problem:

Check Your Understanding If the electric field in [\[link\]](#) is $\vec{\mathbf{E}} = mx\hat{\mathbf{k}}$, what is the flux through the rectangular area?

Solution:

$$mab^2/2$$

Summary

- The electric flux through a surface is proportional to the number of field lines crossing that surface. Note that this means the magnitude is proportional to the portion of the field perpendicular to the area.
- The electric flux is obtained by evaluating the surface integral

Equation:

$$\Phi = \oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} \, dA = \oint_S \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}},$$

where the notation used here is for a closed surface S .

Conceptual Questions

Exercise:

Problem:

Discuss how to orient a planar surface of area A in a uniform electric field of magnitude E_0 to obtain (a) the maximum flux and (b) the minimum flux through the area.

Solution:

a. If the planar surface is perpendicular to the electric field vector, the maximum flux would be obtained. b. If the planar surface were parallel to the electric field vector, the minimum flux would be obtained.

Exercise:**Problem:**

What are the maximum and minimum values of the flux in the preceding question?

Exercise:**Problem:**

The net electric flux crossing a closed surface is always zero. True or false?

Solution:

False. The net electric flux crossing a closed surface is always zero if and only if the net charge enclosed is zero.

Exercise:**Problem:**

The net electric flux crossing an open surface is never zero. True or false?

Problems**Exercise:**

Problem:

A uniform electric field of magnitude $1.1 \times 10^4 \text{ N/C}$ is perpendicular to a square sheet with sides 2.0 m long. What is the electric flux through the sheet?

Exercise:**Problem:**

Calculate the flux through the sheet of the previous problem if the plane of the sheet is at an angle of 60° to the field. Find the flux for both directions of the unit normal to the sheet.

Solution:

$\Phi = \vec{E} \cdot \vec{A} \rightarrow EA \cos \theta = 2.2 \times 10^4 \text{ N} \cdot \text{m}^2/\text{C}$ electric field in direction of unit normal; $\Phi = \vec{E} \cdot \vec{A} \rightarrow EA \cos \theta = -2.2 \times 10^4 \text{ N} \cdot \text{m}^2/\text{C}$ electric field opposite to unit normal

Exercise:**Problem:**

Find the electric flux through a rectangular area $3 \text{ cm} \times 2 \text{ cm}$ between two parallel plates where there is a constant electric field of 30 N/C for the following orientations of the area: (a) parallel to the plates, (b) perpendicular to the plates, and (c) the normal to the area making a 30° angle with the direction of the electric field. Note that this angle can also be given as $180^\circ + 30^\circ$.

Exercise:**Problem:**

The electric flux through a square-shaped area of side 5 cm near a large charged sheet is found to be $3 \times 10^{-5} \text{ N} \cdot \text{m}^2/\text{C}$ when the area is parallel to the plate. Find the charge density on the sheet.

Solution:

$$\frac{3 \times 10^{-5} \text{ N} \cdot \text{m}^2/\text{C}}{(0.05 \text{ m})^2} = E \Rightarrow \sigma = 2.12 \times 10^{-13} \text{ C/m}^2$$

Exercise:**Problem:**

Two large rectangular aluminum plates of area 150 cm^2 face each other with a separation of 3 mm between them. The plates are charged with equal amount of opposite charges, $\pm 20 \mu\text{C}$. The charges on the plates face each other. Find the flux through a circle of radius 3 cm between the plates when the normal to the circle makes an angle of 5° with a line perpendicular to the plates. Note that this angle can also be given as $180^\circ + 5^\circ$.

Exercise:**Problem:**

A square surface of area 2 cm^2 is in a space of uniform electric field of magnitude 10^3 N/C . The amount of flux through it depends on how the square is oriented relative to the direction of the electric field. Find the electric flux through the square, when the normal to it makes the following angles with electric field: (a) 30° , (b) 90° , and (c) 0° . Note that these angles can also be given as $180^\circ + \theta$.

Solution:

a. $\Phi = 0.17 \text{ N} \cdot \text{m}^2/\text{C}$;

b. $\Phi = 0$; c.

$$\Phi = EA \cos 0^\circ = 1.0 \times 10^3 \text{ N/C} (2.0 \times 10^{-4} \text{ m})^2 \cos 0^\circ = 0.20 \text{ N} \cdot \text{m}^2/\text{C}$$

Exercise:**Problem:**

A vector field is pointed along the z-axis, $\vec{v} = \frac{\alpha}{x^2+y^2} \hat{z}$. (a) Find the flux of the vector field through a rectangle in the xy-plane between $a < x < b$ and $c < y < d$. (b) Do the same through a rectangle in the yz-plane between $a < z < b$ and $c < y < d$. (Leave your answer as an integral.)

Exercise:**Problem:**

Consider the uniform electric field $\vec{E} = (4.0\hat{j} + 3.0\hat{k}) \times 10^3 \text{ N/C}$. What is its electric flux through a circular area of radius 2.0 m that lies in the xy-plane?

Solution:

$$\Phi = 3.8 \times 10^4 \text{ N} \cdot \text{m}^2/\text{C}$$

Exercise:**Problem:**

Repeat the previous problem, given that the circular area is (a) in the yz -plane and (b) 45° above the xy -plane.

Exercise:**Problem:**

An infinite charged wire with charge per unit length λ lies along the central axis of a cylindrical surface of radius r and length l . What is the flux through the surface due to the electric field of the charged wire?

Solution:

$$\vec{E}(z) = \frac{1}{4\pi\epsilon_0} \frac{2\lambda}{z} \hat{\mathbf{k}}, \quad \int \vec{E} \cdot \hat{\mathbf{n}} dA = \frac{\lambda}{\epsilon_0} l$$

Glossary

area vector

vector with magnitude equal to the area of a surface and direction perpendicular to the surface

electric flux

dot product of the electric field and the area through which it is passing

flux

quantity of something passing through a given area

Explaining Gauss's Law

By the end of this section, you will be able to:

- State Gauss's law
- Explain the conditions under which Gauss's law may be used
- Apply Gauss's law in appropriate systems

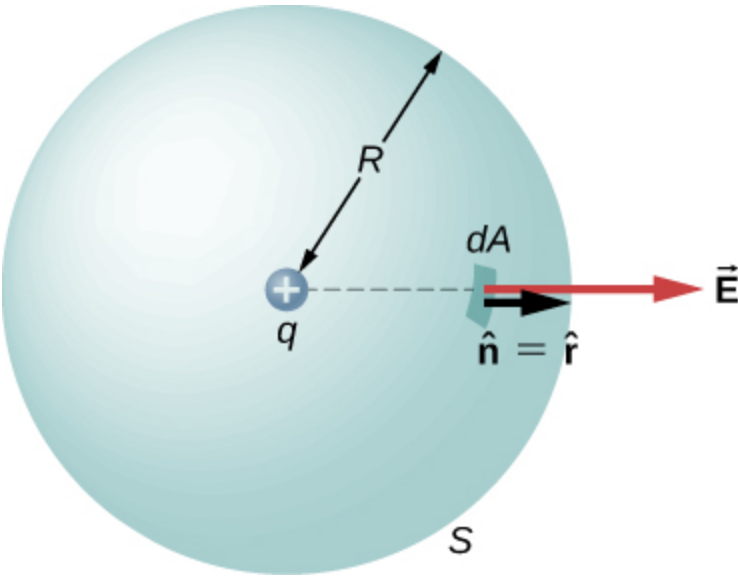
We can now determine the electric flux through an arbitrary closed surface due to an arbitrary charge distribution. We found that if a closed surface does not have any charge inside where an electric field line can terminate, then any electric field line entering the surface at one point must necessarily exit at some other point of the surface. Therefore, if a closed surface does not have any charges inside the enclosed volume, then the electric flux through the surface is zero. Now, what happens to the electric flux if there are some charges inside the enclosed volume? Gauss's law gives a quantitative answer to this question.

To get a feel for what to expect, let's calculate the electric flux through a spherical surface around a positive point charge q , since we already know the electric field in such a situation. Recall that when we place the point charge at the origin of a coordinate system, the electric field at a point P that is at a distance r from the charge at the origin is given by

Equation:

$$\vec{\mathbf{E}}_P = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \hat{\mathbf{r}},$$

where $\hat{\mathbf{r}}$ is the radial vector from the charge at the origin to the point P . We can use this electric field to find the flux through the spherical surface of radius r , as shown in [\[link\]](#).



A closed spherical surface surrounding a point charge q .

Then we apply $\Phi = \int_S \vec{E} \cdot \hat{n} dA$ to this system and substitute known values. On the sphere, $\hat{n} = \hat{r}$ and $r = R$, so for an infinitesimal area dA ,
Equation:

$$d\Phi = \vec{E} \cdot \hat{n} dA = \frac{1}{4\pi\epsilon_0} \frac{q}{R^2} \hat{r} \cdot \hat{r} dA = \frac{1}{4\pi\epsilon_0} \frac{q}{R^2} dA.$$

We now find the net flux by integrating this flux over the surface of the sphere:

Equation:

$$\Phi = \frac{1}{4\pi\epsilon_0} \frac{q}{R^2} \oint_S dA = \frac{1}{4\pi\epsilon_0} \frac{q}{R^2} (4\pi R^2) = \frac{q}{\epsilon_0}.$$

where the total surface area of the spherical surface is $4\pi R^2$. This gives the flux through the closed spherical surface at radius r as

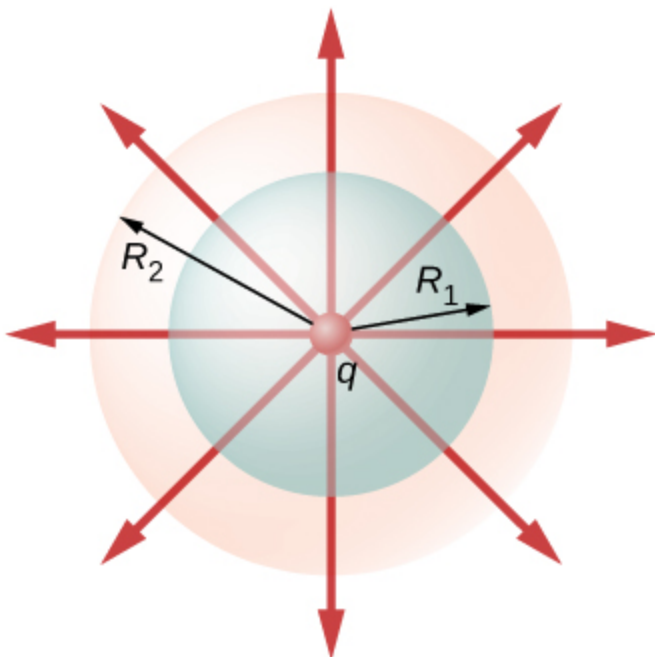
Equation:

$$\Phi = \frac{q}{\epsilon_0}.$$

A remarkable fact about this equation is that the flux is independent of the size of the spherical surface. This can be directly attributed to the fact that the electric field of a point charge decreases as $1/r^2$ with distance, which just cancels the r^2 rate of increase of the surface area.

Electric Field Lines Picture

An alternative way to see why the flux through a closed spherical surface is independent of the radius of the surface is to look at the electric field lines. Note that every field line from q that pierces the surface at radius R_1 also pierces the surface at R_2 ([link](#)).

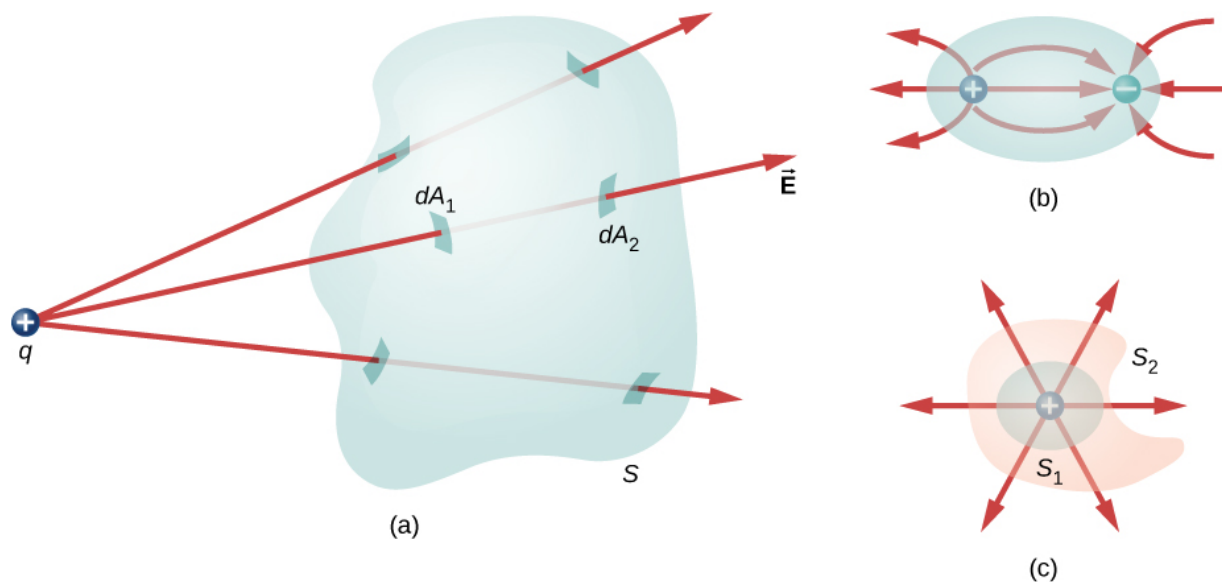


Flux through spherical surfaces of radii R_1 and R_2 enclosing a charge

q are equal, independent of the size of the surface, since all E -field lines that pierce one surface from the inside to outside direction also pierce the other surface in the same direction.

Therefore, the net number of electric field lines passing through the two surfaces from the inside to outside direction is equal. This net number of electric field lines, which is obtained by subtracting the number of lines in the direction from outside to inside from the number of lines in the direction from inside to outside gives a visual measure of the electric flux through the surfaces.

You can see that if no charges are included within a closed surface, then the electric flux through it must be zero. A typical field line enters the surface at dA_1 and leaves at dA_2 . Every line that enters the surface must also leave that surface. Hence the net “flow” of the field lines into or out of the surface is zero ([link](#)(a)). The same thing happens if charges of equal and opposite sign are included inside the closed surface, so that the total charge included is zero (part (b)). A surface that includes the same amount of charge has the same number of field lines crossing it, regardless of the shape or size of the surface, as long as the surface encloses the same amount of charge (part (c)).



Understanding the flux in terms of field lines. (a) The electric flux through a closed surface due to a charge outside that surface is zero. (b) Charges are enclosed, but because the net charge included is zero, the net flux through the closed surface is also zero. (c) The shape and size of the surfaces that enclose a charge does not matter because all surfaces enclosing the same charge have the same flux.

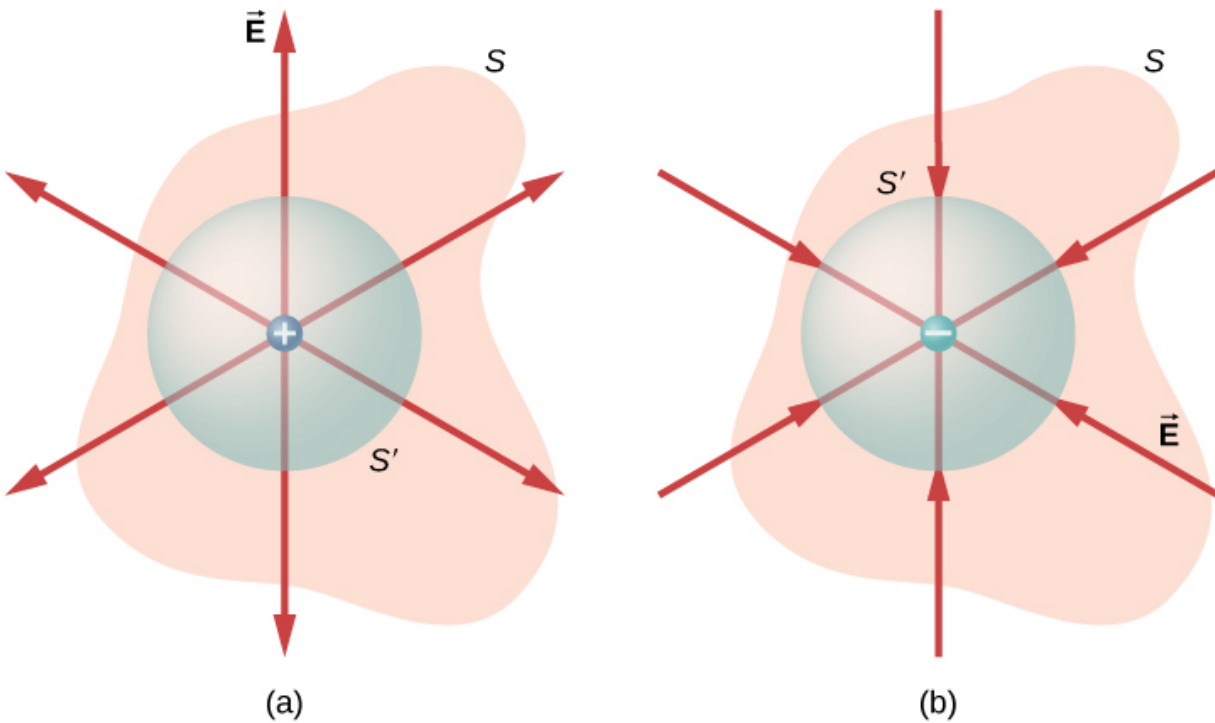
Statement of Gauss's Law

Gauss's law generalizes this result to the case of any number of charges and any location of the charges in the space inside the closed surface. According to Gauss's law, the flux of the electric field \vec{E} through any closed surface, also called a **Gaussian surface**, is equal to the net charge enclosed (q_{enc}) divided by the permittivity of free space (ϵ_0):

Equation:

$$\Phi_{\text{Closed Surface}} = \frac{q_{\text{enc}}}{\epsilon_0}.$$

This equation holds for *charges of either sign*, because we define the area vector of a closed surface to point outward. If the enclosed charge is negative (see [\[link\]](#)(b)), then the flux through either S or S' is negative.

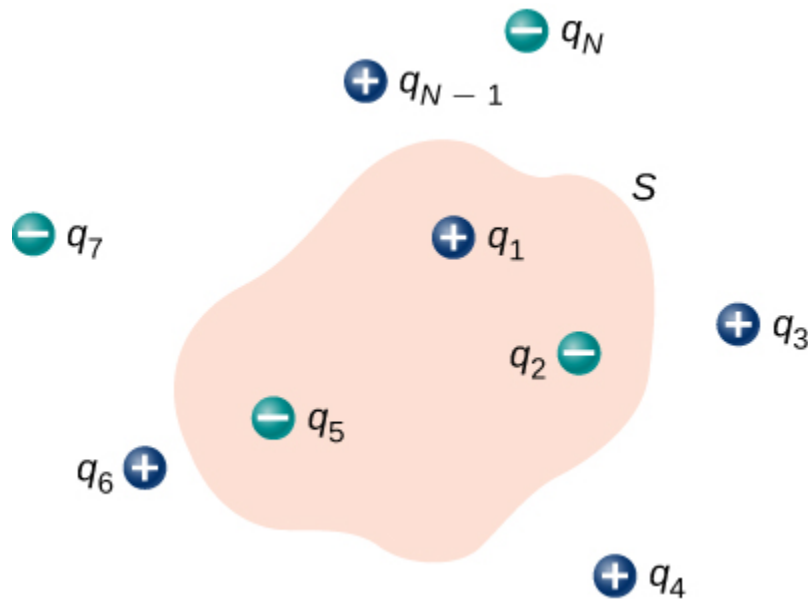


The electric flux through any closed surface surrounding a point charge q is given by Gauss's law. (a) Enclosed charge is positive. (b) Enclosed charge is negative.

The Gaussian surface does not need to correspond to a real, physical object; indeed, it rarely will. It is a mathematical construct that may be of any shape, provided that it is closed. However, since our goal is to integrate the flux over it, we tend to choose shapes that are highly symmetrical.

If the charges are discrete point charges, then we just add them. If the charge is described by a continuous distribution, then we need to integrate appropriately to find the total charge that resides inside the enclosed volume. For example, the flux through the Gaussian surface S of [\[link\]](#) is

$\Phi = (q_1 + q_2 + q_5)/\epsilon_0$. Note that q_{enc} is simply the sum of the point charges. If the charge distribution were continuous, we would need to integrate appropriately to compute the total charge within the Gaussian surface.



The flux through the Gaussian surface shown, due to the charge distribution, is

$$\Phi = |q_1| + |q_2| + |q_5|/\epsilon_0.$$

Recall that the principle of superposition holds for the electric field. Therefore, the total electric field at any point, including those on the chosen Gaussian surface, is the sum of all the electric fields present at this point. This allows us to write Gauss's law in terms of the total electric field.

Note:
Gauss's Law

The flux Φ of the electric field $\vec{\mathbf{E}}$ through any closed surface S (a Gaussian surface) is equal to the net charge enclosed (q_{enc}) divided by the permittivity of free space (ϵ_0) :

Equation:

$$\Phi = \oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = \frac{q_{\text{enc}}}{\epsilon_0}.$$

To use Gauss's law effectively, you must have a clear understanding of what each term in the equation represents. The field $\vec{\mathbf{E}}$ is the *total electric field* at every point on the Gaussian surface. This total field includes contributions from charges both inside and outside the Gaussian surface. However, q_{enc} is just the charge *inside* the Gaussian surface. Finally, the Gaussian surface is any closed surface in space. That surface can coincide with the actual surface of a conductor, or it can be an imaginary geometric surface. The only requirement imposed on a Gaussian surface is that it be closed ([\[link\]](#)).

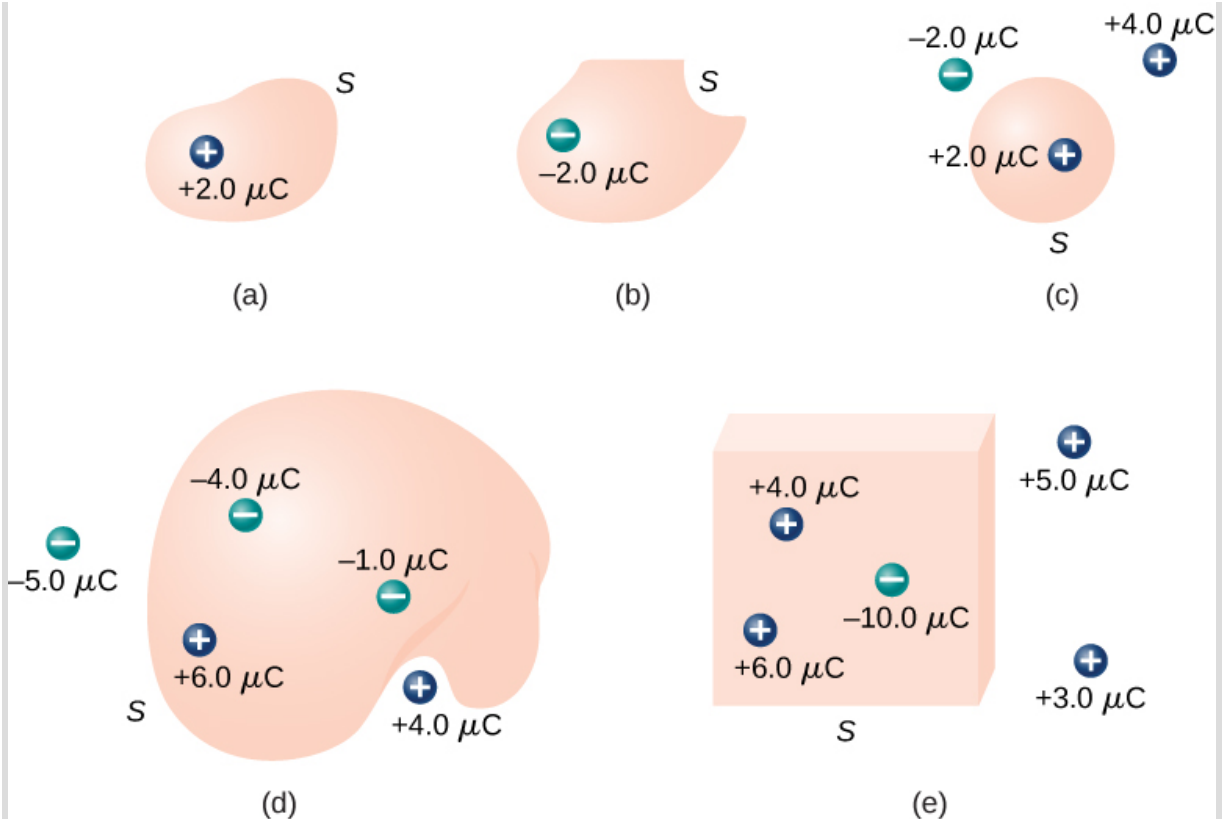


A Klein bottle partially filled with a liquid. Could the Klein bottle be used as a Gaussian surface?

Example:

Electric Flux through Gaussian Surfaces

Calculate the electric flux through each Gaussian surface shown in [\[link\]](#).



Various Gaussian surfaces and charges.

Strategy

From Gauss's law, the flux through each surface is given by $q_{\text{enc}}/\epsilon_0$, where q_{enc} is the charge enclosed by that surface.

Solution

For the surfaces and charges shown, we find

$$\text{a. } \Phi = \frac{2.0 \mu\text{C}}{\epsilon_0} = 2.3 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}.$$

$$\text{b. } \Phi = \frac{-2.0 \mu\text{C}}{\epsilon_0} = -2.3 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}.$$

$$\text{c. } \Phi = \frac{2.0 \mu\text{C}}{\epsilon_0} = 2.3 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}.$$

$$\text{d. } \Phi = \frac{-4.0 \mu\text{C} + 6.0 \mu\text{C} - 1.0 \mu\text{C}}{\epsilon_0} = 1.1 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}.$$

$$\text{e. } \Phi = \frac{4.0 \mu\text{C} + 6.0 \mu\text{C} - 10.0 \mu\text{C}}{\epsilon_0} = 0.$$

Significance

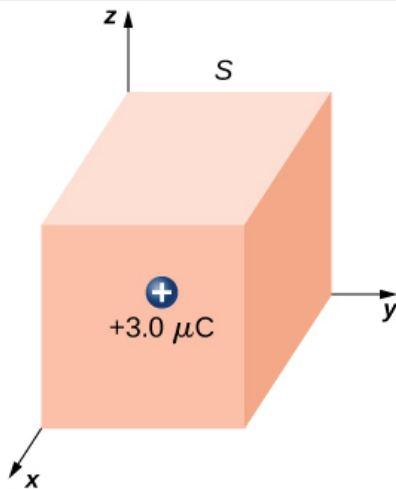
In the special case of a closed surface, the flux calculations become a sum of charges. In the next section, this will allow us to work with more complex systems.

Note:

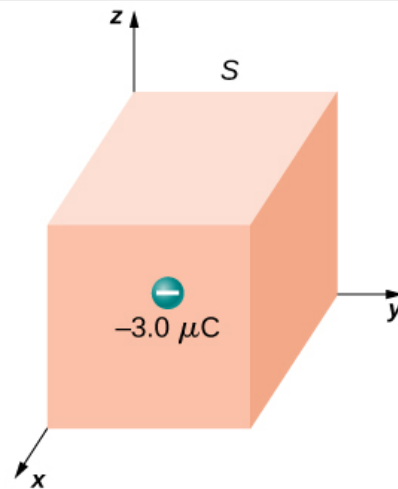
Exercise:

Problem:

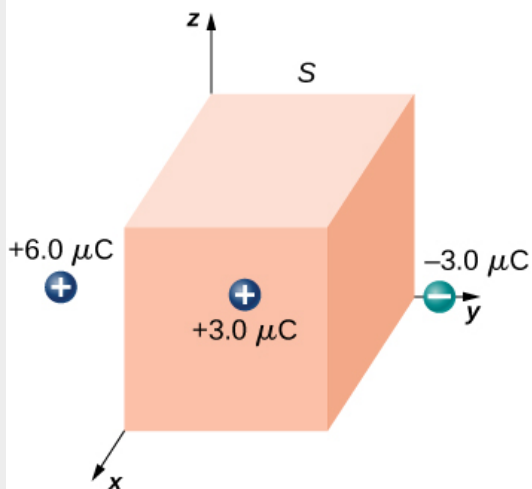
Check Your Understanding Calculate the electric flux through the closed cubical surface for each charge distribution shown in [\[link\]](#).



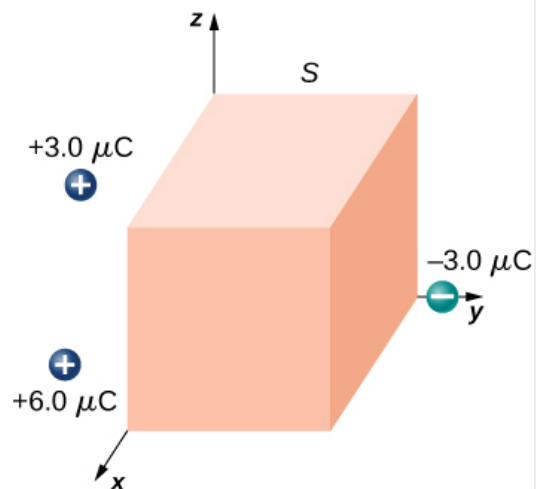
(a)



(b)



(c)



(d)

A cubical Gaussian surface with various charge distributions.

Solution:

a. $3.4 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}$; b. $-3.4 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}$; c.
 $3.4 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}$; d. 0

Note:

Use this [simulation](#) to adjust the magnitude of the charge and the radius of the Gaussian surface around it. See how this affects the total flux and the magnitude of the electric field at the Gaussian surface.

Summary

- Gauss's law relates the electric flux through a closed surface to the net charge within that surface,

Equation:

$$\Phi = \oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = \frac{q_{\text{enc}}}{\epsilon_0},$$

where q_{enc} is the total charge inside the Gaussian surface S .

- All surfaces that include the same amount of charge have the same number of field lines crossing it, regardless of the shape or size of the surface, as long as the surfaces enclose the same amount of charge.

Conceptual Questions

Exercise:

Problem:

Two concentric spherical surfaces enclose a point charge q . The radius of the outer sphere is twice that of the inner one. Compare the electric fluxes crossing the two surfaces.

Solution:

Since the electric field vector has a $\frac{1}{r^2}$ dependence, the fluxes are the same since $A = 4\pi r^2$.

Exercise:**Problem:**

Compare the electric flux through the surface of a cube of side length a that has a charge q at its center to the flux through a spherical surface of radius a with a charge q at its center.

Exercise:**Problem:**

(a) If the electric flux through a closed surface is zero, is the electric field necessarily zero at all points on the surface? (b) What is the net charge inside the surface?

Solution:

a. no; b. zero

Exercise:**Problem:**

Discuss how Gauss's law would be affected if the electric field of a point charge did not vary as $1/r^2$.

Exercise:

Problem:

Discuss the similarities and differences between the gravitational field of a point mass m and the electric field of a point charge q .

Solution:

Both fields vary as $\frac{1}{r^2}$. Because the gravitational constant is so much smaller than $\frac{1}{4\pi\epsilon_0}$, the gravitational field is orders of magnitude weaker than the electric field. Also, the gravitational flux through a closed surface is zero or positive; however, the electric flux is positive, negative, or zero, depending on the definition of flux for the given situation.

Exercise:**Problem:**

Discuss whether Gauss's law can be applied to other forces, and if so, which ones.

Exercise:**Problem:**

Is the term $\vec{\mathbf{E}}$ in Gauss's law the electric field produced by just the charge inside the Gaussian surface?

Solution:

No, it is produced by all charges both inside and outside the Gaussian surface.

Exercise:**Problem:**

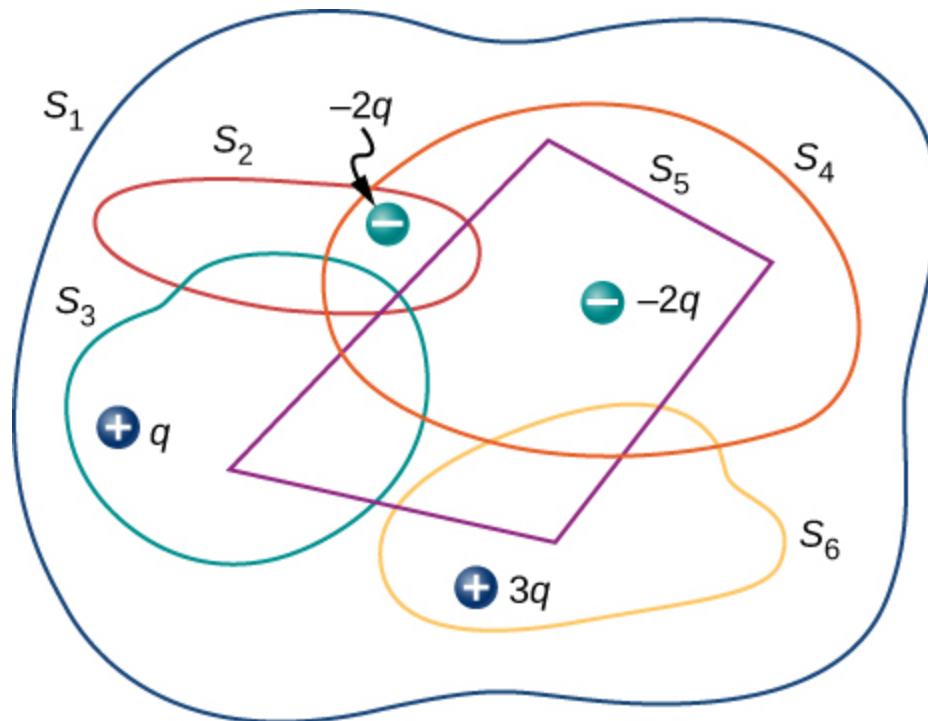
Reformulate Gauss's law by choosing the unit normal of the Gaussian surface to be the one directed inward.

Problems

Exercise:

Problem:

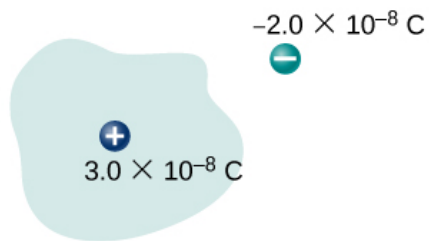
Determine the electric flux through each closed surface where the cross-section inside the surface is shown below.



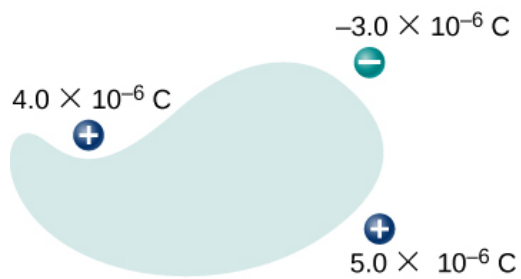
Exercise:

Problem:

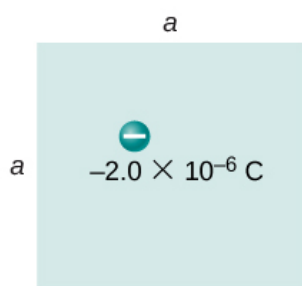
Find the electric flux through the closed surface whose cross-sections are shown below.



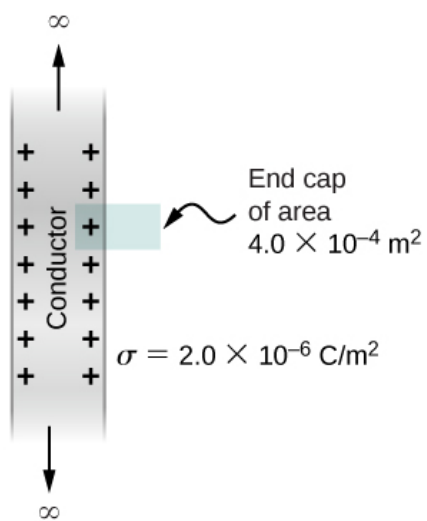
(a)



(b)



(c)



(d)

Solution:

- a. $\Phi = 3.39 \times 10^3 \text{ N} \cdot \text{m}^2/\text{C}$; b. $\Phi = 0$;
c. $\Phi = -2.25 \times 10^5 \text{ N} \cdot \text{m}^2/\text{C}$;
d. $\Phi = 90.4 \text{ N} \cdot \text{m}^2/\text{C}$

Exercise:**Problem:**

A point charge q is located at the center of a cube whose sides are of length a . If there are no other charges in this system, what is the electric flux through one face of the cube?

Exercise:**Problem:**

A point charge of $10 \mu\text{C}$ is at an unspecified location inside a cube of side 2 cm. Find the net electric flux through the surfaces of the cube.

Solution:

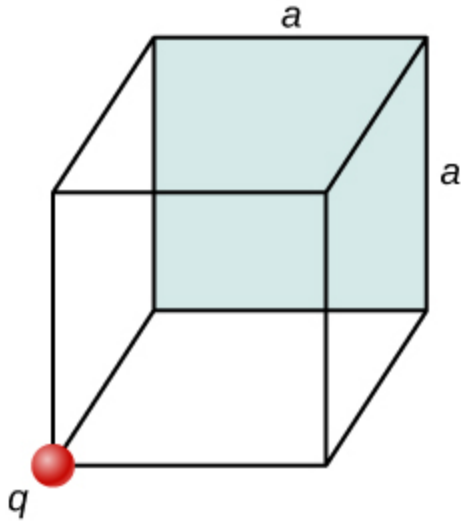
$$\Phi = 1.13 \times 10^6 \text{ N} \cdot \text{m}^2/\text{C}$$

Exercise:**Problem:**

A net flux of $1.0 \times 10^4 \text{ N} \cdot \text{m}^2/\text{C}$ passes inward through the surface of a sphere of radius 5 cm. (a) How much charge is inside the sphere? (b) How precisely can we determine the location of the charge from this information?

Exercise:**Problem:**

A charge q is placed at one of the corners of a cube of side a , as shown below. Find the magnitude of the electric flux through the shaded face due to q . Assume $q > 0$.



Solution:

Make a cube with q at the center, using the cube of side a . This would take four cubes of side a to make one side of the large cube. The shaded side of the small cube would be 1/24th of the total area of the large cube; therefore, the flux through the shaded area would be $\Phi = \frac{1}{24} \frac{q}{\epsilon_0}$.

Exercise:

Problem:

The electric flux through a cubical box 8.0 cm on a side is $1.2 \times 10^3 \text{ N} \cdot \text{m}^2/\text{C}$. What is the total charge enclosed by the box?

Exercise:

Problem:

The electric flux through a spherical surface is $4.0 \times 10^4 \text{ N} \cdot \text{m}^2/\text{C}$. What is the net charge enclosed by the surface?

Solution:

$$q = 3.54 \times 10^{-7} \text{ C}$$

Exercise:**Problem:**

A cube whose sides are of length d is placed in a uniform electric field of magnitude $E = 4.0 \times 10^3 \text{ N/C}$ so that the field is perpendicular to two opposite faces of the cube. What is the net flux through the cube?

Exercise:**Problem:**

Repeat the previous problem, assuming that the electric field is directed along a body diagonal of the cube.

Solution:

zero, also because flux in equals flux out

Exercise:**Problem:**

A total charge $5.0 \times 10^{-6} \text{ C}$ is distributed uniformly throughout a cubical volume whose edges are 8.0 cm long. (a) What is the charge density in the cube? (b) What is the electric flux through a cube with 12.0-cm edges that is concentric with the charge distribution? (c) Do the same calculation for cubes whose edges are 10.0 cm long and 5.0 cm long. (d) What is the electric flux through a spherical surface of radius 3.0 cm that is also concentric with the charge distribution?

Glossary

Gaussian surface

any enclosed (usually imaginary) surface

Applying Gauss's Law

By the end of this section, you will be able to:

- Explain what spherical, cylindrical, and planar symmetry are
- Recognize whether or not a given system possesses one of these symmetries
- Apply Gauss's law to determine the electric field of a system with one of these symmetries

Gauss's law is very helpful in determining expressions for the electric field, even though the law is not directly about the electric field; it is about the electric flux. It turns out that in situations that have certain symmetries (spherical, cylindrical, or planar) in the charge distribution, we can deduce the electric field based on knowledge of the electric flux. In these systems, we can find a

Gaussian surface S over which the electric field has constant magnitude. Furthermore, if \vec{E} is parallel to \hat{n} everywhere on the surface, then $\vec{E} \cdot \hat{n} = E$. (If \vec{E} and \hat{n} are antiparallel everywhere on the surface, then $\vec{E} \cdot \hat{n} = -E$.) Gauss's law then simplifies to

Note:

Equation:

$$\Phi = \oint_S \vec{E} \cdot \hat{n} dA = E \oint_S dA = EA = \frac{q_{\text{enc}}}{\epsilon_0},$$

where A is the area of the surface. Note that these symmetries lead to the transformation of the flux integral into a product of the magnitude of the electric field and an appropriate area. When you use this flux in the expression for Gauss's law, you obtain an algebraic equation that you can solve for the magnitude of the electric field, which looks like

Equation:

$$E \sim \frac{q_{\text{enc}}}{\epsilon_0 \text{ area}}.$$

The direction of the electric field at point P is obtained from the symmetry of the charge distribution and the type of charge in the distribution. Therefore, Gauss's law can be used to determine \vec{E} . Here is a summary of the steps we will follow:

Note:

Gauss's Law

1. *Identify the spatial symmetry of the charge distribution.* This is an important first step that allows us to choose the appropriate Gaussian surface. As examples, an isolated point charge has spherical symmetry, and an infinite line of charge has cylindrical symmetry.
2. *Choose a Gaussian surface with the same symmetry as the charge distribution and identify its consequences.* With this choice, $\vec{\mathbf{E}} \cdot \hat{\mathbf{n}}$ is easily determined over the Gaussian surface.
3. *Evaluate the integral $\oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA$ over the Gaussian surface, that is, calculate the flux through the surface.* The symmetry of the Gaussian surface allows us to factor $\vec{\mathbf{E}} \cdot \hat{\mathbf{n}}$ outside the integral.
4. *Determine the amount of charge enclosed by the Gaussian surface.* This is an evaluation of the right-hand side of the equation representing Gauss's law. It is often necessary to perform an integration to obtain the net enclosed charge.
5. *Evaluate the electric field of the charge distribution.* The field may now be found using the results of steps 3 and 4.

Basically, there are only three types of symmetry that allow Gauss's law to be used to deduce the electric field. They are

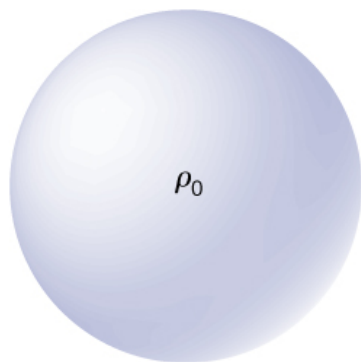
- A charge distribution with spherical symmetry
- A charge distribution with cylindrical symmetry
- A charge distribution with planar symmetry

To exploit the symmetry, we perform the calculations in appropriate coordinate systems and use the right kind of Gaussian surface for that symmetry, applying the remaining four steps.

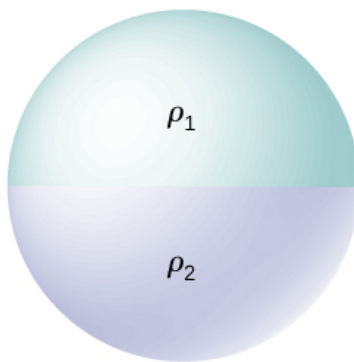
Charge Distribution with Spherical Symmetry

A charge distribution has **spherical symmetry** if the density of charge depends only on the distance from a point in space and not on the direction. In other words, if you rotate the system, it doesn't look different. For instance, if a sphere of radius R is uniformly charged with charge density ρ_0 then the distribution has spherical symmetry ([link](a)). On the other hand, if a sphere of radius R is charged so that the top half of the sphere has uniform charge density ρ_1 and the bottom half has a uniform charge density $\rho_2 \neq \rho_1$, then the sphere does not have spherical symmetry because the charge density depends on the direction ([link](b)). Thus, it is not the shape of the object but rather the shape of the charge distribution that determines whether or not a system has spherical symmetry.

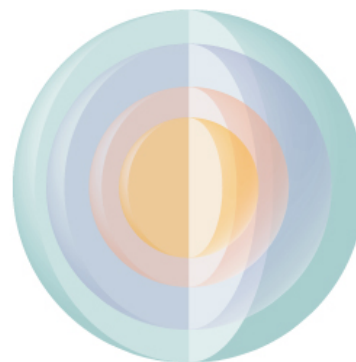
[link](c) shows a sphere with four different shells, each with its own uniform charge density. Although this is a situation where charge density in the full sphere is not uniform, the charge density function depends only on the distance from the center and not on the direction. Therefore, this charge distribution does have spherical symmetry.



(a) Spherically symmetric



(b) Not spherically symmetric



(c) Spherically symmetric

Illustrations of spherically symmetrical and nonsymmetrical systems. Different shadings indicate different charge densities. Charges on spherically shaped objects do not necessarily mean the charges are distributed with spherical symmetry. The spherical symmetry occurs only when the charge density does not depend on the direction. In (a), charges are distributed uniformly in a sphere. In (b), the upper half of the sphere has a different charge density from the lower half; therefore, (b) does not have spherical symmetry. In (c), the charges are in spherical shells of different charge densities, which means that charge density is only a function of the radial distance from the center; therefore, the system has spherical symmetry.

One good way to determine whether or not your problem has spherical symmetry is to look at the charge density function in spherical coordinates, $\rho(r, \theta, \phi)$. If the charge density is only a function of r , that is $\rho = \rho(r)$, then you have spherical symmetry. If the density depends on θ or ϕ , you could change it by rotation; hence, you would not have spherical symmetry.

Consequences of symmetry

In all spherically symmetrical cases, the electric field at any point must be radially directed, because the charge and, hence, the field must be invariant under rotation. Therefore, using spherical coordinates with their origins at the center of the spherical charge distribution, we can write down the expected form of the electric field at a point P located at a distance r from the center:

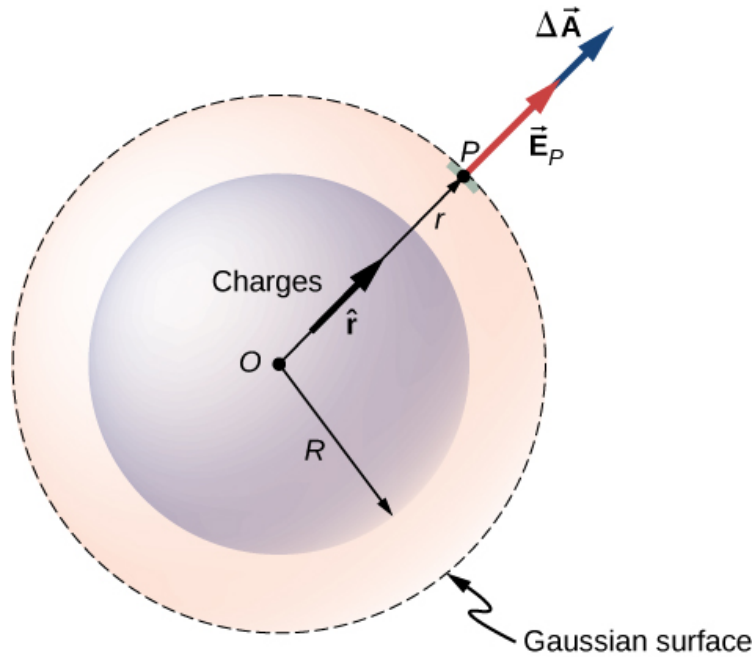
Equation:

$$\text{Spherical symmetry: } \vec{E}_P = E_P(r)\hat{r},$$

where \hat{r} is the unit vector pointed in the direction from the origin to the field point P . The radial component E_P of the electric field can be positive or negative. When $E_P > 0$, the electric field at P points away from the origin, and when $E_P < 0$, the electric field at P points toward the origin.

Gaussian surface and flux calculations

We can now use this form of the electric field to obtain the flux of the electric field through the Gaussian surface. For spherical symmetry, the Gaussian surface is a closed spherical surface that has the same center as the center of the charge distribution. Thus, the direction of the area vector of an area element on the Gaussian surface at any point is parallel to the direction of the electric field at that point, since they are both radially directed outward ([link](#)).



The electric field at any point of the spherical Gaussian surface for a spherically symmetrical charge distribution is parallel to the area element vector at that point, giving flux as the product of the magnitude of electric field and the value of the area. Note that the radius R of the charge distribution and the radius r of the Gaussian surface are different quantities.

The magnitude of the electric field \vec{E} must be the same everywhere on a spherical Gaussian surface concentric with the distribution. For a spherical surface of radius r ,

Equation:

$$\Phi = \oint_S \vec{E}_P \cdot \hat{n} dA = E_P \oint_S dA = E_P 4\pi r^2.$$

Using Gauss's law

According to Gauss's law, the flux through a closed surface is equal to the total charge enclosed within the closed surface divided by the permittivity of vacuum ε_0 . Let q_{enc} be the total charge enclosed inside the distance r from the origin, which is the space inside the Gaussian spherical surface of radius r . This gives the following relation for Gauss's law:

Equation:

$$4\pi r^2 E = \frac{q_{\text{enc}}}{\varepsilon_0}.$$

Hence, the electric field at point P that is a distance r from the center of a spherically symmetrical charge distribution has the following magnitude and direction:

Equation:

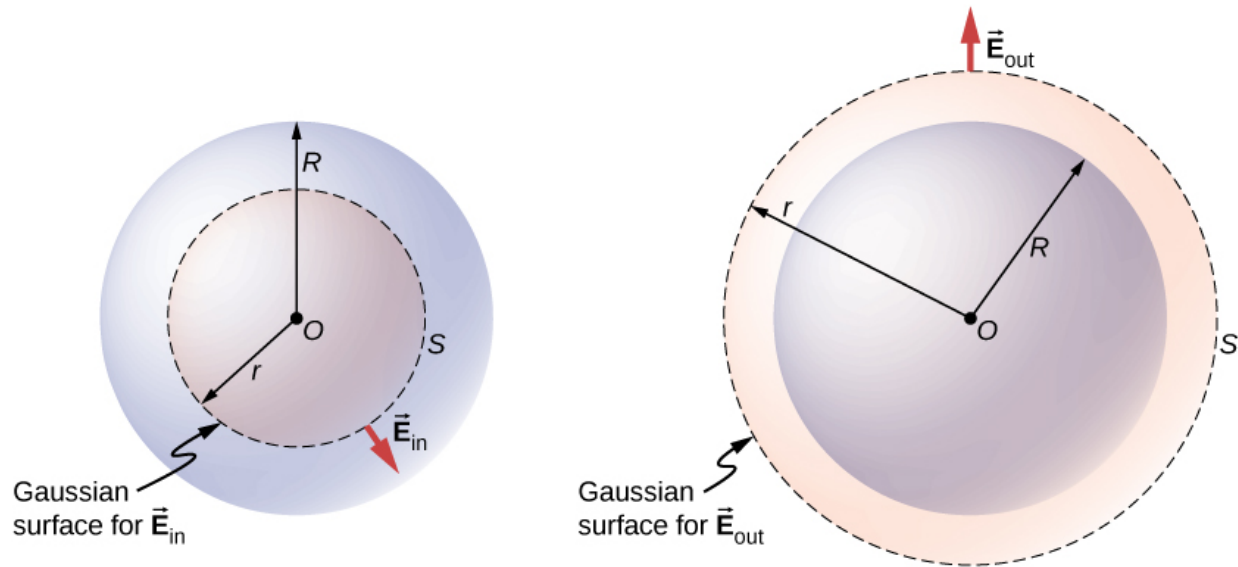
$$\text{Magnitude: } E(r) = \frac{1}{4\pi\varepsilon_0} \frac{q_{\text{enc}}}{r^2}$$

Direction: radial from O to P or from P to O .

The direction of the field at point P depends on whether the charge in the sphere is positive or negative. For a net positive charge enclosed within the Gaussian surface, the direction is from O to P , and for a net negative charge, the direction is from P to O . This is all we need for a point charge, and you will notice that the result above is identical to that for a point charge. However, Gauss's law becomes truly useful in cases where the charge occupies a finite volume.

Computing enclosed charge

The more interesting case is when a spherical charge distribution occupies a volume, and asking what the electric field inside the charge distribution is thus becomes relevant. In this case, the charge enclosed depends on the distance r of the field point relative to the radius of the charge distribution R , such as that shown in [\[link\]](#).



A spherically symmetrical charge distribution and the Gaussian surface used for finding the field (a) inside and (b) outside the distribution.

If point P is located outside the charge distribution—that is, if $r \geq R$ —then the Gaussian surface containing P encloses all charges in the sphere. In this case, q_{enc} equals the total charge in the sphere. On the other hand, if point P is within the spherical charge distribution, that is, if $r < R$, then the Gaussian surface encloses a smaller sphere than the sphere of charge distribution. In this case, q_{enc} is less than the total charge present in the sphere. Referring to [\[link\]](#), we can write q_{enc} as

Equation:

$$q_{\text{enc}} = \begin{cases} q_{\text{tot}} (\text{total charge}) & \text{if } r \geq R \\ q_{\text{within } r < R} (\text{only charge within } r < R) & \text{if } r < R \end{cases}$$

The field at a point outside the charge distribution is also called \vec{E}_{out} , and the field at a point inside the charge distribution is called \vec{E}_{in} . Focusing on the two types of field points, either inside or outside the charge distribution, we can now write the magnitude of the electric field as

Equation:

$$P \text{ outside sphere } E_{\text{out}} = \frac{1}{4\pi\epsilon_0} \frac{q_{\text{tot}}}{r^2}$$

Equation:

$$P \text{ inside sphere } E_{\text{in}} = \frac{1}{4\pi\epsilon_0} \frac{q_{\text{within } r < R}}{r^2}.$$

Note that the electric field outside a spherically symmetrical charge distribution is identical to that of a point charge at the center that has a charge equal to the total charge of the spherical charge distribution. This is remarkable since the charges are not located at the center only. We now work out specific examples of spherical charge distributions, starting with the case of a uniformly charged sphere.

Example:

Uniformly Charged Sphere

A sphere of radius R , such as that shown in [\[link\]](#), has a uniform volume charge density ρ_0 . Find the electric field at a point outside the sphere and at a point inside the sphere.

Strategy

Apply the Gauss's law problem-solving strategy, where we have already worked out the flux calculation.

Solution

The charge enclosed by the Gaussian surface is given by

Equation:

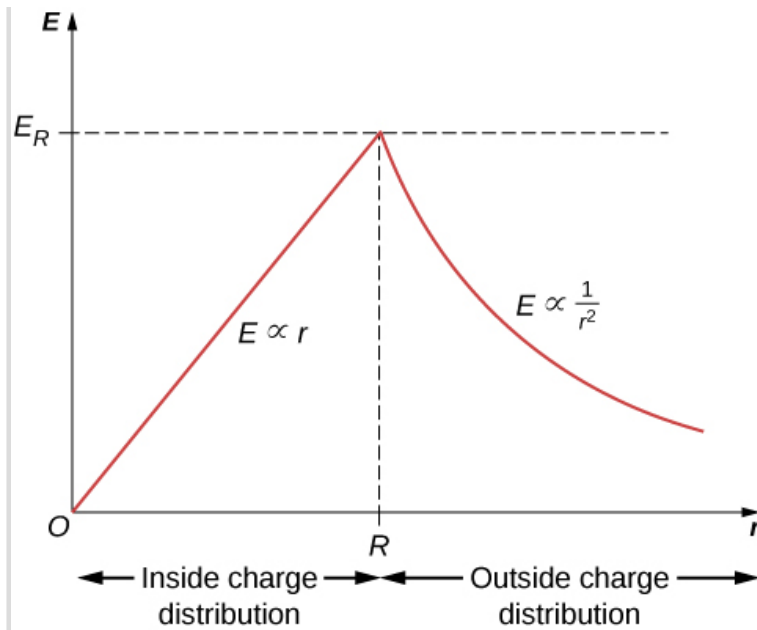
$$q_{\text{enc}} = \int_0^r \rho_0 dV = \int_0^r \rho_0 4\pi r'^2 dr' = \rho_0 \left(\frac{4}{3} \pi r^3 \right).$$

The answer for electric field amplitude can then be written down immediately for a point outside the sphere, labeled E_{out} , and a point inside the sphere, labeled E_{in} .

Equation:

$$\begin{aligned} E_{\text{out}} &= \frac{1}{4\pi\epsilon_0} \frac{q_{\text{tot}}}{r^2}, \quad q_{\text{tot}} = \frac{4}{3} \pi R^3 \rho_0, \\ E_{\text{in}} &= \frac{q_{\text{enc}}}{4\pi\epsilon_0 r^2} = \frac{\rho_0 r}{3\epsilon_0}, \quad \text{since } q_{\text{enc}} = \frac{4}{3} \pi r^3 \rho_0. \end{aligned}$$

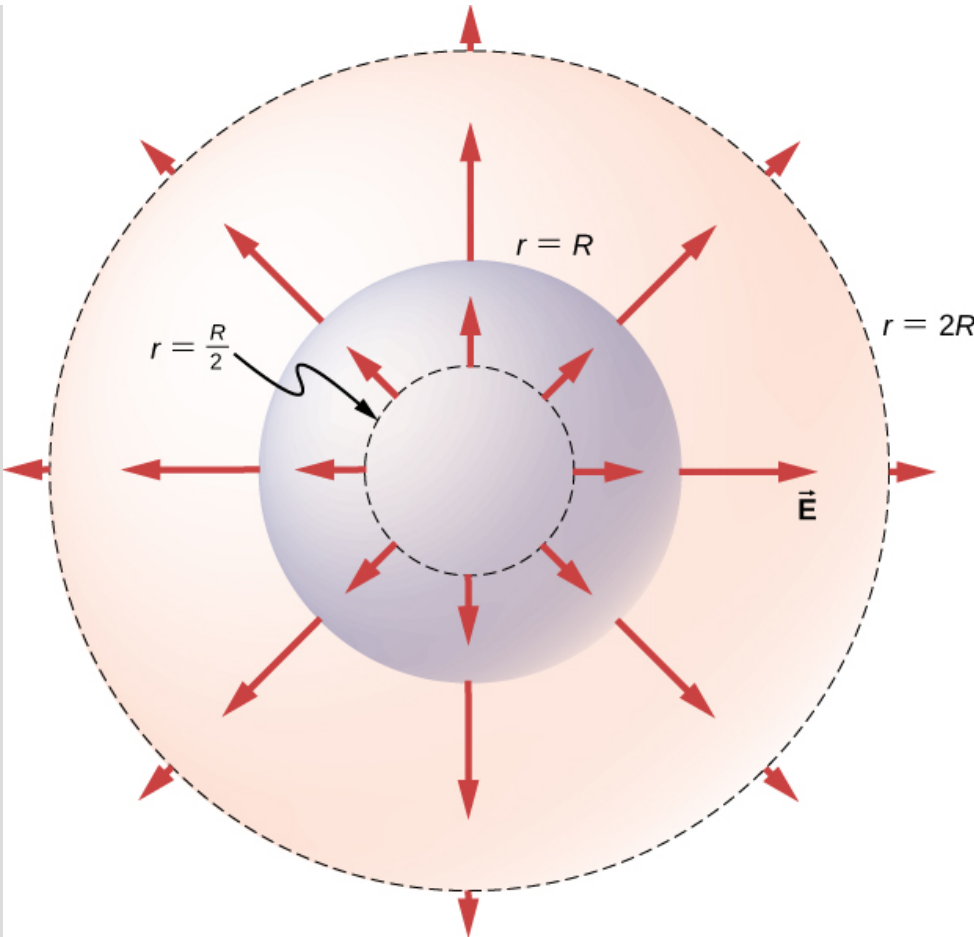
It is interesting to note that the magnitude of the electric field increases inside the material as you go out, since the amount of charge enclosed by the Gaussian surface increases with the volume. Specifically, the charge enclosed grows $\propto r^3$, whereas the field from each infinitesimal element of charge drops off $\propto 1/r^2$ with the net result that the electric field within the distribution increases in strength linearly with the radius. The magnitude of the electric field outside the sphere decreases as you go away from the charges, because the included charge remains the same but the distance increases. [\[link\]](#) displays the variation of the magnitude of the electric field with distance from the center of a uniformly charged sphere.



Electric field of a uniformly charged, non-conducting sphere increases inside the sphere to a maximum at the surface and then decreases as $1/r^2$. Here,

$E_R = \frac{\rho_0 R}{3\epsilon_0}$. The electric field is due to a spherical charge distribution of uniform charge density and total charge Q as a function of distance from the center of the distribution.

The direction of the electric field at any point P is radially outward from the origin if ρ_0 is positive, and inward (i.e., toward the center) if ρ_0 is negative. The electric field at some representative space points are displayed in [\[link\]](#) whose radial coordinates r are $r = R/2$, $r = R$, and $r = 2R$.



Electric field vectors inside and outside a uniformly charged sphere.

Significance

Notice that E_{out} has the same form as the equation of the electric field of an isolated point charge. In determining the electric field of a uniform spherical charge distribution, we can therefore assume that all of the charge inside the appropriate spherical Gaussian surface is located at the center of the distribution.

Example:

Non-Uniformly Charged Sphere

A non-conducting sphere of radius R has a non-uniform charge density that varies with the distance from its center as given by

Equation:

$$\rho(r) = ar^n \quad (r \leq R; n \geq 0),$$

where a is a constant. We require $n \geq 0$ so that the charge density is not undefined at $r = 0$. Find the electric field at a point outside the sphere and at a point inside the sphere.

Strategy

Apply the Gauss's law strategy given above, where we work out the enclosed charge integrals separately for cases inside and outside the sphere.

Solution

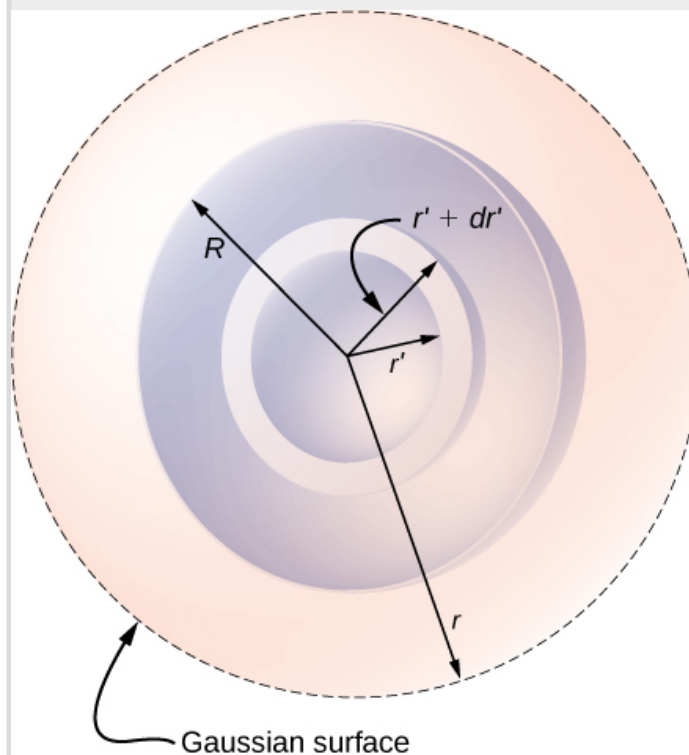
Since the given charge density function has only a radial dependence and no dependence on direction, we have a spherically symmetrical situation. Therefore, the magnitude of the electric field at any point is given above and the direction is radial. We just need to find the enclosed charge q_{enc} , which depends on the location of the field point.

A note about symbols: We use r' for locating charges in the charge distribution and r for locating the field point(s) at the Gaussian surface(s). The letter R is used for the radius of the charge distribution.

As charge density is not constant here, we need to integrate the charge density function over the volume enclosed by the Gaussian surface. Therefore, we set up the problem for charges in one spherical shell, say between r' and $r' + dr'$, as shown in [\[link\]](#). The volume of charges in the shell of infinitesimal width is equal to the product of the area of surface $4\pi r'^2$ and the thickness dr' . Multiplying the volume with the density at this location, which is ar'^n , gives the charge in the shell:

Equation:

$$dq = ar'^n 4\pi r'^2 dr'.$$



Spherical symmetry with non-uniform charge distribution. In this type of problem, we need four radii: R is the radius of the charge distribution, r is the radius of the Gaussian surface, r' is the inner radius of the spherical

shell, and $r' + dr'$ is the outer radius of the spherical shell. The spherical shell is used to calculate the charge enclosed within the Gaussian surface. The range for r' is from 0 to r for the field at a point inside the charge distribution and from 0 to R for the field at a point outside the charge distribution. If $r > R$, then the Gaussian surface encloses more volume than the charge distribution, but the additional volume does not contribute to q_{enc} .

(a) **Field at a point outside the charge distribution.** In this case, the Gaussian surface, which contains the field point P , has a radius r that is greater than the radius R of the charge distribution, $r > R$. Therefore, all charges of the charge distribution are enclosed within the Gaussian surface. Note that the space between $r' = R$ and $r' = r$ is empty of charges and therefore does not contribute to the integral over the volume enclosed by the Gaussian surface:

Equation:

$$q_{\text{enc}} = \int dq = \int_0^R ar'^n 4\pi r'^2 dr' = \frac{4\pi a}{n+3} R^{n+3}.$$

This is used in the general result for $\vec{\mathbf{E}}_{\text{out}}$ above to obtain the electric field at a point outside the charge distribution as

Equation:

$$\vec{\mathbf{E}}_{\text{out}} = \left[\frac{aR^{n+3}}{\epsilon_0(n+3)} \right] \frac{1}{r^2} \hat{\mathbf{r}},$$

where $\hat{\mathbf{r}}$ is a unit vector in the direction from the origin to the field point at the Gaussian surface.

(b) **Field at a point inside the charge distribution.** The Gaussian surface is now buried inside the charge distribution, with $r < R$. Therefore, only those charges in the distribution that are within a distance r of the center of the spherical charge distribution count in q_{enc} :

Equation:

$$q_{\text{enc}} = \int_0^r ar'^n 4\pi r'^2 dr' = \frac{4\pi a}{n+3} r^{n+3}.$$

Now, using the general result above for $\vec{\mathbf{E}}_{\text{in}}$, we find the electric field at a point that is a distance r from the center and lies within the charge distribution as

Equation:

$$\vec{\mathbf{E}}_{\text{in}} = \left[\frac{a}{\epsilon_0(n+3)} \right] r^{n+1} \hat{\mathbf{r}},$$

where the direction information is included by using the unit radial vector.

Note:

Exercise:

Problem:

Check Your Understanding Check that the electric fields for the sphere reduce to the correct values for a point charge.

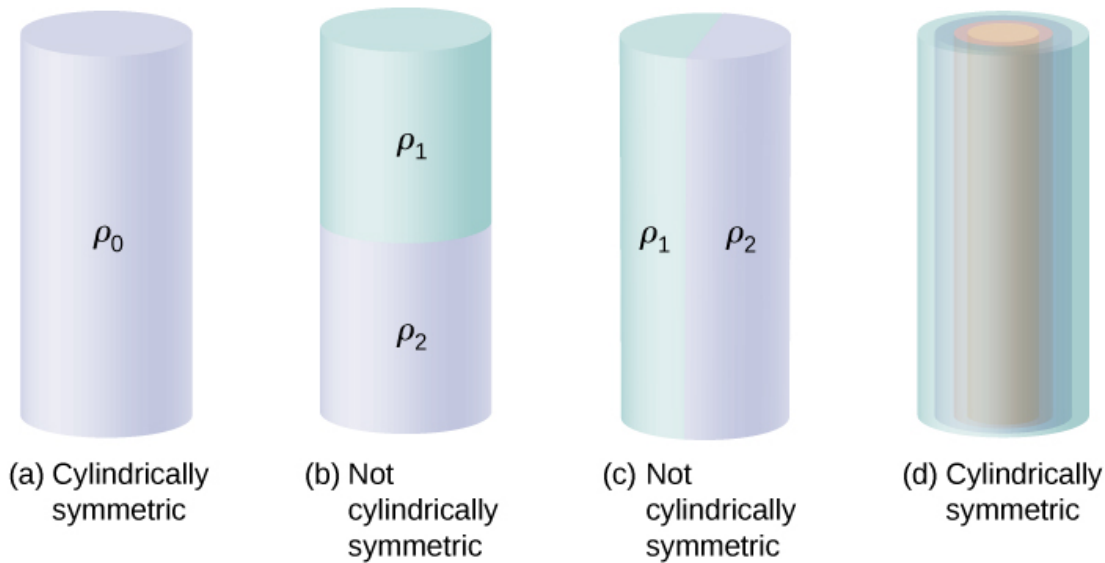
Solution:

In this case, there is only \vec{E}_{out} . So, yes.

Charge Distribution with Cylindrical Symmetry

A charge distribution has **cylindrical symmetry** if the charge density depends only upon the distance r from the axis of a cylinder and must not vary along the axis or with direction about the axis. In other words, if your system varies if you rotate it around the axis, or shift it along the axis, you do not have cylindrical symmetry.

[\[link\]](#) shows four situations in which charges are distributed in a cylinder. A uniform charge density ρ_0 in an infinite straight wire has a cylindrical symmetry, and so does an infinitely long cylinder with constant charge density ρ_0 . An infinitely long cylinder that has different charge densities along its length, such as a charge density ρ_1 for $z > 0$ and $\rho_2 \neq \rho_1$ for $z < 0$, does not have a usable cylindrical symmetry for this course. Neither does a cylinder in which charge density varies with the direction, such as a charge density ρ_1 for $0 \leq \theta < \pi$ and $\rho_2 \neq \rho_1$ for $\pi \leq \theta < 2\pi$. A system with concentric cylindrical shells, each with uniform charge densities, albeit different in different shells, as in [\[link\]\(d\)](#), does have cylindrical symmetry if they are infinitely long. The infinite length requirement is due to the charge density changing along the axis of a finite cylinder. In real systems, we don't have infinite cylinders; however, if the cylindrical object is considerably longer than the radius from it that we are interested in, then the approximation of an infinite cylinder becomes useful.



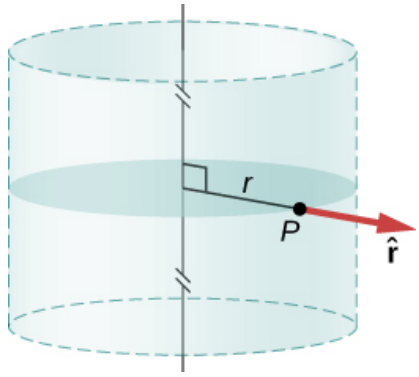
To determine whether a given charge distribution has cylindrical symmetry, look at the cross-section of an “infinitely long” cylinder. If the charge density does not depend on the polar angle of the cross-section or along the axis, then you have cylindrical symmetry. (a) Charge density is constant in the cylinder; (b) upper half of the cylinder has a different charge density from the lower half; (c) left half of the cylinder has a different charge density from the right half; (d) charges are constant in different cylindrical rings, but the density does not depend on the polar angle. Cases (a) and (d) have cylindrical symmetry, whereas (b) and (c) do not.

Consequences of symmetry

In all cylindrically symmetrical cases, the electric field \vec{E}_P at any point P must also display cylindrical symmetry.

Cylindrical symmetry: $\vec{E}_P = E_P(r)\hat{r}$,

where r is the distance from the axis and \hat{r} is a unit vector directed perpendicularly away from the axis ([link](#)).

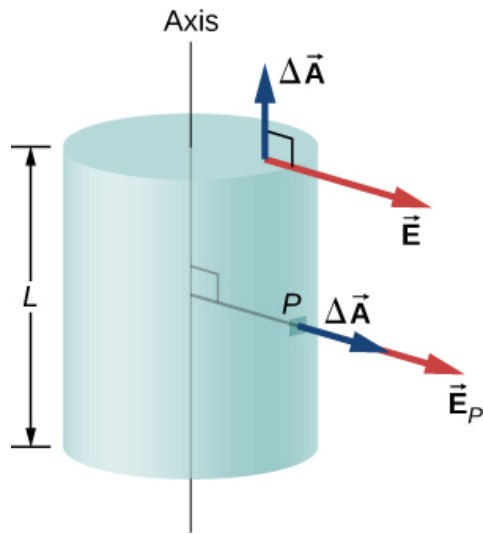


The electric field in a cylindrically symmetrical situation depends only on the distance from the axis.

The direction of the electric field is pointed away from the axis for positive charges and toward the axis for negative charges.

Gaussian surface and flux calculation

To make use of the direction and functional dependence of the electric field, we choose a closed Gaussian surface in the shape of a cylinder with the same axis as the axis of the charge distribution. The flux through this surface of radius s and height L is easy to compute if we divide our task into two parts: (a) a flux through the flat ends and (b) a flux through the curved surface ([link](#)).



The Gaussian surface in the case of cylindrical symmetry. The electric field at a patch is either parallel or perpendicular to the normal to the patch of the Gaussian surface.

The electric field is perpendicular to the cylindrical side and parallel to the planar end caps of the surface. The flux through the cylindrical part is

Equation:

$$\int_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = E \int_S dA = E(2\pi rL),$$

whereas the flux through the end caps is zero because $\vec{\mathbf{E}} \cdot \hat{\mathbf{n}} = 0$ there. Thus, the flux is

Equation:

$$\int_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = E(2\pi rL) + 0 + 0 = 2\pi rLE.$$

Using Gauss's law

According to Gauss's law, the flux must equal the amount of charge within the volume enclosed by this surface, divided by the permittivity of free space. When you do the calculation for a cylinder of length L , you find that q_{enc} of Gauss's law is directly proportional to L . Let us write it as charge per unit length (λ_{enc}) times length L :

Equation:

$$q_{\text{enc}} = \lambda_{\text{enc}} L.$$

Hence, Gauss's law for any cylindrically symmetrical charge distribution yields the following magnitude of the electric field a distance s away from the axis:

Equation:

$$\text{Magnitude: } E(r) = \frac{\lambda_{\text{enc}}}{2\pi\epsilon_0} \frac{1}{r}.$$

The charge per unit length λ_{enc} depends on whether the field point is inside or outside the cylinder of charge distribution, just as we have seen for the spherical distribution.

Computing enclosed charge

Let R be the radius of the cylinder within which charges are distributed in a cylindrically symmetrical way. Let the field point P be at a distance s from the axis. (The side of the Gaussian surface includes the field point P .) When $r > R$ (that is, when P is outside the charge distribution), the Gaussian surface includes all the charge in the cylinder of radius R and length L . When $r < R$ (P is located inside the charge distribution), then only the charge within a cylinder of radius s and length L is enclosed by the Gaussian surface:

Equation:

$$\lambda_{\text{enc}} L = \begin{cases} (\text{total charge}) & \text{if } r \geq R \\ (\text{only charge within } r < R) & \text{if } r < R \end{cases}$$

Example:**Uniformly Charged Cylindrical Shell**

A very long non-conducting cylindrical shell of radius R has a uniform surface charge density σ_0 . Find the electric field (a) at a point outside the shell and (b) at a point inside the shell.

Strategy

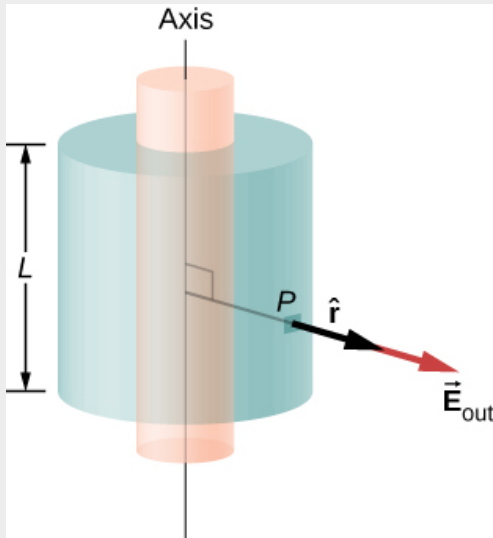
Apply the Gauss's law strategy given earlier, where we treat the cases inside and outside the shell separately.

Solution

- a. **Electric field at a point outside the shell.** For a point outside the cylindrical shell, the Gaussian surface is the surface of a cylinder of radius $r > R$ and length L , as shown in [\[link\]](#). The charge enclosed by the Gaussian cylinder is equal to the charge on the cylindrical shell of length L . Therefore, λ_{enc} is given by

Equation:

$$\lambda_{\text{enc}} = \frac{\sigma_0 2\pi R L}{L} = 2\pi R \sigma_0.$$



A Gaussian surface surrounding a cylindrical shell.

Hence, the electric field at a point P outside the shell at a distance r away from the axis is
Equation:

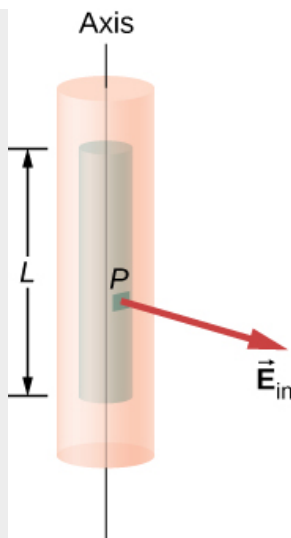
$$\vec{\mathbf{E}} = \frac{2\pi R \sigma_0}{2\pi \epsilon_0} \frac{1}{r} \hat{\mathbf{r}} = \frac{R \sigma_0}{\epsilon_0} \frac{1}{r} \hat{\mathbf{r}} \quad (r > R)$$

where $\hat{\mathbf{r}}$ is a unit vector, perpendicular to the axis and pointing away from it, as shown in the figure. The electric field at P points in the direction of $\hat{\mathbf{r}}$ given in [\[link\]](#) if $\sigma_0 > 0$ and in the opposite direction to $\hat{\mathbf{r}}$ if $\sigma_0 < 0$.

- b. **Electric field at a point inside the shell.** For a point inside the cylindrical shell, the Gaussian surface is a cylinder whose radius r is less than R ([\[link\]](#)). This means no charges are included inside the Gaussian surface:

Equation:

$$\lambda_{\text{enc}} = 0.$$



A Gaussian surface within a cylindrical shell.

This gives the following equation for the magnitude of the electric field E_{in} at a point whose r is less than R of the shell of charges.

Equation:

$$E_{in} 2\pi r L = 0 \quad (r < R),$$

This gives us

Equation:

$$E_{in} = 0 \quad (r < R).$$

Significance

Notice that the result inside the shell is exactly what we should expect: No enclosed charge means zero electric field. Outside the shell, the result becomes identical to a wire with uniform charge $R\sigma_0$.

Note:

Exercise:

Problem:

Check Your Understanding A thin straight wire has a uniform linear charge density λ_0 . Find the electric field at a distance d from the wire, where d is much less than the length of the wire.

Solution:

$\vec{\mathbf{E}} = \frac{\lambda_0}{2\pi\epsilon_0} \frac{1}{d} \hat{\mathbf{r}}$; This agrees with the calculation of [\[link\]](#) where we found the electric field by integrating over the charged wire. Notice how much simpler the calculation of this electric field is with Gauss's law.

Charge Distribution with Planar Symmetry

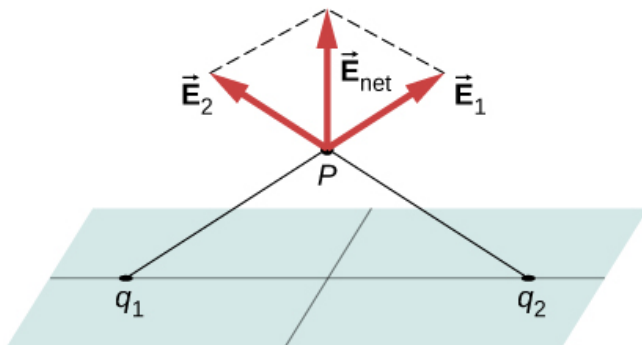
A **planar symmetry** of charge density is obtained when charges are uniformly spread over a large flat surface. In planar symmetry, all points in a plane parallel to the plane of charge are identical with respect to the charges.

Consequences of symmetry

We take the plane of the charge distribution to be the xy -plane and we find the electric field at a space point P with coordinates (x, y, z) . Since the charge density is the same at all (x, y) -coordinates in the $z = 0$ plane, by symmetry, the electric field at P cannot depend on the x - or y -coordinates of point P , as shown in [\[link\]](#). Therefore, the electric field at P can only depend on the distance from the plane and has a direction either toward the plane or away from the plane. That is, the electric field at P has only a nonzero z -component.

Uniform charges in xy plane: $\vec{\mathbf{E}} = E(z)\hat{\mathbf{z}}$

where z is the distance from the plane and $\hat{\mathbf{z}}$ is the unit vector normal to the plane. Note that in this system, $E(z) = E(-z)$, although of course they point in opposite directions.

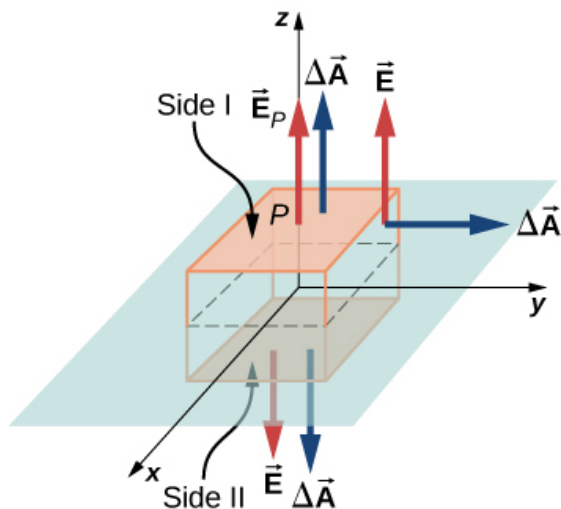


The components of the electric field parallel to a plane of charges cancel out the two charges located symmetrically from the field point P . Therefore, the field at any point is

pointed vertically from the plane of charges.
 For any point P and charge q_1 , we can always
 find a q_2 with this effect.

Gaussian surface and flux calculation

In the present case, a convenient Gaussian surface is a box, since the expected electric field points in one direction only. To keep the Gaussian box symmetrical about the plane of charges, we take it to straddle the plane of the charges, such that one face containing the field point P is taken parallel to the plane of the charges. In [\[link\]](#), sides I and II of the Gaussian surface (the box) that are parallel to the infinite plane have been shaded. They are the only surfaces that give rise to nonzero flux because the electric field and the area vectors of the other faces are perpendicular to each other.



A thin charged sheet and the Gaussian box for finding the electric field at the field point P . The normal to each face of the box is from inside the box to outside. On two faces of the box, the electric fields are parallel to the area vectors, and on the other four faces, the electric fields are perpendicular to the area vectors.

Let A be the area of the shaded surface on each side of the plane and E_P be the magnitude of the electric field at point P . Since sides I and II are at the same distance from the plane, the electric field has the same magnitude at points in these planes, although the directions of the electric field at these points in the two planes are opposite to each other.

Magnitude at I or II: $E(z) = E_P$.

If the charge on the plane is positive, then the direction of the electric field and the area vectors are as shown in [\[link\]](#). Therefore, we find for the flux of electric field through the box

Equation:

$$\Phi = \oint_S \vec{\mathbf{E}}_P \cdot \hat{\mathbf{n}} dA = E_P A + E_P A + 0 + 0 + 0 + 0 = 2E_P A$$

where the zeros are for the flux through the other sides of the box. Note that if the charge on the plane is negative, the directions of electric field and area vectors for planes I and II are opposite to each other, and we get a negative sign for the flux. According to Gauss's law, the flux must equal $q_{\text{enc}}/\epsilon_0$. From [\[link\]](#), we see that the charges inside the volume enclosed by the Gaussian box reside on an area A of the xy -plane. Hence,

Equation:

$$q_{\text{enc}} = \sigma_0 A.$$

Using the equations for the flux and enclosed charge in Gauss's law, we can immediately determine the electric field at a point at height z from a uniformly charged plane in the xy -plane:

Equation:

$$\vec{\mathbf{E}}_P = \frac{\sigma_0}{2\epsilon_0} \hat{\mathbf{n}}.$$

The direction of the field depends on the sign of the charge on the plane and the side of the plane where the field point P is located. Note that above the plane, $\hat{\mathbf{n}} = +\hat{\mathbf{z}}$, while below the plane, $\hat{\mathbf{n}} = -\hat{\mathbf{z}}$.

You may be surprised to note that the electric field does not actually depend on the distance from the plane; this is an effect of the assumption that the plane is infinite. In practical terms, the result given above is still a useful approximation for finite planes near the center.

Summary

- For a charge distribution with certain spatial symmetries (spherical, cylindrical, and planar), we can find a Gaussian surface over which $\vec{\mathbf{E}} \cdot \hat{\mathbf{n}} = E$, where E is constant over the surface. The electric field is then determined with Gauss's law.
- For spherical symmetry, the Gaussian surface is also a sphere, and Gauss's law simplifies to $4\pi r^2 E = \frac{q_{\text{enc}}}{\epsilon_0}$.

- For cylindrical symmetry, we use a cylindrical Gaussian surface, and find that Gauss's law simplifies to $2\pi rLE = \frac{q_{\text{enc}}}{\epsilon_0}$.
- For planar symmetry, a convenient Gaussian surface is a box penetrating the plane, with two faces parallel to the plane and the remainder perpendicular, resulting in Gauss's law being $2AE = \frac{q_{\text{enc}}}{\epsilon_0}$.

Conceptual Questions

Exercise:

Problem:

Would Gauss's law be helpful for determining the electric field of two equal but opposite charges a fixed distance apart?

Solution:

No, since the situation does not have symmetry, making Gauss's law challenging to simplify.

Exercise:

Problem:

Discuss the role that symmetry plays in the application of Gauss's law. Give examples of continuous charge distributions in which Gauss's law is useful and not useful in determining the electric field.

Exercise:

Problem:

Discuss the restrictions on the Gaussian surface used to discuss planar symmetry. For example, is its length important? Does the cross-section have to be square? Must the end faces be on opposite sides of the sheet?

Solution:

Any shape of the Gaussian surface can be used. The only restriction is that the Gaussian integral must be calculable; therefore, a box or a cylinder are the most convenient geometrical shapes for the Gaussian surface.

Problems

Exercise:

Problem:

Recall that in the example of a uniform charged sphere, $\rho_0 = Q/(\frac{4}{3}\pi R^3)$. Rewrite the answers in terms of the total charge Q on the sphere.

Solution:

$$r > R, E = \frac{Q}{4\pi\epsilon_0 r^2}; r < R, E = \frac{qr}{4\pi\epsilon_0 R^3}$$

Exercise:**Problem:**

Suppose that the charge density of the spherical charge distribution shown in [\[link\]](#) is $\rho(r) = \rho_0 r/R$ for $r \leq R$ and zero for $r > R$. Obtain expressions for the electric field both inside and outside the distribution.

Exercise:**Problem:**

A very long, thin wire has a uniform linear charge density of $50 \mu\text{C}/\text{m}$. What is the electric field at a distance 2.0 cm from the wire?

Solution:

$$EA = \frac{\lambda l}{\epsilon_0} \Rightarrow E = 4.50 \times 10^7 \text{ N/C}$$

Exercise:**Problem:**

A charge of $-30 \mu\text{C}$ is distributed uniformly throughout a spherical volume of radius 10.0 cm. Determine the electric field due to this charge at a distance of (a) 2.0 cm, (b) 5.0 cm, and (c) 20.0 cm from the center of the sphere.

Exercise:**Problem:**

Repeat your calculations for the preceding problem, given that the charge is distributed uniformly over the surface of a spherical conductor of radius 10.0 cm.

Solution:

$$\text{a. } 0; \text{ b. } 0; \text{ c. } \vec{E} = 6.74 \times 10^6 \text{ N/C}(-\hat{r})$$

Exercise:**Problem:**

A total charge Q is distributed uniformly throughout a spherical shell of inner and outer radii r_1 and r_2 , respectively. Show that the electric field due to the charge is

$$\begin{aligned}\vec{E} &= \vec{0} & (r \leq r_1); \\ \vec{E} &= \frac{Q}{4\pi\epsilon_0 r^2} \left(\frac{r^3 - r_1^3}{r_2^3 - r_1^3} \right) \hat{r} & (r_1 \leq r \leq r_2); \\ \vec{E} &= \frac{Q}{4\pi\epsilon_0 r^2} \hat{r} & (r \geq r_2).\end{aligned}$$

Exercise:

Problem:

When a charge is placed on a metal sphere, it ends up in equilibrium at the outer surface. Use this information to determine the electric field of $+3.0 \mu\text{C}$ charge put on a 5.0-cm aluminum spherical ball at the following two points in space: (a) a point 1.0 cm from the center of the ball (an inside point) and (b) a point 10 cm from the center of the ball (an outside point).

Solution:

a. 0; b. $E = 2.70 \times 10^6 \text{ N/C}$

Exercise:

Problem:

A large sheet of charge has a uniform charge density of $10 \mu\text{C}/\text{m}^2$. What is the electric field due to this charge at a point just above the surface of the sheet?

Exercise:

Problem:

Determine if approximate cylindrical symmetry holds for the following situations. State why or why not. (a) A 300-cm long copper rod of radius 1 cm is charged with $+500 \text{ nC}$ of charge and we seek electric field at a point 5 cm from the center of the rod. (b) A 10-cm long copper rod of radius 1 cm is charged with $+500 \text{ nC}$ of charge and we seek electric field at a point 5 cm from the center of the rod. (c) A 150-cm wooden rod is glued to a 150-cm plastic rod to make a 300-cm long rod, which is then painted with a charged paint so that one obtains a uniform charge density. The radius of each rod is 1 cm, and we seek an electric field at a point that is 4 cm from the center of the rod. (d) Same rod as (c), but we seek electric field at a point that is 500 cm from the center of the rod.

Solution:

a. Yes, the length of the rod is much greater than the distance to the point in question. b. No, The length of the rod is of the same order of magnitude as the distance to the point in question. c. Yes, the length of the rod is much greater than the distance to the point in question. d. No. The length of the rod is of the same order of magnitude as the distance to the point in question.

Exercise:

Problem:

A long silver rod of radius 3 cm has a charge of $-5 \mu\text{C}/\text{cm}$ on its surface. (a) Find the electric field at a point 5 cm from the center of the rod (an outside point). (b) Find the electric field at a point 2 cm from the center of the rod (an inside point).

Exercise:**Problem:**

The electric field at 2 cm from the center of long copper rod of radius 1 cm has a magnitude 3 N/C and directed outward from the axis of the rod. (a) How much charge per unit length exists on the copper rod? (b) What would be the electric flux through a cube of side 5 cm situated such that the rod passes through opposite sides of the cube perpendicularly?

Solution:

$$\begin{aligned} \text{a. } \vec{E} &= \frac{R\sigma_0}{\varepsilon_0} \frac{1}{r} \hat{r} \Rightarrow \sigma_0 = 5.31 \times 10^{-11} \text{ C/m}^2, \\ \lambda &= 3.33 \times 10^{-12} \text{ C/m}; \\ \text{b. } \Phi &= \frac{q_{\text{enc}}}{\varepsilon_0} = \frac{3.33 \times 10^{-12} \text{ C/m}(0.05 \text{ m})}{\varepsilon_0} = 0.019 \text{ N} \cdot \text{m}^2/\text{C} \end{aligned}$$

Exercise:**Problem:**

A long copper cylindrical shell of inner radius 2 cm and outer radius 3 cm surrounds concentrically a charged long aluminum rod of radius 1 cm with a charge density of 4 pC/m. All charges on the aluminum rod reside at its surface. The inner surface of the copper shell has exactly opposite charge to that of the aluminum rod while the outer surface of the copper shell has the same charge as the aluminum rod. Find the magnitude and direction of the electric field at points that are at the following distances from the center of the aluminum rod: (a) 0.5 cm, (b) 1.5 cm, (c) 2.5 cm, (d) 3.5 cm, and (e) 7 cm.

Exercise:**Problem:**

Charge is distributed uniformly with a density ρ throughout an infinitely long cylindrical volume of radius R . Show that the field of this charge distribution is directed radially with respect to the cylinder and that

$$\begin{aligned} E &= \frac{\rho r}{2\varepsilon_0} & (r \leq R); \\ E &= \frac{\rho R^2}{2\varepsilon_0 r} & (r \geq R). \end{aligned}$$

Solution:

$$E2\pi rl = \frac{\rho\pi r^2 l}{\varepsilon_0} \Rightarrow E = \frac{\rho r}{2\varepsilon_0} \quad (r \leq R);$$

$$E2\pi rl = \frac{\rho\pi R^2 l}{\varepsilon_0} \Rightarrow E = \frac{\rho R^2}{2\varepsilon_0 r} \quad (r \geq R)$$

Exercise:

Problem:

Charge is distributed throughout a very long cylindrical volume of radius R such that the charge density increases with the distance r from the central axis of the cylinder according to $\rho = \alpha r$, where α is a constant. Show that the field of this charge distribution is directed radially with respect to the cylinder and that

$$E = \frac{\alpha r^2}{3\varepsilon_0} \quad (r \leq R);$$

$$E = \frac{\alpha R^3}{3\varepsilon_0 r} \quad (r \geq R).$$

Exercise:

Problem:

The electric field 10.0 cm from the surface of a copper ball of radius 5.0 cm is directed toward the ball's center and has magnitude 4.0×10^2 N/C. How much charge is on the surface of the ball?

Solution:

$$\Phi = \frac{q_{\text{enc}}}{\varepsilon_0} \Rightarrow q_{\text{enc}} = -1.0 \times 10^{-9} \text{ C}$$

Exercise:

Problem:

Charge is distributed throughout a spherical shell of inner radius r_1 and outer radius r_2 with a volume density given by $\rho = \rho_0 r_1/r$, where ρ_0 is a constant. Determine the electric field due to this charge as a function of r , the distance from the center of the shell.

Exercise:

Problem:

Charge is distributed throughout a spherical volume of radius R with a density $\rho = \alpha r^2$, where α is a constant. Determine the electric field due to the charge at points both inside and outside the sphere.

Solution:

$$q_{\text{enc}} = \frac{4}{5} \pi \alpha r^5,$$

$$E4\pi r^2 = \frac{4\pi \alpha r^5}{5\varepsilon_0} \Rightarrow E = \frac{\alpha r^3}{5\varepsilon_0} \quad (r \leq R),$$

$$q_{\text{enc}} = \frac{4}{5} \pi \alpha R^5, \quad E4\pi r^2 = \frac{4\pi \alpha R^5}{5\varepsilon_0} \Rightarrow E = \frac{\alpha R^5}{5\varepsilon_0 r^2} \quad (r \geq R)$$

Exercise:**Problem:**

Consider a uranium nucleus to be sphere of radius $R = 7.4 \times 10^{-15}$ m with a charge of $92e$ distributed uniformly throughout its volume. (a) What is the electric force exerted on an electron when it is 3.0×10^{-15} m from the center of the nucleus? (b) What is the acceleration of the electron at this point?

Exercise:**Problem:**

The volume charge density of a spherical charge distribution is given by $\rho(r) = \rho_0 e^{-\alpha r}$, where ρ_0 and α are constants. What is the electric field produced by this charge distribution?

Solution:

integrate by parts:

$$q_{\text{enc}} = 4\pi\rho_0 \left[-e^{-\alpha r} \left(\frac{(r)^2}{\alpha} + \frac{2r}{\alpha^2} + \frac{2}{\alpha^3} \right) + \frac{2}{\alpha^3} \right] \Rightarrow E = \frac{\rho_0}{r^2 \epsilon_0} \left[-e^{-\alpha r} \left(\frac{(r)^2}{\alpha} + \frac{2r}{\alpha^2} + \frac{2}{\alpha^3} \right) + \frac{2}{\alpha^3} \right]$$

Glossary

cylindrical symmetry

system only varies with distance from the axis, not direction

planar symmetry

system only varies with distance from a plane

spherical symmetry

system only varies with the distance from the origin, not in direction

Conductors in Electrostatic Equilibrium

By the end of this section, you will be able to:

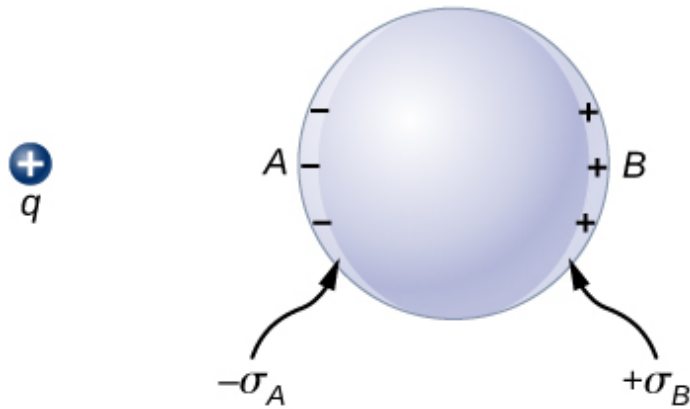
- Describe the electric field within a conductor at equilibrium
- Describe the electric field immediately outside the surface of a charged conductor at equilibrium
- Explain why if the field is not as described in the first two objectives, the conductor is not at equilibrium

So far, we have generally been working with charges occupying a volume within an insulator. We now study what happens when free charges are placed on a conductor. Generally, in the presence of a (generally external) electric field, the free charge in a conductor redistributes and very quickly reaches electrostatic equilibrium. The resulting charge distribution and its electric field have many interesting properties, which we can investigate with the help of Gauss's law and the concept of electric potential.

The Electric Field inside a Conductor Vanishes

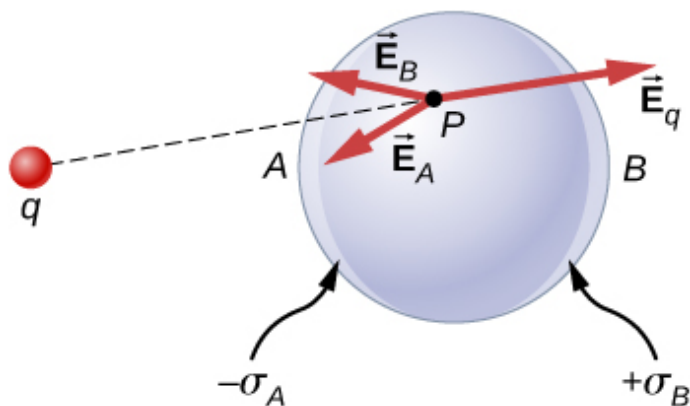
If an electric field is present inside a conductor, it exerts forces on the **free electrons** (also called conduction electrons), which are electrons in the material that are not bound to an atom. These free electrons then accelerate. However, moving charges by definition means nonstatic conditions, contrary to our assumption. Therefore, when electrostatic equilibrium is reached, the charge is distributed in such a way that the electric field inside the conductor vanishes.

If you place a piece of a metal near a positive charge, the free electrons in the metal are attracted to the external positive charge and migrate freely toward that region. The region the electrons move to then has an excess of electrons over the protons in the atoms and the region from where the electrons have migrated has more protons than electrons. Consequently, the metal develops a negative region near the charge and a positive region at the far end ([\[link\]](#)). As we saw in the preceding chapter, this separation of equal magnitude and opposite type of electric charge is called polarization. If you remove the external charge, the electrons migrate back and neutralize the positive region.



Polarization of a metallic sphere by an external point charge $+q$. The near side of the metal has an opposite surface charge compared to the far side of the metal. The sphere is said to be polarized. When you remove the external charge, the polarization of the metal also disappears.

The polarization of the metal happens only in the presence of external charges. You can think of this in terms of electric fields. The external charge creates an external electric field. When the metal is placed in the region of this electric field, the electrons and protons of the metal experience electric forces due to this external electric field, but only the conduction electrons are free to move in the metal over macroscopic distances. The movement of the conduction electrons leads to the polarization, which creates an induced electric field in addition to the external electric field ([\[link\]](#)). The net electric field is a vector sum of the fields of $+q$ and the surface charge densities $-\sigma_A$ and $+\sigma_B$. This means that the net field inside the conductor is different from the field outside the conductor.



In the presence of an external charge q , the charges in a metal redistribute.

The electric field at any point has three contributions, from $+q$ and the induced charges $-σ_A$ and $+σ_B$. Note that the surface charge distribution will not be uniform in this case.

The redistribution of charges is such that the sum of the three contributions at any point P inside the conductor is

Equation:

$$\vec{E}_P = \vec{E}_q + \vec{E}_B + \vec{E}_A = \vec{0}.$$

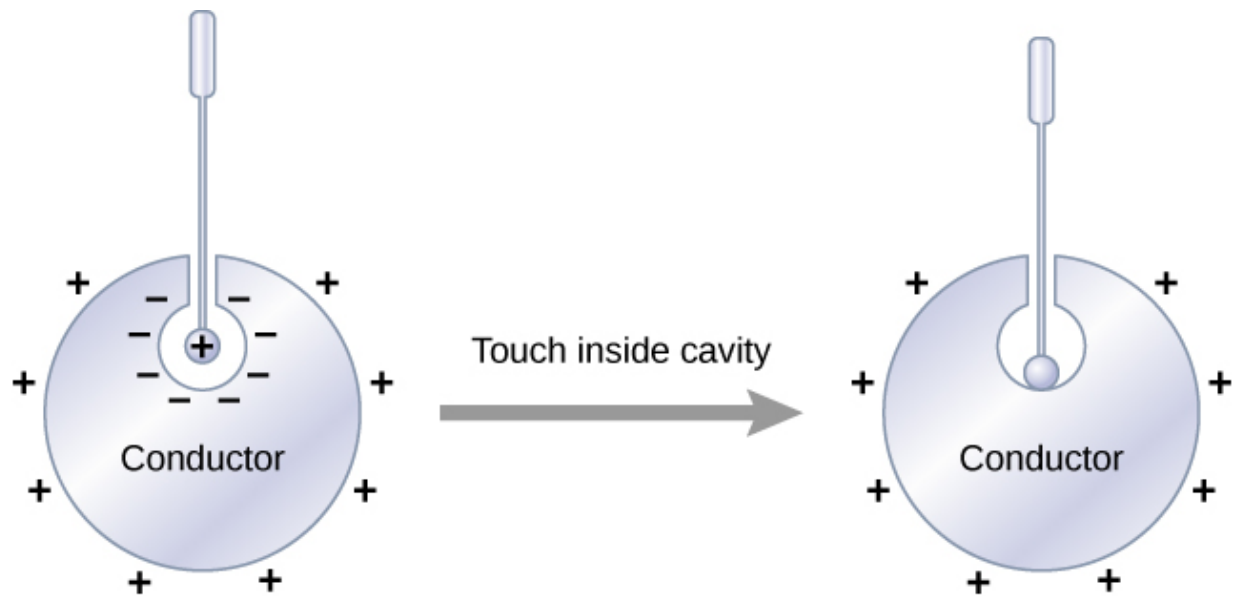
Now, thanks to Gauss's law, we know that there is no net charge enclosed by a Gaussian surface that is solely within the volume of the conductor at equilibrium. That is, $q_{\text{enc}} = 0$ and hence

Equation:

$$\vec{E}_{\text{net}} = \vec{0} \text{ (at points inside a conductor).}$$

Charge on a Conductor

An interesting property of a conductor in static equilibrium is that extra charges on the conductor end up on the outer surface of the conductor, regardless of where they originate. [\[link\]](#) illustrates a system in which we bring an external positive charge inside the cavity of a metal and then touch it to the inside surface. Initially, the inside surface of the cavity is negatively charged and the outside surface of the conductor is positively charged. When we touch the inside surface of the cavity, the induced charge is neutralized, leaving the outside surface and the whole metal charged with a net positive charge.



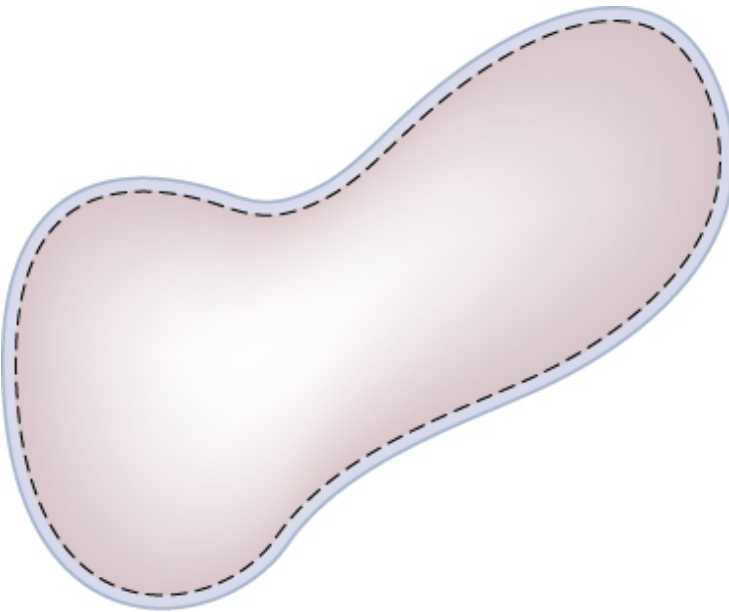
Electric charges on a conductor migrate to the outside surface no matter where you put them initially.

To see why this happens, note that the Gaussian surface in [\[link\]](#) (the dashed line) follows the contour of the actual surface of the conductor and is located an infinitesimal distance *within* it. Since $E = 0$ everywhere inside a conductor,

Equation:

$$\oint_s \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = 0.$$

Thus, from Gauss's law, there is no net charge inside the Gaussian surface. But the Gaussian surface lies just below the actual surface of the conductor; consequently, there is no net charge inside the conductor. Any excess charge must lie on its surface.



The dashed line represents a Gaussian surface that is just beneath the actual surface of the conductor.

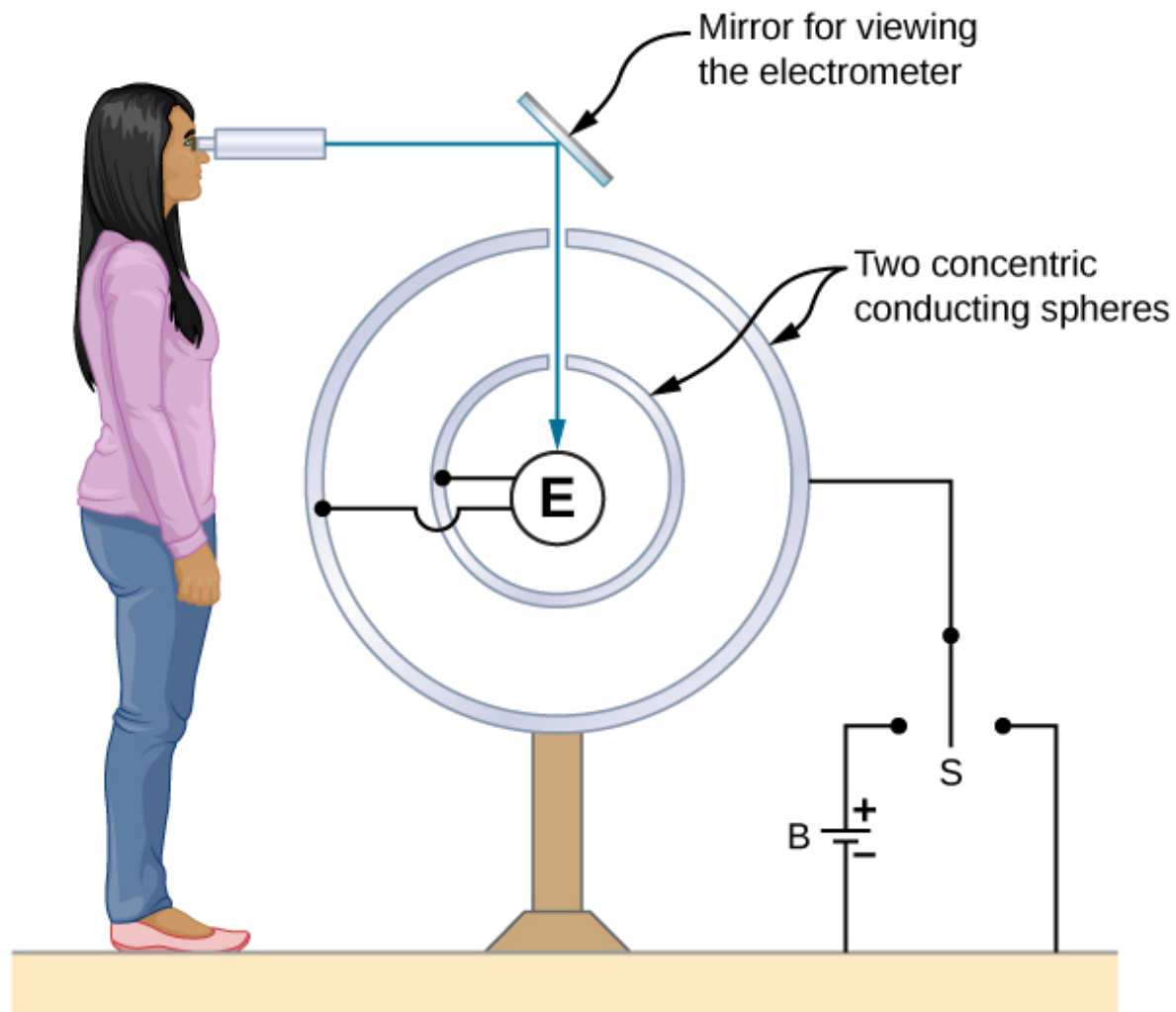
This particular property of conductors is the basis for an extremely accurate method developed by Plimpton and Lawton in 1936 to verify Gauss's law and, correspondingly, Coulomb's law. A sketch of their apparatus is shown in [\[link\]](#). Two spherical shells are connected to one another through an electrometer E, a device that can detect a very slight amount of charge flowing from one shell to the other. When switch S is thrown to the left,

charge is placed on the outer shell by the battery B. Will charge flow through the electrometer to the inner shell?

No. Doing so would mean a violation of Gauss's law. Plimpton and Lawton did not detect any flow and, knowing the sensitivity of their electrometer, concluded that if the radial dependence in Coulomb's law were $1/r^{(2+\delta)}$, δ would be less than 2×10^{-9} [\[footnote\]](#). More recent measurements place δ at less than 3×10^{-16} [\[footnote\]](#), a number so small that the validity of Coulomb's law seems indisputable.

S. Plimpton and W. Lawton. 1936. "A Very Accurate Test of Coulomb's Law of Force between Charges." *Physical Review* 50, No. 11: 1066, doi:10.1103/PhysRev.50.1066

E. Williams, J. Faller, and H. Hill. 1971. "New Experimental Test of Coulomb's Law: A Laboratory Upper Limit on the Photon Rest Mass." *Physical Review Letters* 26, No. 12: 721, doi:10.1103/PhysRevLett.26.721



A representation of the apparatus used by Plimpton and Lawton. Any transfer of charge between the spheres is detected by the electrometer E.

The Electric Field at the Surface of a Conductor

If the electric field had a component parallel to the surface of a conductor, free charges on the surface would move, a situation contrary to the assumption of electrostatic equilibrium. Therefore, the electric field is always perpendicular to the surface of a conductor.

At any point just above the surface of a conductor, the surface charge density σ and the magnitude of the electric field E are related by

Note:

Equation:

$$E = \frac{\sigma}{\epsilon_0}.$$

To see this, consider an infinitesimally small Gaussian cylinder that surrounds a point on the surface of the conductor, as in [\[link\]](#). The cylinder has one end face inside and one end face outside the surface. The height and cross-sectional area of the cylinder are δ and ΔA , respectively. The cylinder's sides are perpendicular to the surface of the conductor, and its end faces are parallel to the surface. Because the cylinder is infinitesimally small, the charge density σ is essentially constant over the surface enclosed, so the total charge inside the Gaussian cylinder is $\sigma\Delta A$. Now E is perpendicular to the surface of the conductor outside the conductor and vanishes within it, because otherwise, the charges would accelerate, and we would not be in equilibrium. Electric flux therefore crosses only the outer end face of the Gaussian surface and may be written as $E\Delta A$, since the cylinder is assumed to be small enough that E is approximately constant over that area. From Gauss' law,

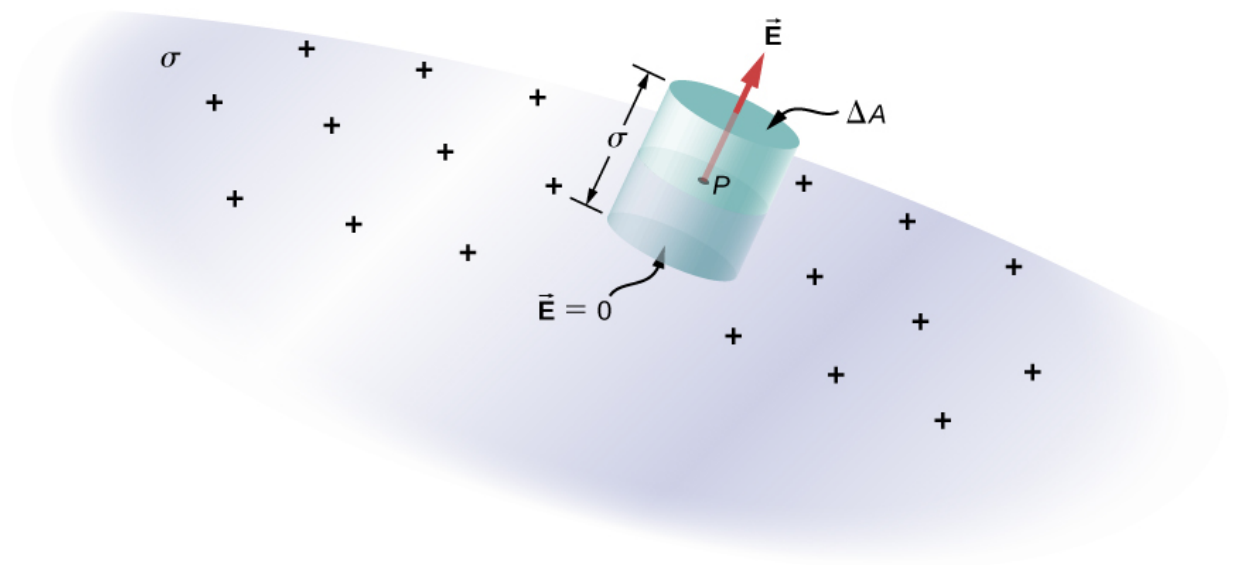
Equation:

$$E\Delta A = \frac{\sigma\Delta A}{\epsilon_0}.$$

Thus,

Equation:

$$E = \frac{\sigma}{\epsilon_0}.$$

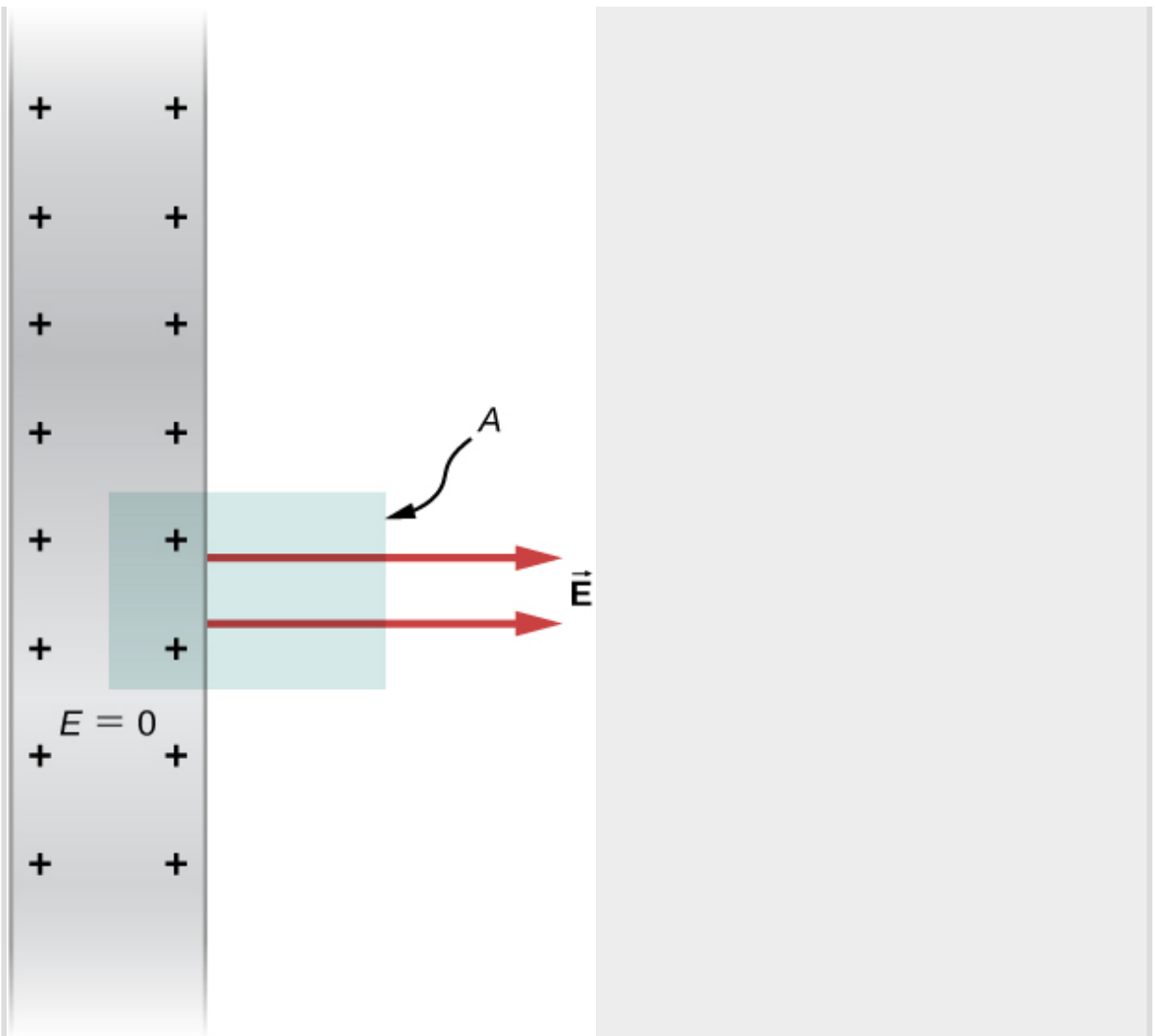


An infinitesimally small cylindrical Gaussian surface surrounds point P , which is on the surface of the conductor. The field \vec{E} is perpendicular to the surface of the conductor outside the conductor and vanishes within it.

Example:

Electric Field of a Conducting Plate

The infinite conducting plate in [\[link\]](#) has a uniform surface charge density σ . Use Gauss' law to find the electric field outside the plate. Compare this result with that previously calculated directly.



A side view of an infinite conducting plate and Gaussian cylinder with cross-sectional area A .

Strategy

For this case, we use a cylindrical Gaussian surface, a side view of which is shown.

Solution

The flux calculation is similar to that for an infinite sheet of charge from the previous chapter with one major exception: The left face of the Gaussian

surface is inside the conductor where $\vec{E} = \vec{0}$, so the total flux through the Gaussian surface is EA rather than $2EA$. Then from Gauss' law,

Equation:

$$EA = \frac{\sigma A}{\epsilon_0}$$

and the electric field outside the plate is

Equation:

$$E = \frac{\sigma}{\epsilon_0}.$$

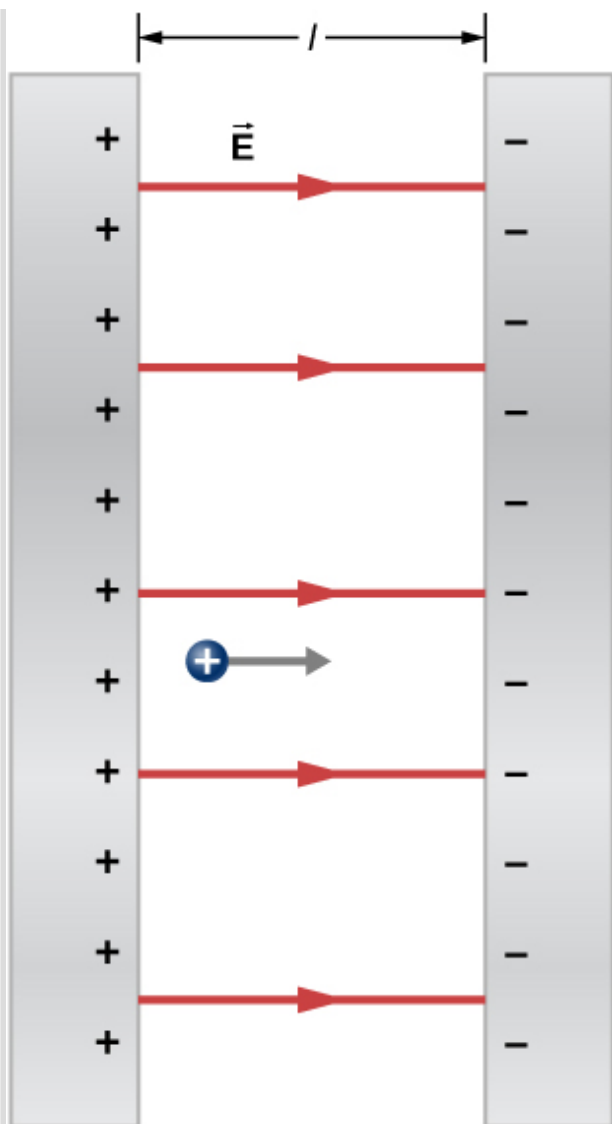
Significance

This result is in agreement with the result from the previous section, and consistent with the rule stated above.

Example:

Electric Field between Oppositely Charged Parallel Plates

Two large conducting plates carry equal and opposite charges, with a surface charge density σ of magnitude $6.81 \times 10^{-7} \text{ C/m}^2$, as shown in [\[link\]](#). The separation between the plates is $l = 6.50 \text{ mm}$. What is the electric field between the plates?



The electric field between oppositely charged parallel plates. A test charge is released at the positive plate.

Strategy

Note that the electric field at the surface of one plate only depends on the charge on that plate. Thus, apply $E = \sigma/\epsilon_0$ with the given values.

Solution

The electric field is directed from the positive to the negative plate, as shown in the figure, and its magnitude is given by

Equation:

$$E = \frac{\sigma}{\epsilon_0} = \frac{6.81 \times 10^{-7} \text{ C/m}^2}{8.85 \times 10^{-12} \text{ C}^2/\text{N m}^2} = 7.69 \times 10^4 \text{ N/C}.$$

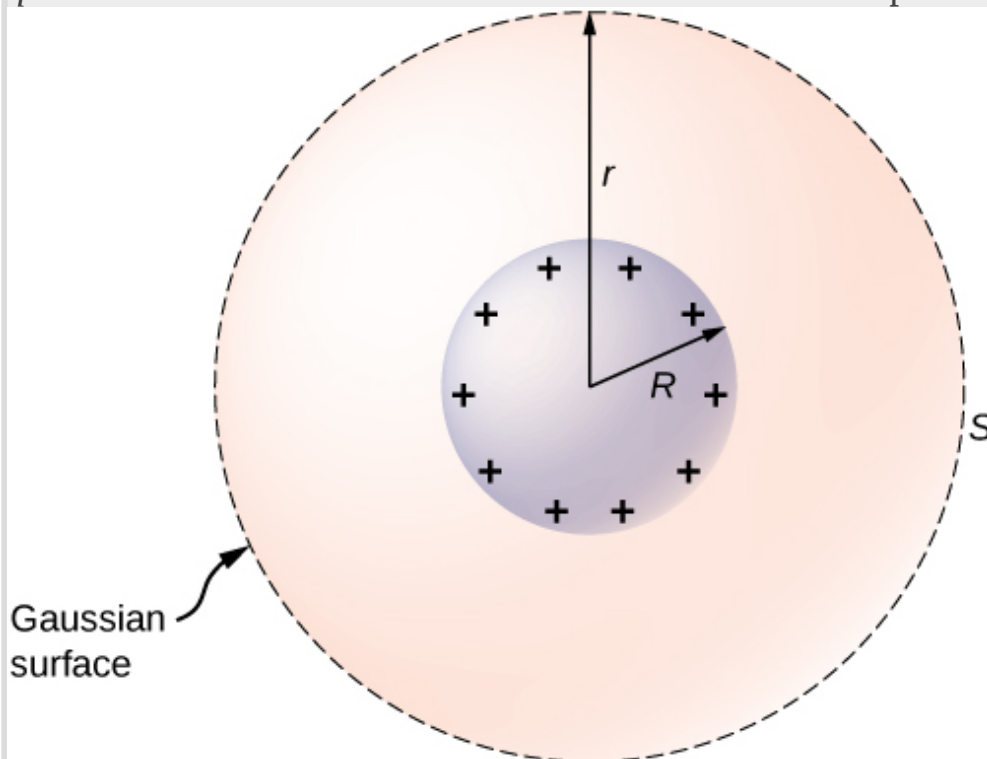
Significance

This formula is applicable to more than just a plate. Furthermore, two-plate systems will be important later.

Example:

A Conducting Sphere

The isolated conducting sphere ([link](#)) has a radius R and an excess charge q . What is the electric field both inside and outside the sphere?



An isolated conducting sphere.

Strategy

The sphere is isolated, so its surface charge distribution and the electric field of that distribution are spherically symmetrical. We can therefore represent the field as $\vec{\mathbf{E}} = E(r)\hat{\mathbf{r}}$. To calculate $E(r)$, we apply Gauss's law over a closed spherical surface S of radius r that is concentric with the conducting sphere.

Solution

Since r is constant and $\hat{\mathbf{n}} = \hat{\mathbf{r}}$ on the sphere,

Equation:

$$\oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = E(r) \oint_S dA = E(r) 4\pi r^2.$$

For $r < R$, S is within the conductor, so $q_{\text{enc}} = 0$, and Gauss's law gives

Equation:

$$E(r) = 0,$$

as expected inside a conductor. If $r > R$, S encloses the conductor so $q_{\text{enc}} = q$. From Gauss's law,

Equation:

$$E(r) 4\pi r^2 = \frac{q}{\epsilon_0}.$$

The electric field of the sphere may therefore be written as

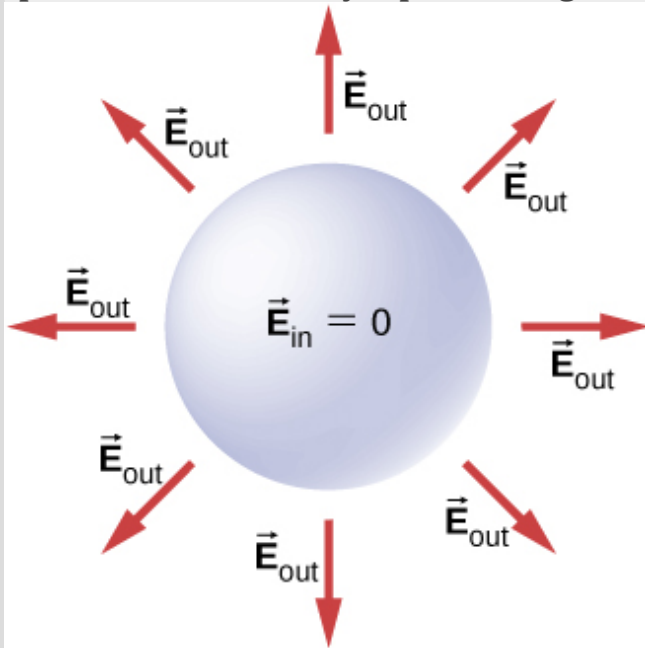
Equation:

$$\begin{aligned} \vec{\mathbf{E}} &= \vec{\mathbf{0}} & (r < R), \\ \vec{\mathbf{E}} &= \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \hat{\mathbf{r}} & (r \geq R). \end{aligned}$$

Significance

Notice that in the region $r \geq R$, the electric field due to a charge q placed on an isolated conducting sphere of radius R is identical to the electric field of a point charge q located at the center of the sphere. The difference between the charged metal and a point charge occurs only at the space

points inside the conductor. For a point charge placed at the center of the sphere, the electric field is not zero at points of space occupied by the sphere, but a conductor with the same amount of charge has a zero electric field at those points ([\[link\]](#)). However, there is no distinction at the outside points in space where $r > R$, and we can replace the isolated charged spherical conductor by a point charge at its center with impunity.



Electric field of a positively charged metal sphere. The electric field inside is zero, and the electric field outside is same as the electric field of a point charge at the center, although the charge on the metal sphere is at the surface.

Note:

Exercise:

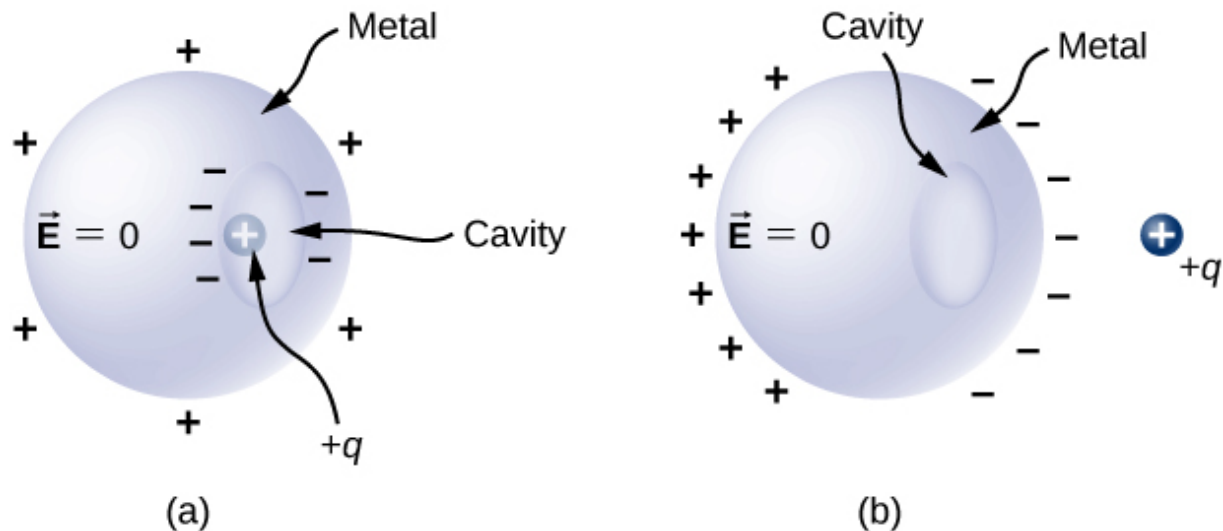
Problem:

Check Your Understanding How will the system above change if there are charged objects external to the sphere?

Solution:

If there are other charged objects around, then the charges on the surface of the sphere will not necessarily be spherically symmetrical; there will be more in certain direction than in other directions.

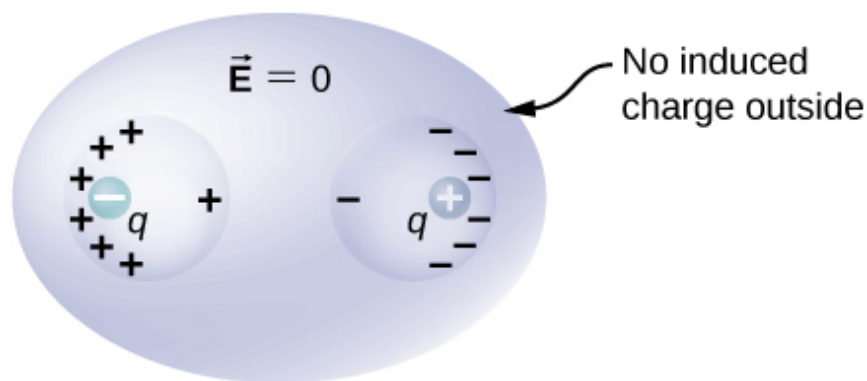
For a conductor with a cavity, if we put a charge $+q$ inside the cavity, then the charge separation takes place in the conductor, with $-q$ amount of charge on the inside surface and a $+q$ amount of charge at the outside surface ([link](#) (a)). For the same conductor with a charge $+q$ outside it, there is no excess charge on the inside surface; both the positive and negative induced charges reside on the outside surface ([link](#) (b)).



(a) A charge inside a cavity in a metal. The distribution of charges at the outer surface does not depend on how the charges are distributed at the inner surface, since the E -field inside the body of the metal is zero.

That magnitude of the charge on the outer surface does depend on the magnitude of the charge inside, however. (b) A charge outside a conductor containing an inner cavity. The cavity remains free of charge. The polarization of charges on the conductor happens at the surface.

If a conductor has two cavities, one of them having a charge $+q_a$ inside it and the other a charge $-q_b$, the polarization of the conductor results in $-q_a$ on the inside surface of the cavity a , $+q_b$ on the inside surface of the cavity b , and $q_a - q_b$ on the outside surface ([link](#)). The charges on the surfaces may not be uniformly spread out; their spread depends upon the geometry. The only rule obeyed is that when the equilibrium has been reached, the charge distribution in a conductor is such that the electric field by the charge distribution in the conductor cancels the electric field of the external charges at all space points inside the body of the conductor.



The charges induced by two equal and opposite charges in two separate cavities of a conductor.

If the net charge on the cavity is nonzero, the external surface becomes charged to the amount of the net charge.

Summary

- The electric field inside a conductor vanishes.
- Any excess charge placed on a conductor resides entirely on the surface of the conductor.
- The electric field is perpendicular to the surface of a conductor everywhere on that surface.
- The magnitude of the electric field just above the surface of a conductor is given by $E = \frac{\sigma}{\epsilon_0}$.

Key Equations

Definition of electric flux, for uniform electric field	$\Phi = \vec{\mathbf{E}} \cdot \vec{\mathbf{A}} \rightarrow EA \cos \theta$
Electric flux through an open surface	$\Phi = \int_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = \int_S \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$
Electric flux through a closed surface	$\Phi = \oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = \oint_S \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$
Gauss's law	$\Phi = \oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = \frac{q_{\text{enc}}}{\epsilon_0}$
Gauss's Law for systems with symmetry	$\Phi = \oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = E \oint_S dA = EA = \frac{q_{\text{enc}}}{\epsilon_0}$
The magnitude of the electric field	$E = \frac{\sigma}{\epsilon_0}$

just outside the surface of a conductor	
---	--

Conceptual Questions

Exercise:

Problem: Is the electric field inside a metal always zero?

Exercise:

Problem:

Under electrostatic conditions, the excess charge on a conductor resides on its surface. Does this mean that all the conduction electrons in a conductor are on the surface?

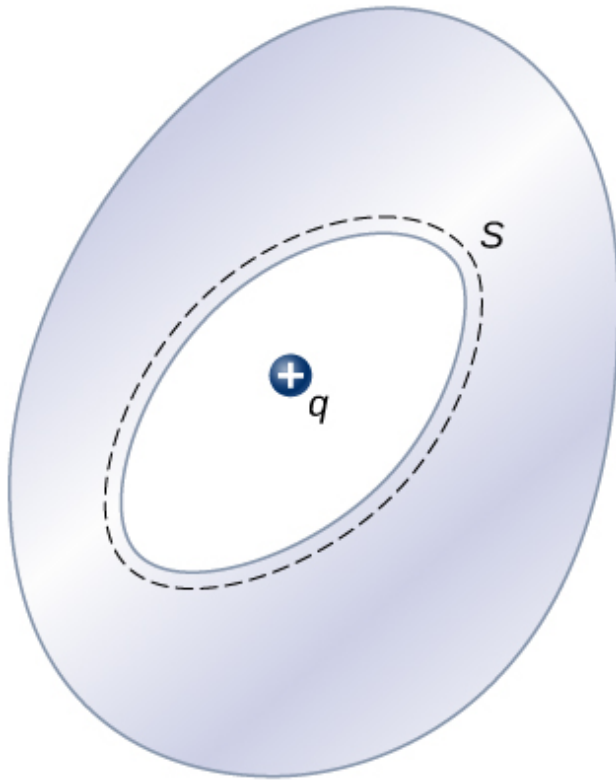
Solution:

No. If a metal was in a region of zero electric field, all the conduction electrons would be distributed uniformly throughout the metal.

Exercise:

Problem:

A charge q is placed in the cavity of a conductor as shown below. Will a charge outside the conductor experience an electric field due to the presence of q ?



Exercise:

Problem:

The conductor in the preceding figure has an excess charge of $-5.0\ \mu\text{C}$. If a $2.0\text{-}\mu\text{C}$ point charge is placed in the cavity, what is the net charge on the surface of the cavity and on the outer surface of the conductor?

Solution:

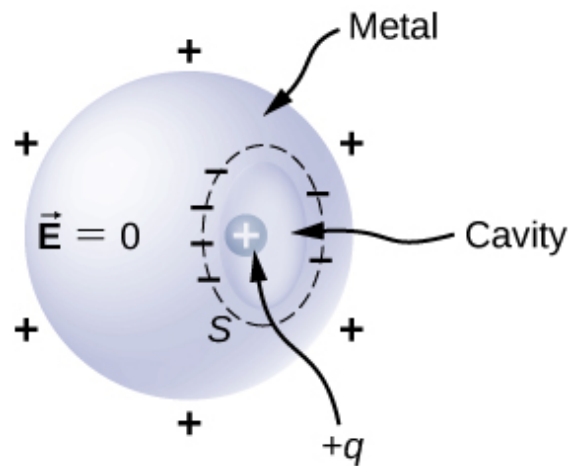
Since the electric field is zero inside a conductor, a charge of $-2.0\ \mu\text{C}$ is induced on the inside surface of the cavity. This will put a charge of $+2.0\ \mu\text{C}$ on the outside surface leaving a net charge of $-3.0\ \mu\text{C}$ on the surface.

Problems

Exercise:

Problem:

An uncharged conductor with an internal cavity is shown in the following figure. Use the closed surface S along with Gauss' law to show that when a charge q is placed in the cavity a total charge $-q$ is induced on the inner surface of the conductor. What is the charge on the outer surface of the conductor?

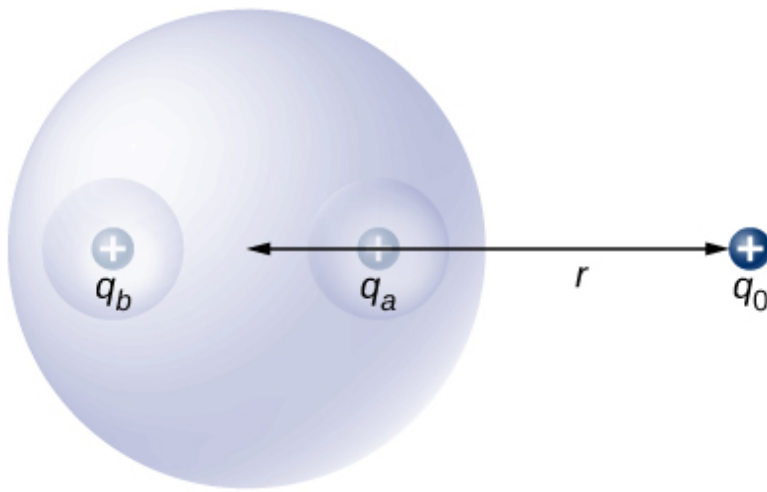


A charge inside a cavity of a metal. Charges at the outer surface do not depend on how the charges are distributed at the inner surface since E field inside the body of the metal is zero.

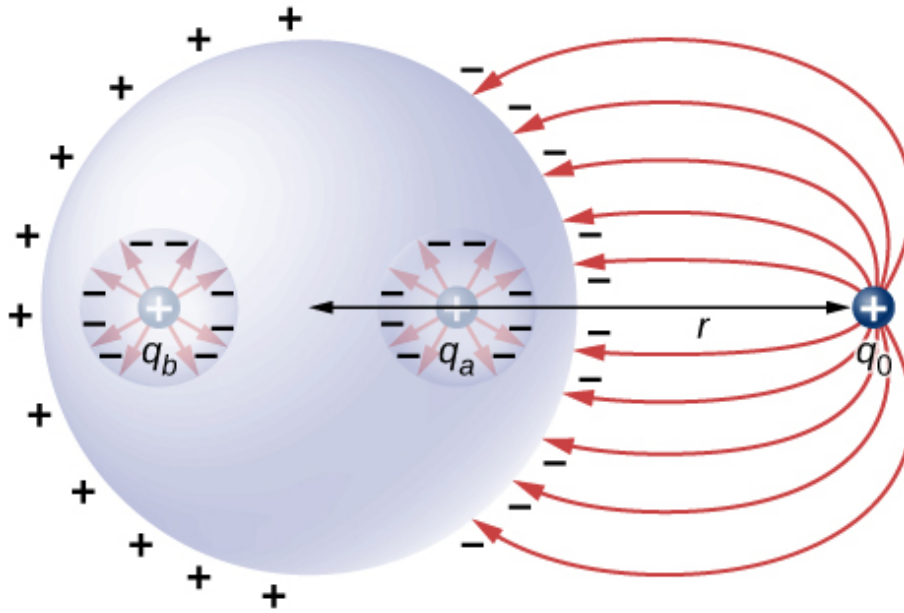
Exercise:

Problem:

An uncharged spherical conductor S of radius R has two spherical cavities A and B of radii a and b , respectively as shown below. Two point charges $+q_a$ and $+q_b$ are placed at the center of the two cavities by using non-conducting supports. In addition, a point charge $+q_0$ is placed outside at a distance r from the center of the sphere. (a) Draw approximate charge distributions in the metal although metal sphere has no net charge. (b) Draw electric field lines. Draw enough lines to represent all distinctly different places.



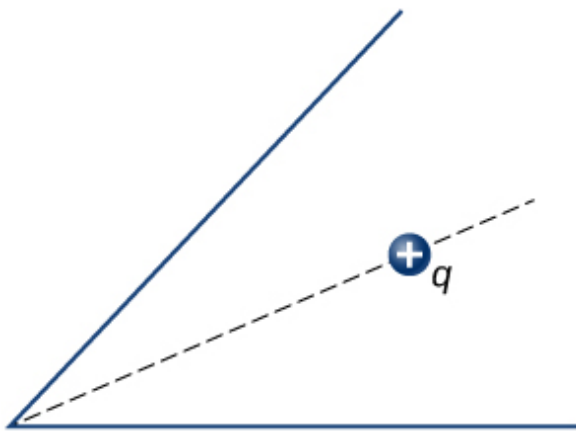
Solution:



Exercise:

Problem:

A positive point charge is placed at the angle bisector of two uncharged plane conductors that make an angle of 45° . See below. Draw the electric field lines.



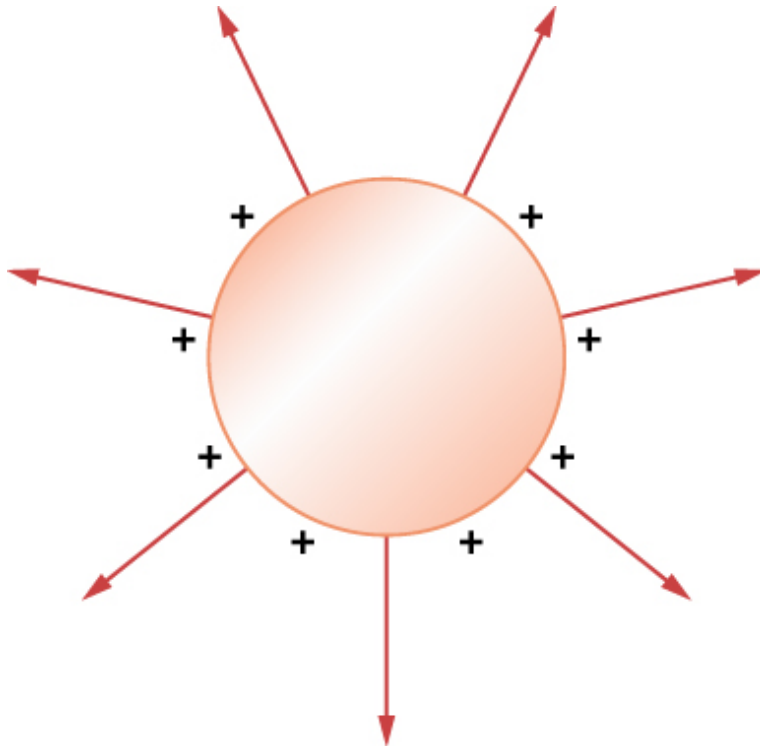
Exercise:

Problem:

A long cylinder of copper of radius 3 cm is charged so that it has a uniform charge per unit length on its surface of 3 C/m. (a) Find the electric field inside and outside the cylinder. (b) Draw electric field lines in a plane perpendicular to the rod.

Solution:

a. Outside: $E2\pi rl = \frac{\lambda l}{\epsilon_0} \Rightarrow E = \frac{3.0 \text{ C/m}}{2\pi\epsilon_0 r}$; Inside $E_{\text{in}} = 0$; b.

**Exercise:**

Problem:

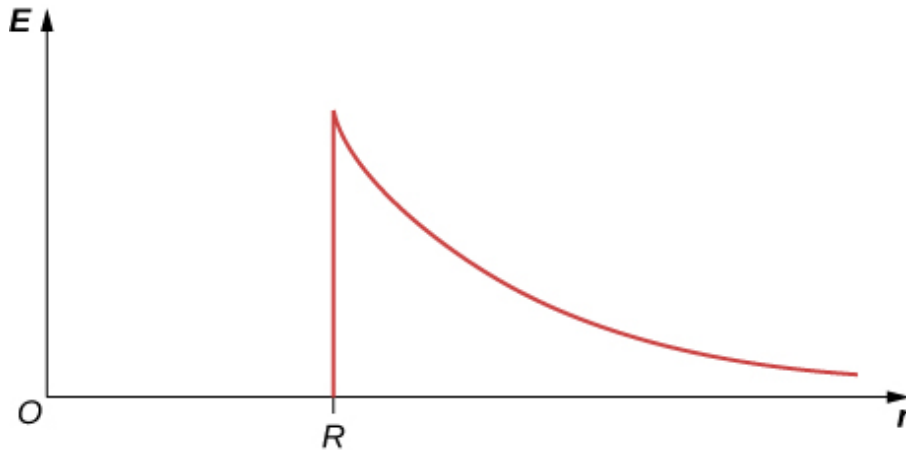
An aluminum spherical ball of radius 4 cm is charged with $5 \mu\text{C}$ of charge. A copper spherical shell of inner radius 6 cm and outer radius 8 cm surrounds it. A total charge of $-8 \mu\text{C}$ is put on the copper shell. (a) Find the electric field at all points in space, including points inside the aluminum and copper shell when copper shell and aluminum sphere are concentric. (b) Find the electric field at all points in space, including points inside the aluminum and copper shell when the centers of copper shell and aluminum sphere are 1 cm apart.

Exercise:**Problem:**

A long cylinder of aluminum of radius R meters is charged so that it has a uniform charge per unit length on its surface of λ . (a) Find the electric field inside and outside the cylinder. (b) Plot electric field as a function of distance from the center of the rod.

Solution:

a. $E 2\pi r l = \frac{\lambda l}{\epsilon_0} \Rightarrow E = \frac{\lambda}{2\pi\epsilon_0 r} \quad r \geq R$ E inside equals 0; b.

**Exercise:**

Problem:

At the surface of any conductor in electrostatic equilibrium, $E = \sigma/\epsilon_0$. Show that this equation is consistent with the fact that $E = kq/r^2$ at the surface of a spherical conductor.

Exercise:**Problem:**

Two parallel plates 10 cm on a side are given equal and opposite charges of magnitude 5.0×10^{-9} C. The plates are 1.5 mm apart. What is the electric field at the center of the region between the plates?

Solution:

$$E = 5.65 \times 10^4 \text{ N/C}$$

Exercise:**Problem:**

Two parallel conducting plates, each of cross-sectional area 400 cm^2 , are 2.0 cm apart and uncharged. If 1.0×10^{12} electrons are transferred from one plate to the other, what are (a) the charge density on each plate? (b) The electric field between the plates?

Exercise:**Problem:**

The surface charge density on a long straight metallic pipe is σ . What is the electric field outside and inside the pipe? Assume the pipe has a diameter of $2a$.



Solution:

$$\lambda = \frac{\lambda l}{\varepsilon_0} \Rightarrow E = \frac{a\sigma}{\varepsilon_0 r} \quad r \geq a, \quad E = 0 \text{ inside since } q_{\text{enclosed}} = 0$$

Exercise:

Problem:

A point charge $q = -5.0 \times 10^{-12} \text{ C}$ is placed at the center of a spherical conducting shell of inner radius 3.5 cm and outer radius 4.0 cm. The electric field just above the surface of the conductor is directed radially outward and has magnitude 8.0 N/C. (a) What is the charge density on the inner surface of the shell? (b) What is the charge density on the outer surface of the shell? (c) What is the net charge on the conductor?

Exercise:**Problem:**

A solid cylindrical conductor of radius a is surrounded by a concentric cylindrical shell of inner radius b . The solid cylinder and the shell carry charges $+Q$ and $-Q$, respectively. Assuming that the length L of both conductors is much greater than a or b , determine the electric field as a function of r , the distance from the common central axis of the cylinders, for (a) $r < a$; (b) $a < r < b$; and (c) $r > b$.

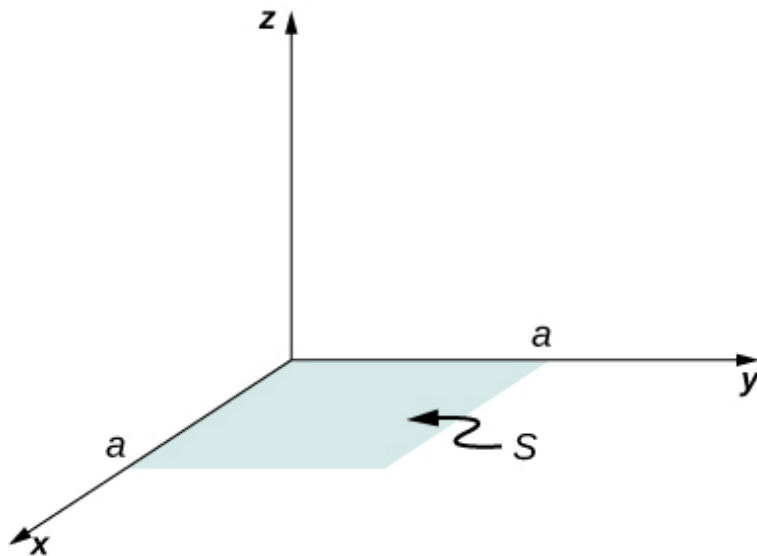
Solution:

a. $E = 0$; b. $E2\pi rL = \frac{Q}{\epsilon_0} \Rightarrow E = \frac{Q}{2\pi\epsilon_0 rL}$; c. $E = 0$ since r would be either inside the second shell or if outside then q enclosed equals 0.

Additional Problems**Exercise:****Problem:**

A vector field \vec{E} (not necessarily an electric field; note units) is given by $\vec{E} = 3x^2\hat{k}$. Calculate $\int_S \vec{E} \cdot \hat{n} da$, where S is the area shown below.

Assume that $\hat{n} = \hat{k}$.



Exercise:

Problem: Repeat the preceding problem, with $\vec{\mathbf{E}} = 2x\hat{\mathbf{i}} + 3x^2\hat{\mathbf{k}}$.

Solution:

$$\int \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = a^4$$

Exercise:

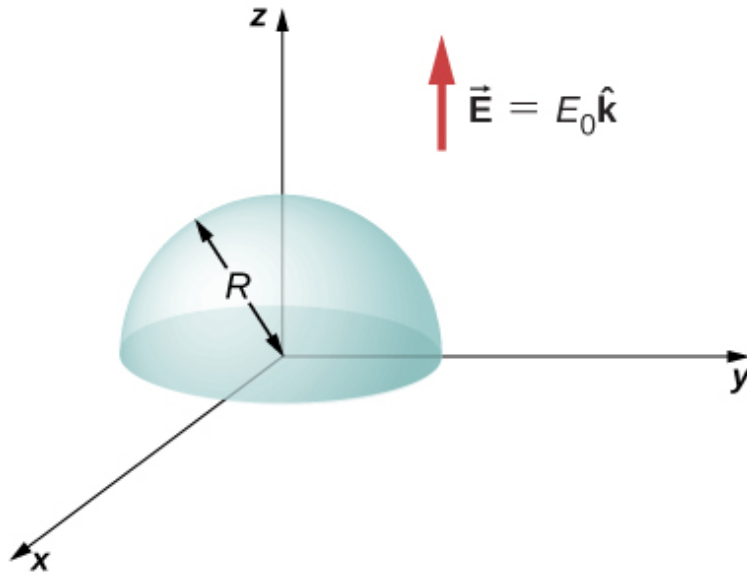
Problem:

A circular area S is concentric with the origin, has radius a , and lies in the yz -plane. Calculate $\int_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA$ for $\vec{\mathbf{E}} = 3z^2\hat{\mathbf{i}}$.

Exercise:

Problem:

(a) Calculate the electric flux through the open hemispherical surface due to the electric field $\vec{\mathbf{E}} = E_0\hat{\mathbf{k}}$ (see below). (b) If the hemisphere is rotated by 90° around the x -axis, what is the flux through it?



Solution:

- a. $\int \vec{E} \cdot \hat{n} dA = E_0 r^2 \pi$; b. zero, since the flux through the upper half cancels the flux through the lower half of the sphere

Exercise:

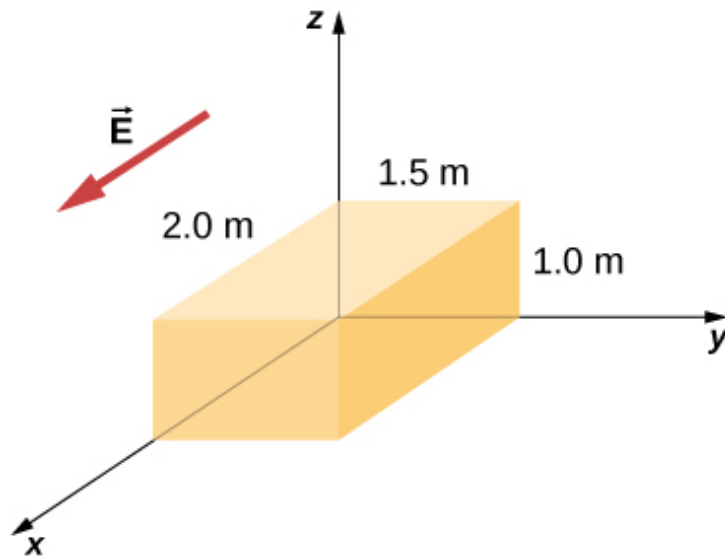
Problem:

Suppose that the electric field of an isolated point charge were proportional to $1/r^{2+\sigma}$ rather than $1/r^2$. Determine the flux that passes through the surface of a sphere of radius R centered at the charge. Would Gauss's law remain valid?

Exercise:

Problem:

The electric field in a region is given by $\vec{E} = a/(b + cx)\hat{i}$, where $a = 200 \text{ N} \cdot \text{m}/\text{C}$, $b = 2.0 \text{ m}$, and $c = 2.0$. What is the net charge enclosed by the shaded volume shown below?



Solution:

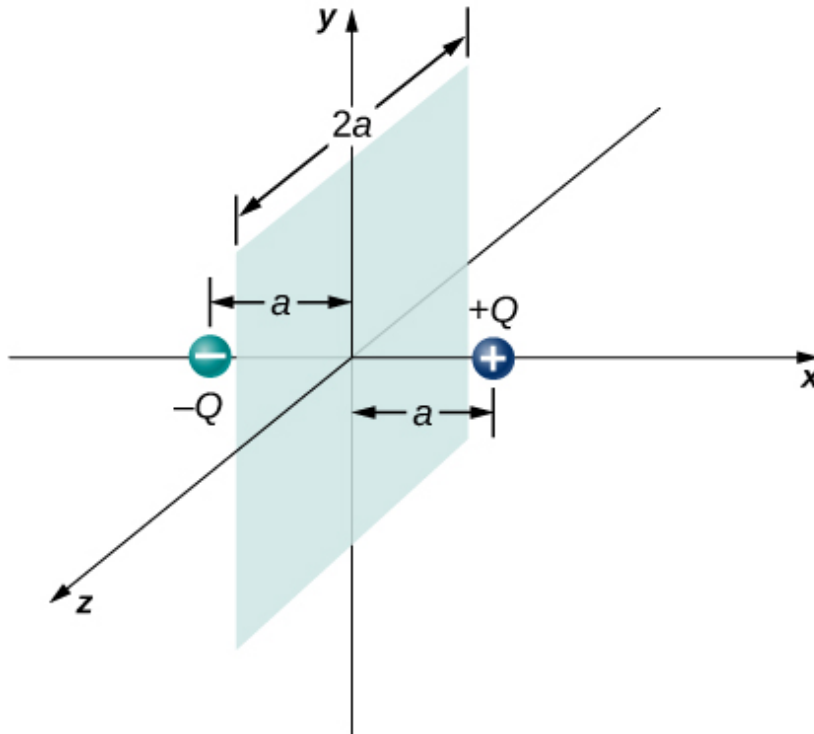
$\Phi = \frac{q_{\text{enc}}}{\epsilon_0}$; There are two contributions to the surface integral: one at the side of the rectangle at $x = 0$ and the other at the side at $x = 2.0 \text{ m}$;
 $-E(0)[1.5 \text{ m}^2] + E(2.0 \text{ m})[1.5 \text{ m}^2] = \frac{q_{\text{enc}}}{\epsilon_0} = -100 \text{ Nm}^2/\text{C}$
 where the minus sign indicates that at $x = 0$, the electric field is along positive x and the unit normal is along negative x . At $x = 2$, the unit normal and the electric field vector are in the same direction:

$$q_{\text{enc}} = \epsilon_0 \Phi = -8.85 \times 10^{-10} \text{ C}.$$

Exercise:

Problem:

Two equal and opposite charges of magnitude Q are located on the x -axis at the points $+a$ and $-a$, as shown below. What is the net flux due to these charges through a square surface of side $2a$ that lies in the yz -plane and is centered at the origin? (*Hint: Determine the flux due to each charge separately, then use the principle of superposition. You may be able to make a symmetry argument.*)



Exercise:

Problem:

A fellow student calculated the flux through the square for the system in the preceding problem and got 0. What went wrong?

Solution:

didn't keep consistent directions for the area vectors, or the electric fields

Exercise:

Problem:

A $10\text{ cm} \times 10\text{ cm}$ piece of aluminum foil of 0.1 mm thickness has a charge of $20\text{ }\mu\text{C}$ that spreads on both wide side surfaces evenly. You may ignore the charges on the thin sides of the edges. (a) Find the charge density. (b) Find the electric field 1 cm from the center, assuming approximate planar symmetry.

Exercise:

Problem:

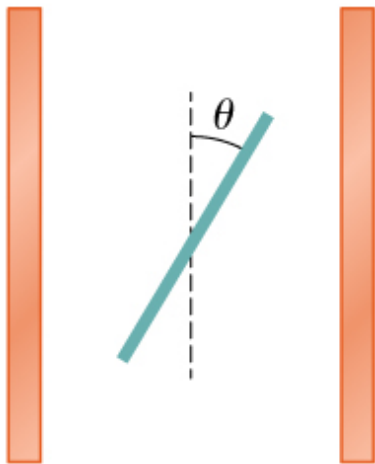
Two $10\text{ cm} \times 10\text{ cm}$ pieces of aluminum foil of thickness 0.1 mm face each other with a separation of 5 mm . One of the foils has a charge of $+30\text{ }\mu\text{C}$ and the other has $-30\text{ }\mu\text{C}$. (a) Find the charge density at all surfaces, i.e., on those facing each other and those facing away. (b) Find the electric field between the plates near the center assuming planar symmetry.

Solution:

a. $\sigma = 3.0 \times 10^{-3}\text{ C/m}^2$, $+3 \times 10^{-3}\text{ C/m}^2$ on one and $-3 \times 10^{-3}\text{ C/m}^2$ on the other; b. $E = 3.39 \times 10^8\text{ N/C}$

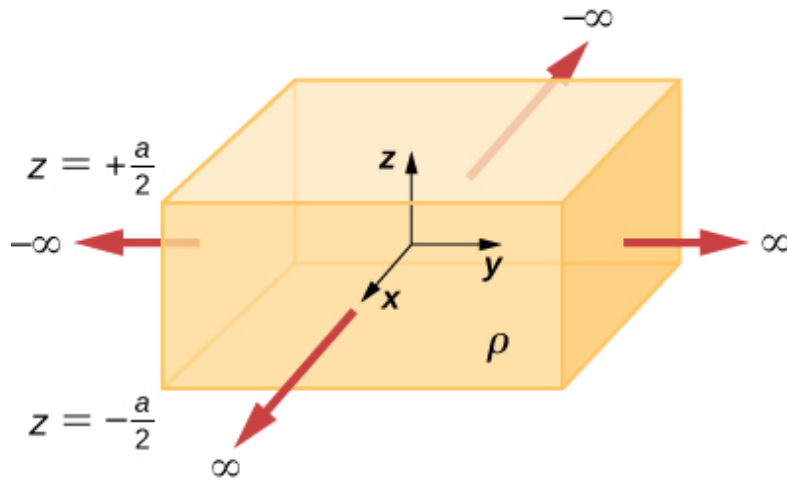
Exercise:**Problem:**

Two large copper plates facing each other have charge densities $\pm 4.0\text{ C/m}^2$ on the surface facing the other plate, and zero in between the plates. Find the electric flux through a $3\text{ cm} \times 4\text{ cm}$ rectangular area between the plates, as shown below, for the following orientations of the area. (a) If the area is parallel to the plates, and (b) if the area is tilted $\theta = 30^\circ$ from the parallel direction. Note, this angle can also be $\theta = 180^\circ + 30^\circ$.

**Exercise:**

Problem:

The infinite slab between the planes defined by $z = -a/2$ and $z = a/2$ contains a uniform volume charge density ρ (see below). What is the electric field produced by this charge distribution, both inside and outside the distribution?

**Solution:**

Construct a Gaussian cylinder along the z -axis with cross-sectional area A .

$$|z| \geq \frac{a}{2} \quad q_{\text{enc}} = \rho Aa, \quad \Phi = \frac{\rho Aa}{\epsilon_0} \Rightarrow E = \frac{\rho a}{2\epsilon_0},$$

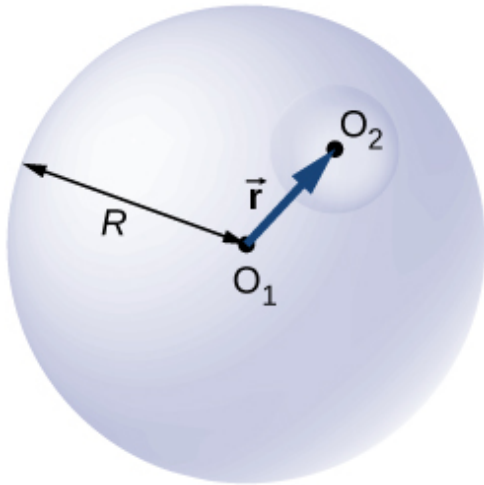
$$|z| \leq \frac{a}{2} \quad q_{\text{enc}} = \rho A2z, \quad E(2A) = \frac{\rho A2z}{\epsilon_0} \Rightarrow E = \frac{\rho z}{\epsilon_0}$$

Exercise:**Problem:**

A total charge Q is distributed uniformly throughout a spherical volume that is centered at O_1 and has a radius R . Without disturbing the charge remaining, charge is removed from the spherical volume that is centered at O_2 (see below). Show that the electric field everywhere in the empty region is given by

$$\vec{E} = \frac{Q\vec{r}}{4\pi\epsilon_0 R^3},$$

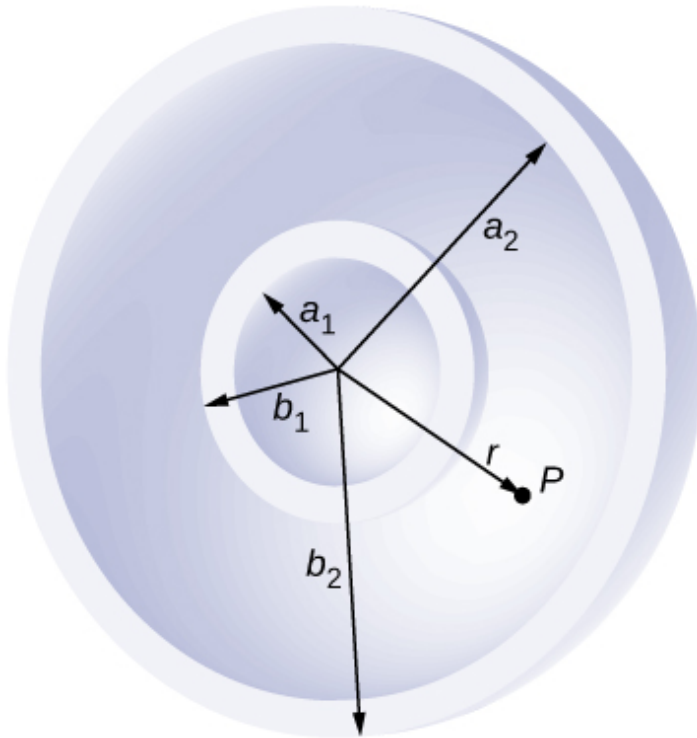
where \vec{r} is the displacement vector directed from O_1 to O_2 .



Exercise:

Problem:

A non-conducting spherical shell of inner radius a_1 and outer radius b_1 is uniformly charged with charged density ρ_1 inside another non-conducting spherical shell of inner radius a_2 and outer radius b_2 that is also uniformly charged with charge density ρ_2 . See below. Find the electric field at space point P at a distance r from the common center such that (a) $r > b_2$, (b) $a_2 < r < b_2$, (c) $b_1 < r < a_2$, (d) $a_1 < r < b_1$, and (e) $r < a_1$.



Solution:

a. $r > b_2$ $E4\pi r^2 = \frac{\frac{4}{3}\pi[\rho_1(b_1^3 - a_1^3) + \rho_2(b_2^3 - a_2^3)]}{\epsilon_0} \Rightarrow E = \frac{\rho_1(b_1^3 - a_1^3) + \rho_2(b_2^3 - a_2^3)}{3\epsilon_0 r^2}$;

b.

$a_2 < r < b_2$ $E4\pi r^2 = \frac{\frac{4}{3}\pi[\rho_1(b_1^3 - a_1^3) + \rho_2(r^3 - a_2^3)]}{\epsilon_0} \Rightarrow E = \frac{\rho_1(b_1^3 - a_1^3) + \rho_2(r^3 - a_2^3)}{3\epsilon_0 r^2}$;

;

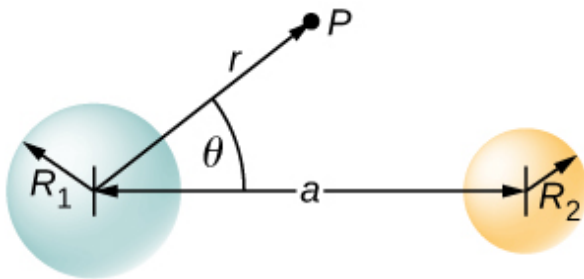
c. $b_1 < r < a_2$ $E4\pi r^2 = \frac{\frac{4}{3}\pi\rho_1(b_1^3 - a_1^3)}{\epsilon_0} \Rightarrow E = \frac{\rho_1(b_1^3 - a_1^3)}{3\epsilon_0 r^2}$;

d. $a_1 < r < b_1$ $E4\pi r^2 = \frac{\frac{4}{3}\pi\rho_1(r^3 - a_1^3)}{\epsilon_0} \Rightarrow E = \frac{\rho_1(r^3 - a_1^3)}{3\epsilon_0 r^2}$; e. 0

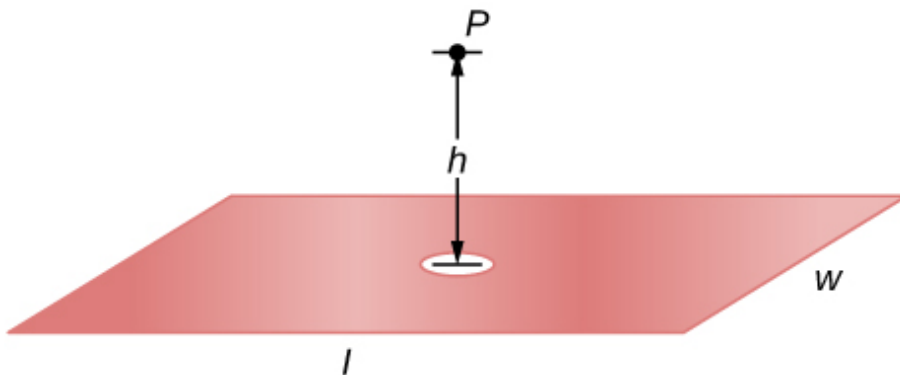
Exercise:

Problem:

Two non-conducting spheres of radii R_1 and R_2 are uniformly charged with charge densities ρ_1 and ρ_2 , respectively. They are separated at center-to-center distance a (see below). Find the electric field at point P located at a distance r from the center of sphere 1 and is in the direction θ from the line joining the two spheres assuming their charge densities are not affected by the presence of the other sphere. (*Hint: Work one sphere at a time and use the superposition principle.*)

**Exercise:****Problem:**

A disk of radius R is cut in a non-conducting large plate that is uniformly charged with charge density σ (coulomb per square meter). See below. Find the electric field at a height h above the center of the disk. ($h \gg R, h \ll l$ or w). (*Hint: Fill the hole with $\pm\sigma$.*)



Solution:

Electric field due to plate without hole: $E = \frac{\sigma}{2\epsilon_0}$.

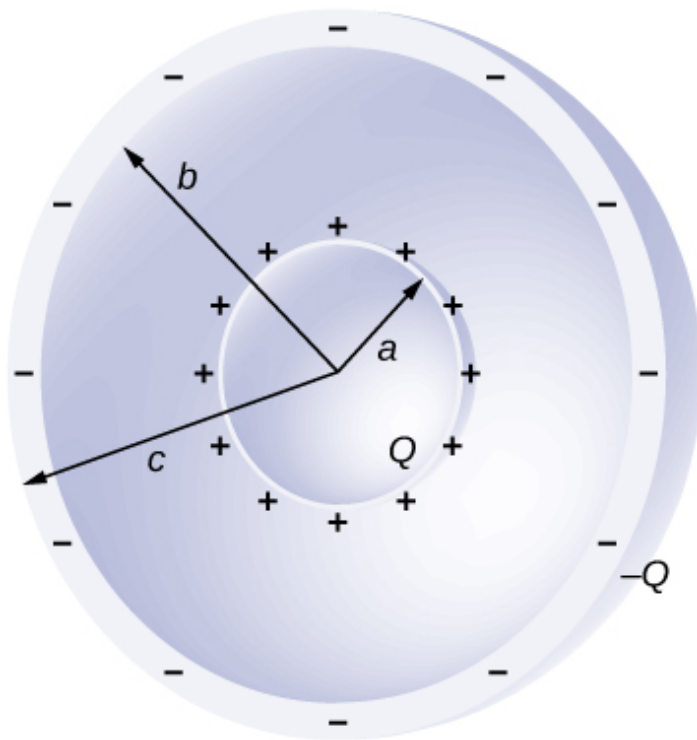
Electric field of just hole filled with $-\sigma$ $E = \frac{-\sigma}{2\epsilon_0} \left(1 - \frac{z}{\sqrt{R^2+z^2}}\right)$.

Thus, $E_{\text{net}} = \frac{\sigma}{2\epsilon_0} \frac{h}{\sqrt{R^2+h^2}}$.

Exercise:

Problem:

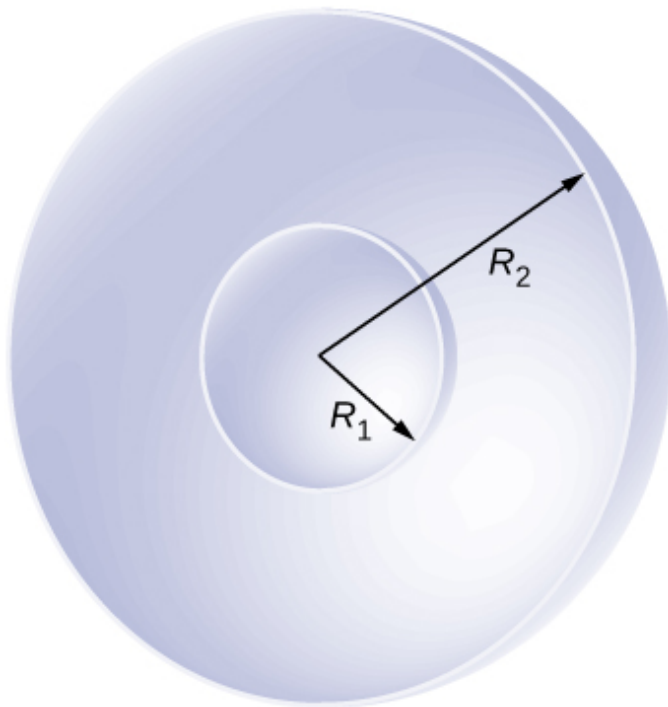
Concentric conducting spherical shells carry charges Q and $-Q$, respectively (see below). The inner shell has negligible thickness. Determine the electric field for (a) $r < a$; (b) $a < r < b$; (c) $b < r < c$; and (d) $r > c$.



Exercise:

Problem:

Shown below are two concentric conducting spherical shells of radii R_1 and R_2 , each of finite thickness much less than either radius. The inner and outer shell carry net charges q_1 and q_2 , respectively, where both q_1 and q_2 are positive. What is the electric field for (a) $r < R_1$; (b) $R_1 < r < R_2$; and (c) $r > R_2$? (d) What is the net charge on the inner surface of the inner shell, the outer surface of the inner shell, the inner surface of the outer shell, and the outer surface of the outer shell?



Solution:

a. $E = 0$; b. $E = \frac{q_1}{4\pi\epsilon_0 r^2}$; c. $E = \frac{q_1 + q_2}{4\pi\epsilon_0 r^2}$; d. 0, q_1 , $-q_1$, $q_1 + q_2$

Exercise:

Problem:

A point charge of $q = 5.0 \times 10^{-8} \text{ C}$ is placed at the center of an uncharged spherical conducting shell of inner radius 6.0 cm and outer radius 9.0 cm. Find the electric field at (a) $r = 4.0 \text{ cm}$, (b) $r = 8.0 \text{ cm}$, and (c) $r = 12.0 \text{ cm}$. (d) What are the charges induced on the inner and outer surfaces of the shell?

Challenge Problems**Exercise:****Problem:**

The Hubble Space Telescope can measure the energy flux from distant objects such as supernovae and stars. Scientists then use this data to calculate the energy emitted by that object. Choose an interstellar object which scientists have observed the flux at the Hubble with (for example, Vega^[footnote]), find the distance to that object and the size of Hubble's primary mirror, and calculate the total energy flux. (*Hint:* The Hubble intercepts only a small part of the total flux.)

<http://adsabs.harvard.edu/abs/2004AJ....127.3508B>

Solution:

Given the referenced link, using a distance to Vega of $237 \times 10^{15} \text{ m}$ ^[footnote] and a diameter of 2.4 m for the primary mirror,^[footnote] we find that at a wavelength of 555.6 nm, Vega is emitting $2.44 \times 10^{24} \text{ J/s}$ at that wavelength. Note that the flux through the mirror is essentially constant.

<http://webviz.u-strasbg.fr/viz-bin/VizieR-5?-source=I/311&HIP=91262>

<http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19910003124.pdf>

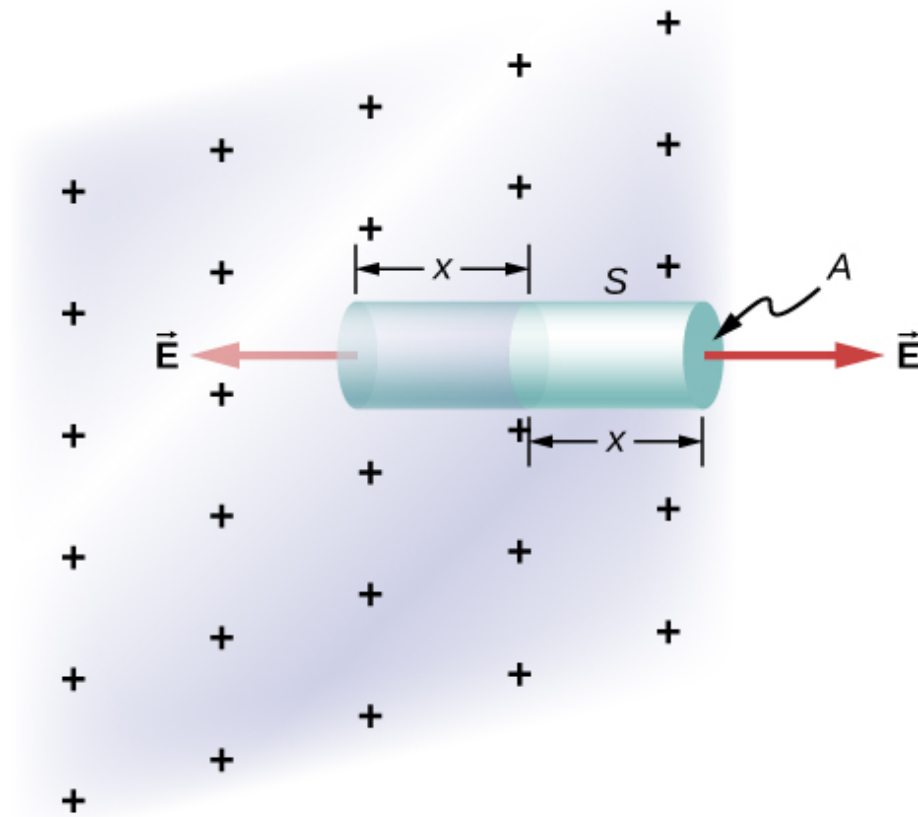
Exercise:

Problem:

Re-derive Gauss's law for the gravitational field, with \vec{g} directed positively outward.

Exercise:**Problem:**

An infinite plate sheet of charge of surface charge density σ is shown below. What is the electric field at a distance x from the sheet? Compare the result of this calculation with that of worked out in the text.



Solution:

The symmetry of the system forces \vec{E} to be perpendicular to the sheet and constant over any plane parallel to the sheet. To calculate the

electric field, we choose the cylindrical Gaussian surface shown. The cross-section area and the height of the cylinder are A and $2x$, respectively, and the cylinder is positioned so that it is bisected by the plane sheet. Since E is perpendicular to each end and parallel to the side of the cylinder, we have EA as the flux through each end and there is no flux through the side. The charge enclosed by the cylinder is σA , so from Gauss's law, $2EA = \frac{\sigma A}{\epsilon_0}$, and the electric field of an infinite sheet of charge is $E = \frac{\sigma}{2\epsilon_0}$, in agreement with the calculation of in the text.

Exercise:

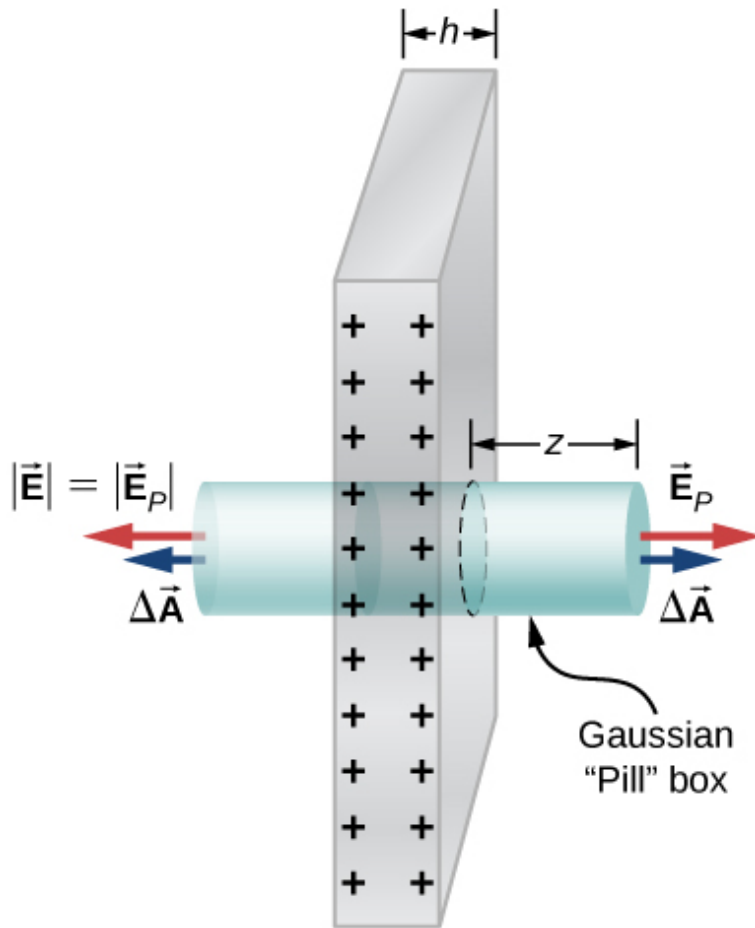
Problem:

A spherical rubber balloon carries a total charge Q distributed uniformly over its surface. At $t = 0$, the radius of the balloon is R . The balloon is then slowly inflated until its radius reaches $2R$ at the time t_0 . Determine the electric field due to this charge as a function of time (a) at the surface of the balloon, (b) at the surface of radius R , and (c) at the surface of radius $2R$. Ignore any effect on the electric field due to the material of the balloon and assume that the radius increases uniformly with time.

Exercise:

Problem:

Find the electric field of a large conducting plate containing a net charge q . Let A be area of one side of the plate and h the thickness of the plate (see below). The charge on the metal plate will distribute mostly on the two planar sides and very little on the edges if the plate is thin.



Solution:

There is $Q/2$ on each side of the plate since the net charge is Q : $\sigma = \frac{Q}{2A}$

$$\oint_S \vec{E} \cdot \hat{n} dA = \frac{2\sigma\Delta A}{\epsilon_0} \Rightarrow E_P = \frac{\sigma}{\epsilon_0} = \frac{Q}{\epsilon_0 2A}$$

Glossary

free electrons

also called conduction electrons, these are the electrons in a conductor that are not bound to any particular atom, and hence are free to move around

Introduction

class="introduction"

The energy released in a lightning strike is an excellent illustration of the vast quantities of energy that may be stored and released by an electric potential difference.

In this chapter, we calculate just how much energy can be released in a lightning strike and how this varies with the height of the clouds from the ground.
(credit: modification of work

by Anthony
Quintano)



In [Electric Charges and Fields](#), we just scratched the surface (or at least rubbed it) of electrical phenomena. Two terms commonly used to describe electricity are its energy and *voltage*, which we show in this chapter is directly related to the potential energy in a system.

We know, for example, that great amounts of electrical energy can be stored in batteries, are transmitted cross-country via currents through power lines, and may jump from clouds to explode the sap of trees. In a similar manner, at the molecular level, ions cross cell membranes and transfer information.

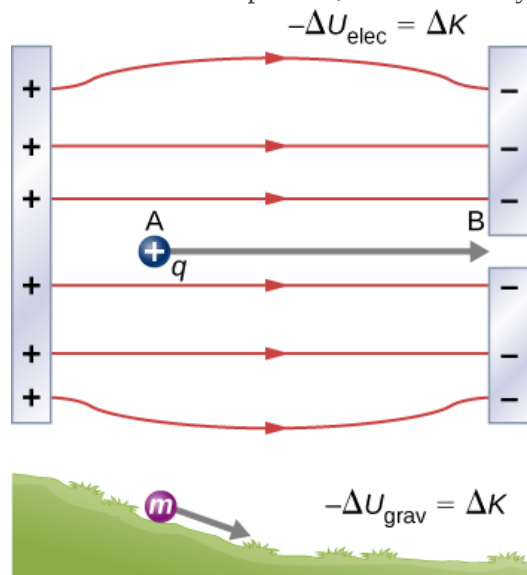
We also know about voltages associated with electricity. Batteries are typically a few volts, the outlets in your home frequently produce 120 volts, and power lines can be as high as hundreds of thousands of volts. But energy and voltage are not the same thing. A motorcycle battery, for example, is small and would not be very successful in replacing a much larger car battery, yet each has the same voltage. In this chapter, we examine the relationship between voltage and electrical energy, and begin to explore some of the many applications of electricity.

Electric Potential Energy

By the end of this section, you will be able to:

- Define the work done by an electric force
- Define electric potential energy
- Apply work and potential energy in systems with electric charges

When a free positive charge q is accelerated by an electric field, it is given kinetic energy ([link](#)). The process is analogous to an object being accelerated by a gravitational field, as if the charge were going down an electrical hill where its electric potential energy is converted into kinetic energy, although of course the sources of the forces are very different. Let us explore the work done on a charge q by the electric field in this process, so that we may develop a definition of electric potential energy.



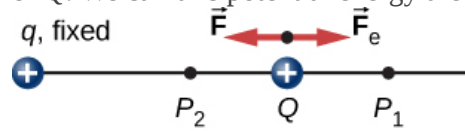
A charge accelerated by an electric field is analogous to a mass going down a hill. In both cases, potential energy decreases as kinetic energy increases, $-\Delta U = \Delta K$. Work is done by a force, but since this force is conservative, we can write

$$W = -\Delta U.$$

The electrostatic or Coulomb force is conservative, which means that the work done on q is independent of the path taken, as we will demonstrate later. This is exactly analogous to the gravitational force. When a force is conservative, it is possible to define a potential energy associated with the force. It is usually easier to work with the potential energy (because it depends only on position) than to calculate the work directly.

To show this explicitly, consider an electric charge $+q$ fixed at the origin and move another charge $+Q$ toward q in such a manner that, at each instant, the applied force \vec{F} exactly balances the electric force

\vec{F}_e on Q ([link](#)). The work done by the applied force \vec{F} on the charge Q changes the potential energy of Q . We call this potential energy the **electrical potential energy** of Q .



Displacement of “test” charge Q
in the presence of fixed “source”
charge q .

The work W_{12} done by the applied force \vec{F} when the particle moves from P_1 to P_2 may be calculated by

Equation:

$$W_{12} = \int_{P_1}^{P_2} \vec{F} \cdot d\vec{l}.$$

Since the applied force \vec{F} balances the electric force \vec{F}_e on Q , the two forces have equal magnitude and opposite directions. Therefore, the applied force is

Equation:

$$\vec{F} = -\vec{F}_e = -\frac{kqQ}{r^2} \hat{r},$$

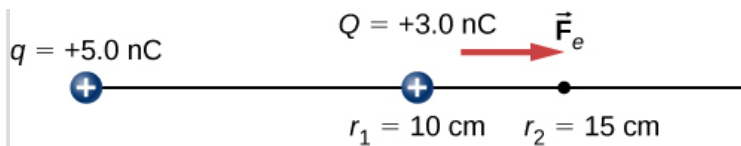
where we have defined positive to be pointing away from the origin and r is the distance from the origin. The directions of both the displacement and the applied force in the system in [link](#) are parallel, and thus the work done on the system is positive.

We use the letter U to denote electric potential energy, which has units of joules (J). When a conservative force does negative work, the system gains potential energy. When a conservative force does positive work, the system loses potential energy, $\Delta U = -W$. In the system in [link](#), the Coulomb force acts in the opposite direction to the displacement; therefore, the work is negative. However, we have increased the potential energy in the two-charge system.

Example:

Kinetic Energy of a Charged Particle

A $+3.0\text{-nC}$ charge Q is initially at rest a distance of 10 cm (r_1) from a $+5.0\text{-nC}$ charge q fixed at the origin ([link](#)). Naturally, the Coulomb force accelerates Q away from q , eventually reaching 15 cm (r_2).



The charge Q is repelled by q , thus having work done on it and gaining kinetic energy.

- What is the work done by the electric field between r_1 and r_2 ?
- How much kinetic energy does Q have at r_2 ?

Strategy

Calculate the work with the usual definition. Since Q started from rest, this is the same as the kinetic energy.

Solution

Integrating force over distance, we obtain

Equation:

$$\begin{aligned}
 W_{12} &= \int_{r_1}^{r_2} \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}} = \int_{r_1}^{r_2} \frac{kqQ}{r^2} dr = \left[-\frac{kqQ}{r} \right]_{r_1}^{r_2} = kqQ \left[\frac{-1}{r_2} + \frac{1}{r_1} \right] \\
 &= (8.99 \times 10^9 \text{ Nm}^2/\text{C}^2) (5.0 \times 10^{-9} \text{ C}) (3.0 \times 10^{-9} \text{ C}) \left[\frac{-1}{0.15 \text{ m}} + \frac{1}{0.10 \text{ m}} \right] \\
 &= 4.5 \times 10^{-7} \text{ J}.
 \end{aligned}$$

This is also the value of the kinetic energy at r_2 .

Significance

Charge Q was initially at rest; the electric field of q did work on Q , so now Q has kinetic energy equal to the work done by the electric field.

Note:

Exercise:

Problem: Check Your Understanding If Q has a mass of $4.00 \mu\text{g}$, what is the speed of Q at r_2 ?

Solution:

$$K = \frac{1}{2} mv^2, v = \sqrt{2 \frac{K}{m}} = \sqrt{2 \frac{4.5 \times 10^{-7} \text{ J}}{4.00 \times 10^{-9} \text{ kg}}} = 15 \text{ m/s}$$

In this example, the work W done to accelerate a positive charge from rest is positive and results from a loss in U , or a negative ΔU . A value for U can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Note:**Electric Potential Energy**

Work W done to accelerate a positive charge from rest is positive and results from a loss in U , or a negative ΔU . Mathematically,

Equation:

$$W = -\Delta U.$$

Gravitational potential energy and electric potential energy are quite analogous. Potential energy accounts for work done by a conservative force and gives added insight regarding energy and energy transformation without the necessity of dealing with the force directly. It is much more common, for example, to use the concept of electric potential energy than to deal with the Coulomb force directly in real-world applications.

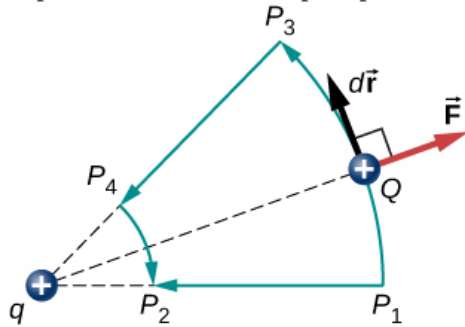
In polar coordinates with q at the origin and Q located at r , the displacement element vector is

$d\vec{l} = \hat{r} dr$ and thus the work becomes

Equation:

$$W_{12} = kqQ \int_{r_1}^{r_2} \frac{1}{r^2} \hat{r} \cdot \hat{r} dr = -kqQ \frac{1}{r_2} + kqQ \frac{1}{r_1}.$$

Notice that this result only depends on the endpoints and is otherwise independent of the path taken. To explore this further, compare path P_1 to P_2 with path $P_1P_3P_4P_2$ in [\[link\]](#).



Two paths for displacement P_1 to P_2 . The work on segments P_1P_3 and P_4P_2 are zero due to the electrical force being perpendicular to the displacement along these paths. Therefore, work on paths P_1P_2 and $P_1P_3P_4P_2$ are equal.

The segments P_1P_3 and P_4P_2 are arcs of circles centered at q . Since the force on Q points either toward or away from q , no work is done by a force balancing the electric force, because it is

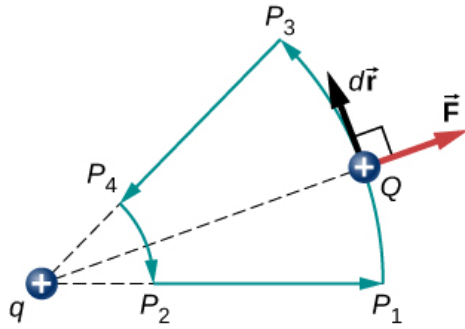
perpendicular to the displacement along these arcs. Therefore, the only work done is along segment P_3P_4 , which is identical to P_1P_2 .

One implication of this work calculation is that if we were to go around the path $P_1P_3P_4P_2P_1$, the net work would be zero ([link](#)). Recall that this is how we determine whether a force is conservative or not. Hence, because the electric force is related to the electric field by $\vec{F} = q\vec{E}$, the electric field is itself conservative. That is,

Equation:

$$\oint \vec{E} \cdot d\vec{l} = 0.$$

Note that Q is a constant.



A closed path in an electric field.
The net work around this path is zero.

Another implication is that we may define an electric potential energy. Recall that the work done by a conservative force is also expressed as the difference in the potential energy corresponding to that force. Therefore, the work W_{ref} to bring a charge from a reference point to a point of interest may be written as

Equation:

$$W_{\text{ref}} = \int_{r_{\text{ref}}}^r \vec{F} \cdot d\vec{l}$$

and, by [link](#), the difference in potential energy ($U_2 - U_1$) of the test charge Q between the two points is

Equation:

$$\Delta U = - \int_{r_{\text{ref}}}^r \vec{F} \cdot d\vec{l}.$$

Therefore, we can write a general expression for the potential energy of two point charges (in spherical coordinates):

Equation:

$$\Delta U = - \int_{r_{\text{ref}}}^r \frac{kqQ}{r^2} dr = - \left[-\frac{kqQ}{r} \right]_{r_{\text{ref}}}^r = kqQ \left[\frac{1}{r} - \frac{1}{r_{\text{ref}}} \right].$$

We may take the second term to be an arbitrary constant reference level, which serves as the zero reference:

Equation:

$$U(r) = k \frac{qQ}{r} - U_{\text{ref}}.$$

A convenient choice of reference that relies on our common sense is that when the two charges are infinitely far apart, there is no interaction between them. (Recall the discussion of reference potential energy in [Potential Energy and Conservation of Energy](#).) Taking the potential energy of this state to be zero removes the term U_{ref} from the equation (just like when we say the ground is zero potential energy in a gravitational potential energy problem), and the potential energy of Q when it is separated from q by a distance r assumes the form

Note:

Equation:

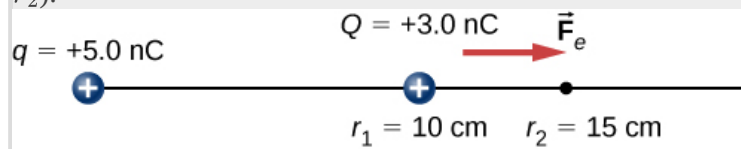
$$U(r) = k \frac{qQ}{r} \text{ (zero reference at } r = \infty \text{)}.$$

This formula is symmetrical with respect to q and Q , so it is best described as the potential energy of the two-charge system.

Example:

Potential Energy of a Charged Particle

A $+3.0\text{-nC}$ charge Q is initially at rest a distance of 10 cm (r_1) from a $+5.0\text{-nC}$ charge q fixed at the origin ([\[link\]](#)). Naturally, the Coulomb force accelerates Q away from q , eventually reaching 15 cm (r_2).



The charge Q is repelled by q , thus having work done on it and losing potential energy.

What is the change in the potential energy of the two-charge system from r_1 to r_2 ?

Strategy

Calculate the potential energy with the definition given above: $\Delta U_{12} = - \int_{r_1}^{r_2} \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}}$. Since Q

started from rest, this is the same as the kinetic energy.

Solution

We have

Equation:

$$\begin{aligned}\Delta U_{12} &= - \int_{r_1}^{r_2} \vec{\mathbf{F}} \cdot d\vec{\mathbf{r}} = - \int_{r_1}^{r_2} \frac{kqQ}{r^2} dr = - \left[-\frac{kqQ}{r} \right]_{r_1}^{r_2} = kqQ \left[\frac{1}{r_2} - \frac{1}{r_1} \right] \\ &= \left(8.99 \times 10^9 \text{ Nm}^2/\text{C}^2 \right) (5.0 \times 10^{-9} \text{ C}) (3.0 \times 10^{-9} \text{ C}) \left[\frac{1}{0.15 \text{ m}} - \frac{1}{0.10 \text{ m}} \right] \\ &= -4.5 \times 10^{-7} \text{ J}.\end{aligned}$$

Significance

The change in the potential energy is negative, as expected, and equal in magnitude to the change in kinetic energy in this system. Recall from [\[link\]](#) that the change in kinetic energy was positive.

Note:

Exercise:

Problem:

Check Your Understanding What is the potential energy of Q relative to the zero reference at infinity at r_2 in the above example?

Solution:

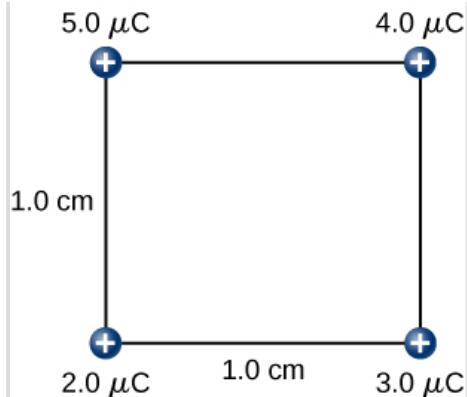
It has kinetic energy of $4.5 \times 10^{-7} \text{ J}$ at point r_2 and potential energy of $9.0 \times 10^{-7} \text{ J}$, which means that as Q approaches infinity, its kinetic energy totals three times the kinetic energy at r_2 , since all of the potential energy gets converted to kinetic.

Due to Coulomb's law, the forces due to multiple charges on a test charge Q superimpose; they may be calculated individually and then added. This implies that the work integrals and hence the resulting potential energies exhibit the same behavior. To demonstrate this, we consider an example of assembling a system of four charges.

Example:

Assembling Four Positive Charges

Find the amount of work an external agent must do in assembling four charges $+2.0 \mu\text{C}$, $+3.0 \mu\text{C}$, $+4.0 \mu\text{C}$, and $+5.0 \mu\text{C}$ at the vertices of a square of side 1.0 cm , starting each charge from infinity ([\[link\]](#)).



How much work is needed to assemble this charge configuration?

Strategy

We bring in the charges one at a time, giving them starting locations at infinity and calculating the work to bring them in from infinity to their final location. We do this in order of increasing charge.

Solution

Step 1. First bring the $+2.0\text{-}\mu\text{C}$ charge to the origin. Since there are no other charges at a finite distance from this charge yet, no work is done in bringing it from infinity,

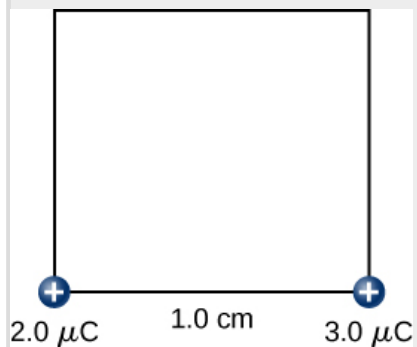
Equation:

$$W_1 = 0.$$

Step 2. While keeping the $+2.0\text{-}\mu\text{C}$ charge fixed at the origin, bring the $+3.0\text{-}\mu\text{C}$ charge to $(x, y, z) = (1.0\text{ cm}, 0, 0)$ ([link](#)). Now, the applied force must do work against the force exerted by the $+2.0\text{-}\mu\text{C}$ charge fixed at the origin. The work done equals the change in the potential energy of the $+3.0\text{-}\mu\text{C}$ charge:

Equation:

$$W_2 = k \frac{q_1 q_2}{r_{12}} = \left(9.0 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \right) \frac{(2.0 \times 10^{-6} \text{ C}) (3.0 \times 10^{-6} \text{ C})}{1.0 \times 10^{-2} \text{ m}} = 5.4 \text{ J}.$$



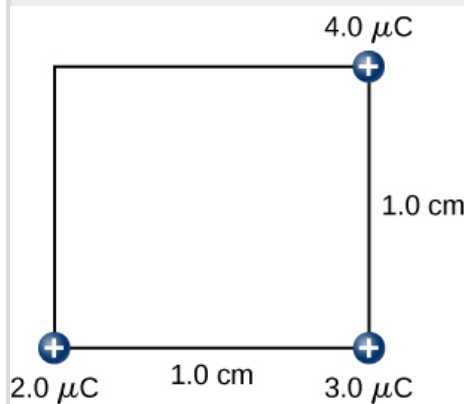
Step 2. Work W_2 to bring the $+3.0\text{-}\mu\text{C}$ charge from

infinity.

Step 3. While keeping the charges of $+2.0\ \mu\text{C}$ and $+3.0\ \mu\text{C}$ fixed in their places, bring in the $+4.0\text{-}\mu\text{C}$ charge to $(x, y, z) = (1.0\ \text{cm}, 1.0\ \text{cm}, 0)$ ([link](#)). The work done in this step is

Equation:

$$\begin{aligned} W_3 &= k \frac{q_1 q_3}{r_{13}} + k \frac{q_2 q_3}{r_{23}} \\ &= \left(9.0 \times 10^9 \frac{\text{N}\cdot\text{m}^2}{\text{C}^2} \right) \left[\frac{(2.0 \times 10^{-6}\ \text{C})(4.0 \times 10^{-6}\ \text{C})}{\sqrt{2} \times 10^{-2}\ \text{m}} + \frac{(3.0 \times 10^{-6}\ \text{C})(4.0 \times 10^{-6}\ \text{C})}{1.0 \times 10^{-2}\ \text{m}} \right] = 15.9\ \text{J}. \end{aligned}$$

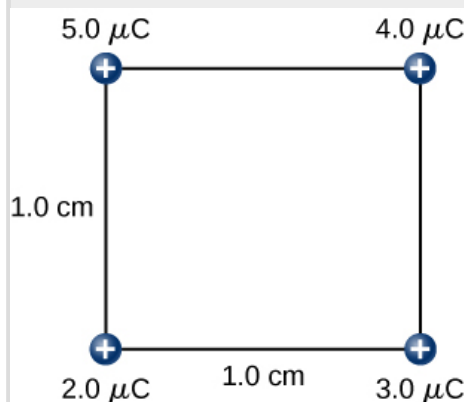


Step 3. The work W_3 to bring the $+4.0\text{-}\mu\text{C}$ charge from infinity.

Step 4. Finally, while keeping the first three charges in their places, bring the $+5.0\text{-}\mu\text{C}$ charge to $(x, y, z) = (0, 1.0\ \text{cm}, 0)$ ([link](#)). The work done here is

Equation:

$$\begin{aligned} W_4 &= k q_4 \left[\frac{q_1}{r_{14}} + \frac{q_2}{r_{24}} + \frac{q_3}{r_{34}} \right], \\ &= \left(9.0 \times 10^9 \frac{\text{N}\cdot\text{m}^2}{\text{C}^2} \right) (5.0 \times 10^{-6}\ \text{C}) \left[\frac{(2.0 \times 10^{-6}\ \text{C})}{1.0 \times 10^{-2}\ \text{m}} + \frac{(3.0 \times 10^{-6}\ \text{C})}{\sqrt{2} \times 10^{-2}\ \text{m}} + \frac{(4.0 \times 10^{-6}\ \text{C})}{1.0 \times 10^{-2}\ \text{m}} \right] = 36.5\ \text{J}. \end{aligned}$$



Step 4. The work W_4 to bring the $+5.0\text{-}\mu\text{C}$ charge from infinity.

Hence, the total work done by the applied force in assembling the four charges is equal to the sum of the work in bringing each charge from infinity to its final position:

Equation:

$$W_T = W_1 + W_2 + W_3 + W_4 = 0 + 5.4 \text{ J} + 15.9 \text{ J} + 36.5 \text{ J} = 57.8 \text{ J}.$$

Significance

The work on each charge depends only on its pairwise interactions with the other charges. No more complicated interactions need to be considered; the work on the third charge only depends on its interaction with the first and second charges, the interaction between the first and second charge does not affect the third.

Note:

Exercise:

Problem:

Check Your Understanding Is the electrical potential energy of two point charges positive or negative if the charges are of the same sign? Opposite signs? How does this relate to the work necessary to bring the charges into proximity from infinity?

Solution:

positive, negative, and these quantities are the same as the work you would need to do to bring the charges in from infinity

Note that the electrical potential energy is positive if the two charges are of the same type, either positive or negative, and negative if the two charges are of opposite types. This makes sense if you think of the change in the potential energy ΔU as you bring the two charges closer or move them farther apart. Depending on the relative types of charges, you may have to work on the system or the system would do work on you, that is, your work is either positive or negative. If you have to do positive work on the system (actually push the charges closer), then the energy of the system should increase. If you bring two positive charges or two negative charges closer, you have to do positive work on the system, which raises their potential energy. Since potential energy is proportional to $1/r$, the potential energy goes up when r goes down between two positive or two negative charges.

On the other hand, if you bring a positive and a negative charge nearer, you have to do negative work on the system (the charges are pulling you), which means that you take energy away from the system. This reduces the potential energy. Since potential energy is negative in the case of a positive and a negative charge pair, the increase in $1/r$ makes the potential energy more negative, which is the same as a reduction in potential energy.

The result from [\[link\]](#) may be extended to systems with any arbitrary number of charges. In this case, it is most convenient to write the formula as

Note:

Equation:

$$W_{12\dots N} = \frac{k}{2} \sum_i^N \sum_j^N \frac{q_i q_j}{r_{ij}} \text{ for } i \neq j.$$

The factor of 1/2 accounts for adding each pair of charges twice.

Summary

- The work done to move a charge from point A to B in an electric field is path independent, and the work around a closed path is zero. Therefore, the electric field and electric force are conservative.
- We can define an electric potential energy, which between point charges is $U(r) = k \frac{qQ}{r}$, with the zero reference taken to be at infinity.
- The superposition principle holds for electric potential energy; the potential energy of a system of multiple charges is the sum of the potential energies of the individual pairs.

Conceptual Questions

Exercise:

Problem:

Would electric potential energy be meaningful if the electric field were not conservative?

Solution:

No. We can only define potential energies for conservative fields.

Exercise:

Problem:

Why do we need to be careful about work done *on* the system versus work done *by* the system in calculations?

Exercise:

Problem:

Does the order in which we assemble a system of point charges affect the total work done?

Solution:

No, though certain orderings may be simpler to compute.

Problems

Exercise:

Problem:

Consider a charge $Q_1 (+5.0 \mu\text{C})$ fixed at a site with another charge Q_2 (charge $+3.0 \mu\text{C}$, mass $6.0 \mu\text{g}$) moving in the neighboring space. (a) Evaluate the potential energy of Q_2 when it is 4.0 cm from Q_1 . (b) If Q_2 starts from rest from a point 4.0 cm from Q_1 , what will be its speed when it is 8.0 cm from Q_1 ? (Note: Q_1 is held fixed in its place.)

Solution:

- a. $U = 3.4 \text{ J}$;
b. $\frac{1}{2}mv^2 = Q_1Q_2 \left(\frac{1}{r_i} - \frac{1}{r_f} \right) \rightarrow v = 2.4 \times 10^4 \text{ m/s}$

Exercise:

Problem:

Two charges $Q_1 (+2.00 \mu\text{C})$ and $Q_2 (+2.00 \mu\text{C})$ are placed symmetrically along the x-axis at $x = \pm 3.00 \text{ cm}$. Consider a charge Q_3 of charge $+4.00 \mu\text{C}$ and mass 10.0 mg moving along the y-axis. If Q_3 starts from rest at $y = 2.00 \text{ cm}$, what is its speed when it reaches $y = 4.00 \text{ cm}$?

Exercise:

Problem:

To form a hydrogen atom, a proton is fixed at a point and an electron is brought from far away to a distance of $0.529 \times 10^{-10} \text{ m}$, the average distance between proton and electron in a hydrogen atom. How much work is done?

Solution:

$$U = 4.36 \times 10^{-18} \text{ J}$$

Exercise:

Problem:

- (a) What is the average power output of a heart defibrillator that dissipates 400 J of energy in 10.0 ms? (b) Considering the high-power output, why doesn't the defibrillator produce serious burns?

Glossary

electric potential energy

potential energy stored in a system of charged objects due to the charges

Electric Potential and Potential Difference

By the end of this section, you will be able to:

- Define electric potential, voltage, and potential difference
- Define the electron-volt
- Calculate electric potential and potential difference from potential energy and electric field
- Describe systems in which the electron-volt is a useful unit
- Apply conservation of energy to electric systems

Recall that earlier we defined electric field to be a quantity independent of the test charge in a given system, which would nonetheless allow us to calculate the force that would result on an arbitrary test charge. (The default assumption in the absence of other information is that the test charge is positive.) We briefly defined a field for gravity, but gravity is always attractive, whereas the electric force can be either attractive or repulsive. Therefore, although potential energy is perfectly adequate in a gravitational system, it is convenient to define a quantity that allows us to calculate the work on a charge independent of the magnitude of the charge. Calculating the work directly may be difficult, since $W = \vec{\mathbf{F}} \cdot \vec{\mathbf{d}}$ and the direction and magnitude of $\vec{\mathbf{F}}$ can be complex for multiple charges, for odd-shaped objects, and along arbitrary paths. But we do know that because $\vec{\mathbf{F}} = q\vec{\mathbf{E}}$, the work, and hence ΔU , is proportional to the test charge q . To have a physical quantity that is independent of test charge, we define **electric potential** V (or simply potential, since electric is understood) to be the potential energy per unit charge:

Note:

Electric Potential

The electric potential energy per unit charge is

Equation:

$$V = \frac{U}{q}.$$

Since U is proportional to q , the dependence on q cancels. Thus, V does not depend on q . The change in potential energy ΔU is crucial, so we are concerned with the

difference in potential or potential difference ΔV between two points, where

Equation:

$$\Delta V = V_B - V_A = \frac{\Delta U}{q}.$$

Note:

Electric Potential Difference

The **electric potential difference** between points A and B , $V_B - V_A$, is defined to be the change in potential energy of a charge q moved from A to B , divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

Equation:

$$1 \text{ V} = 1 \text{ J/C}$$

The familiar term **voltage** is the common name for electric potential difference. Keep in mind that whenever a voltage is quoted, it is understood to be the potential difference between two points. For example, every battery has two terminals, and its voltage is the potential difference between them. More fundamentally, the point you choose to be zero volts is arbitrary. This is analogous to the fact that gravitational potential energy has an arbitrary zero, such as sea level or perhaps a lecture hall floor. It is worthwhile to emphasize the distinction between potential difference and electrical potential energy.

Note:

Potential Difference and Electrical Potential Energy

The relationship between potential difference (or voltage) and electrical potential energy is given by

Equation:

$$\Delta V = \frac{\Delta U}{q} \text{ or } \Delta U = q\Delta V.$$

Voltage is not the same as energy. Voltage is the energy per unit charge. Thus, a motorcycle battery and a car battery can both have the same voltage (more precisely, the same potential difference between battery terminals), yet one stores much more energy than the other because $\Delta U = q\Delta V$. The car battery can move more charge than the motorcycle battery, although both are 12-V batteries.

Example:**Calculating Energy**

You have a 12.0-V motorcycle battery that can move 5000 C of charge, and a 12.0-V car battery that can move 60,000 C of charge. How much energy does each deliver? (Assume that the numerical value of each charge is accurate to three significant figures.)

Strategy

To say we have a 12.0-V battery means that its terminals have a 12.0-V potential difference. When such a battery moves charge, it puts the charge through a potential difference of 12.0 V, and the charge is given a change in potential energy equal to $\Delta U = q\Delta V$. To find the energy output, we multiply the charge moved by the potential difference.

Solution

For the motorcycle battery, $q = 5000 \text{ C}$ and $\Delta V = 12.0 \text{ V}$. The total energy delivered by the motorcycle battery is

Equation:

$$\Delta U_{\text{cycle}} = (5000 \text{ C})(12.0 \text{ V}) = (5000 \text{ C})(12.0 \text{ J/C}) = 6.00 \times 10^4 \text{ J}.$$

Similarly, for the car battery, $q = 60,000 \text{ C}$ and

Equation:

$$\Delta U_{\text{car}} = (60,000 \text{ C})(12.0 \text{ V}) = 7.20 \times 10^5 \text{ J}.$$

Significance

Voltage and energy are related, but they are not the same thing. The voltages of the batteries are identical, but the energy supplied by each is quite different. A car battery has a much larger engine to start than a motorcycle. Note also that as a battery is discharged, some of its energy is used internally and its terminal voltage drops, such as when headlights dim because of a depleted car battery. The energy supplied by the battery is still calculated as in this example, but not all of the energy is available for external use.

Note:

Exercise:

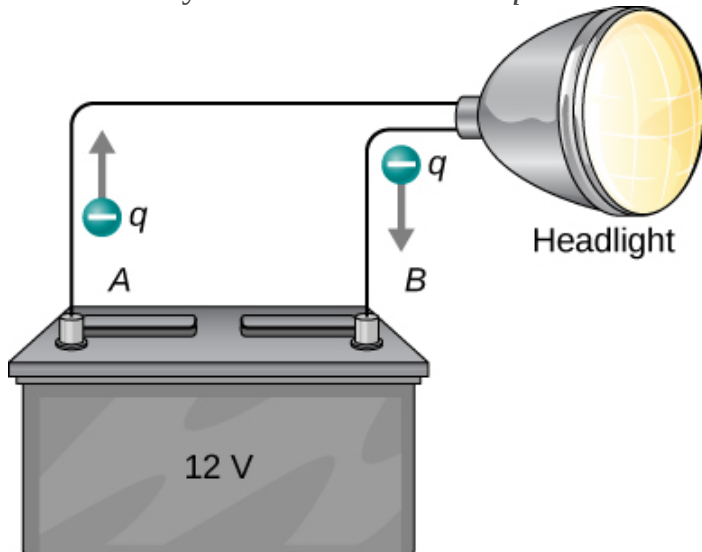
Problem:

Check Your Understanding How much energy does a 1.5-V AAA battery have that can move 100 C?

Solution:

$$\Delta U = q\Delta V = (100 \text{ C})(1.5 \text{ V}) = 150 \text{ J}$$

Note that the energies calculated in the previous example are absolute values. The change in potential energy for the battery is negative, since it loses energy. These batteries, like many electrical systems, actually move negative charge—electrons in particular. The batteries repel electrons from their negative terminals (*A*) through whatever circuitry is involved and attract them to their positive terminals (*B*), as shown in [\[link\]](#). The change in potential is $\Delta V = V_B - V_A = +12 \text{ V}$ and the charge q is negative, so that $\Delta U = q\Delta V$ is negative, meaning the potential energy of the battery has decreased when q has moved from *A* to *B*.



A battery moves negative charge from its negative terminal through a headlight to its positive terminal. Appropriate combinations of chemicals in the battery separate charges so that the negative

terminal has an excess of negative charge, which is repelled by it and attracted to the excess positive charge on the other terminal. In terms of potential, the positive terminal is at a higher voltage than the negative terminal. Inside the battery, both positive and negative charges move.

Example:**How Many Electrons Move through a Headlight Each Second?**

When a 12.0-V car battery powers a single 30.0-W headlight, how many electrons pass through it each second?

Strategy

To find the number of electrons, we must first find the charge that moves in 1.00 s. The charge moved is related to voltage and energy through the equations $\Delta U = q\Delta V$. A 30.0-W lamp uses 30.0 joules per second. Since the battery loses energy, we have $\Delta U = -30 \text{ J}$ and, since the electrons are going from the negative terminal to the positive, we see that $\Delta V = +12.0 \text{ V}$.

Solution

To find the charge q moved, we solve the equation $\Delta U = q\Delta V$:

Equation:

$$q = \frac{\Delta U}{\Delta V}.$$

Entering the values for ΔU and ΔV , we get

Equation:

$$q = \frac{-30.0 \text{ J}}{+12.0 \text{ V}} = \frac{-30.0 \text{ J}}{+12.0 \text{ J/C}} = -2.50 \text{ C}.$$

The number of electrons n_e is the total charge divided by the charge per electron. That is,

Equation:

$$n_e = \frac{-2.50 \text{ C}}{-1.60 \times 10^{-19} \text{ C/e}^-} = 1.56 \times 10^{19} \text{ electrons}.$$

Significance

This is a very large number. It is no wonder that we do not ordinarily observe individual electrons with so many being present in ordinary systems. In fact, electricity had been in use for many decades before it was determined that the moving charges in many circumstances were negative. Positive charge moving in the opposite direction of negative charge often produces identical effects; this makes it difficult to determine which is moving or whether both are moving.

Note:

Exercise:

Problem:

Check Your Understanding How many electrons would go through a 24.0-W lamp each second from a 12-volt car battery?

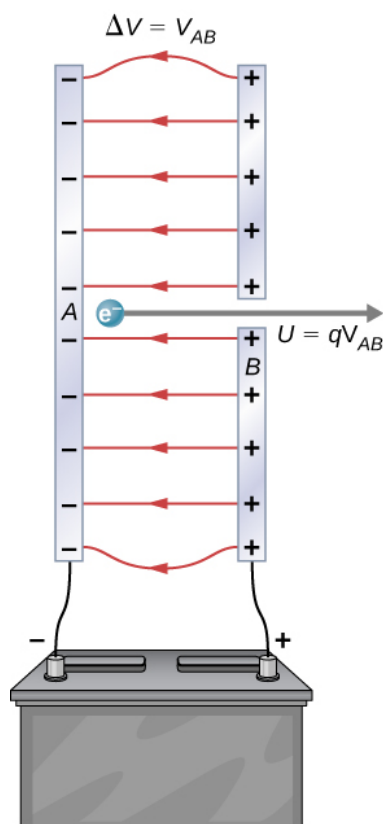
Solution:

$$-2.00 \text{ C}, n_e = 1.25 \times 10^{19} \text{ electrons}$$

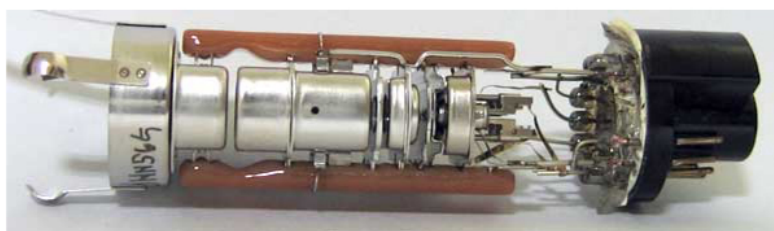
The Electron-Volt

The energy per electron is very small in macroscopic situations like that in the previous example—a tiny fraction of a joule. But on a submicroscopic scale, such energy per particle (electron, proton, or ion) can be of great importance. For example, even a tiny fraction of a joule can be great enough for these particles to destroy organic molecules and harm living tissue. The particle may do its damage by direct collision, or it may create harmful X-rays, which can also inflict damage. It is useful to have an energy unit related to submicroscopic effects.

[\[link\]](#) shows a situation related to the definition of such an energy unit. An electron is accelerated between two charged metal plates, as it might be in an old-model television tube or oscilloscope. The electron gains kinetic energy that is later converted into another form—light in the television tube, for example. (Note that in terms of energy, “downhill” for the electron is “uphill” for a positive charge.) Since energy is related to voltage by $\Delta U = q\Delta V$, we can think of the joule as a coulomb-volt.



(a)



(b)

A typical electron gun accelerates electrons using a potential difference between two separated metal plates. By conservation of energy, the kinetic energy has to equal the change in potential energy, so $KE = qV$. The energy of the electron in electron-volts is numerically the same as the voltage between the plates. For example, a 5000-V potential difference produces 5000-eV electrons. The conceptual construct, namely two parallel plates with a hole in one, is shown in (a), while a real electron gun is shown in (b).

Note:

Electron-Volt

On the submicroscopic scale, it is more convenient to define an energy unit called the **electron-volt** (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

Equation:

$$1 \text{ eV} = (1.60 \times 10^{-19} \text{ C})(1 \text{ V}) = (1.60 \times 10^{-19} \text{ C})(1 \text{ J/C}) = 1.60 \times 10^{-19} \text{ J}.$$

An electron accelerated through a potential difference of 1 V is given an energy of 1 eV. It follows that an electron accelerated through 50 V gains 50 eV. A potential difference of 100,000 V (100 kV) gives an electron an energy of 100,000 eV (100 keV), and so on. Similarly, an ion with a double positive charge accelerated through 100 V gains 200 eV of energy. These simple relationships between accelerating voltage and particle charges make the electron-volt a simple and convenient energy unit in such circumstances.

The electron-volt is commonly employed in submicroscopic processes—chemical valence energies and molecular and nuclear binding energies are among the quantities often expressed in electron-volts. For example, about 5 eV of energy is required to break up certain organic molecules. If a proton is accelerated from rest through a potential difference of 30 kV, it acquires an energy of 30 keV (30,000 eV) and can break up as many as 6000 of these molecules ($30,000 \text{ eV} \div 5 \text{ eV per molecule} = 6000 \text{ molecules}$). Nuclear decay energies are on the order of 1 MeV (1,000,000 eV) per event and can thus produce significant biological damage.

Conservation of Energy

The total energy of a system is conserved if there is no net addition (or subtraction) due to work or heat transfer. For conservative forces, such as the electrostatic force, conservation of energy states that mechanical energy is a constant.

Mechanical energy is the sum of the kinetic energy and potential energy of a system; that is, $K + U = \text{constant}$. A loss of U for a charged particle becomes an increase in its K . Conservation of energy is stated in equation form as

Equation:

$$K + U = \text{constant}$$

or

Equation:

$$K_i + U_i = K_f + U_f$$

where i and f stand for initial and final conditions. As we have found many times before, considering energy can give us insights and facilitate problem solving.

Example:

Electrical Potential Energy Converted into Kinetic Energy

Calculate the final speed of a free electron accelerated from rest through a potential difference of 100 V. (Assume that this numerical value is accurate to three significant figures.)

Strategy

We have a system with only conservative forces. Assuming the electron is accelerated in a vacuum, and neglecting the gravitational force (we will check on this assumption later), all of the electrical potential energy is converted into kinetic energy. We can identify the initial and final forms of energy to be

$$K_i = 0, K_f = \frac{1}{2}mv^2, U_i = qV, U_f = 0.$$

Solution

Conservation of energy states that

Equation:

$$K_i + U_i = K_f + U_f.$$

Entering the forms identified above, we obtain

Equation:

$$qV = \frac{mv^2}{2}.$$

We solve this for v:

Equation:

$$v = \sqrt{\frac{2qV}{m}}.$$

Entering values for q , V , and m gives

Equation:

$$v = \sqrt{\frac{2(-1.60 \times 10^{-19} \text{ C})(-100 \text{ J/C})}{9.11 \times 10^{-31} \text{ kg}}} = 5.93 \times 10^6 \text{ m/s}.$$

Significance

Note that both the charge and the initial voltage are negative, as in [\[link\]](#). From the discussion of electric charge and electric field, we know that electrostatic forces on small particles are generally very large compared with the gravitational force. The large final speed confirms that the gravitational force is indeed negligible here. The large speed also indicates how easy it is to accelerate electrons with small voltages because of their very small mass. Voltages much higher than the 100 V in this problem are typically used in electron guns. These higher voltages produce electron speeds so great that effects from special relativity must be taken into account and hence are reserved for a later chapter ([Relativity](#)). That is why we consider a low voltage (accurately) in this example.

Note:

Exercise:

Problem:

Check Your Understanding How would this example change with a positron? A positron is identical to an electron except the charge is positive.

Solution:

It would be going in the opposite direction, with no effect on the calculations as presented.

Voltage and Electric Field

So far, we have explored the relationship between voltage and energy. Now we want to explore the relationship between voltage and electric field. We will start with the general case for a non-uniform $\vec{\mathbf{E}}$ field. Recall that our general formula for the potential energy of a test charge q at point P relative to reference point R is

Equation:

$$U_P = - \int_R^P \vec{\mathbf{F}} \cdot d\vec{\mathbf{l}}.$$

When we substitute in the definition of electric field ($\vec{\mathbf{E}} = \vec{\mathbf{F}}/q$), this becomes

Equation:

$$U_P = -q \int_R^P \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}.$$

Applying our definition of potential ($V = U/q$) to this potential energy, we find that, in general,

Note:

Equation:

$$V_P = - \int_R^P \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}.$$

From our previous discussion of the potential energy of a charge in an electric field, the result is independent of the path chosen, and hence we can pick the integral path that is most convenient.

Consider the special case of a positive point charge q at the origin. To calculate the potential caused by q at a distance r from the origin relative to a reference of 0 at infinity (recall that we did the same for potential energy), let $P = r$ and $R = \infty$, with $d\vec{\mathbf{l}} = d\vec{\mathbf{r}} = \hat{\mathbf{r}}dr$ and use $\vec{\mathbf{E}} = \frac{kq}{r^2}\hat{\mathbf{r}}$. When we evaluate the integral

Equation:

$$V_P = - \int_R^P \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$$

for this system, we have

Equation:

$$V_r = - \int_{\infty}^r \frac{kq}{r^2} \hat{\mathbf{r}} \cdot \hat{\mathbf{r}} dr,$$

which simplifies to

Equation:

$$V_r = - \int_{\infty}^r \frac{kq}{r^2} dr = \frac{kq}{r} - \frac{kq}{\infty} = \frac{kq}{r}.$$

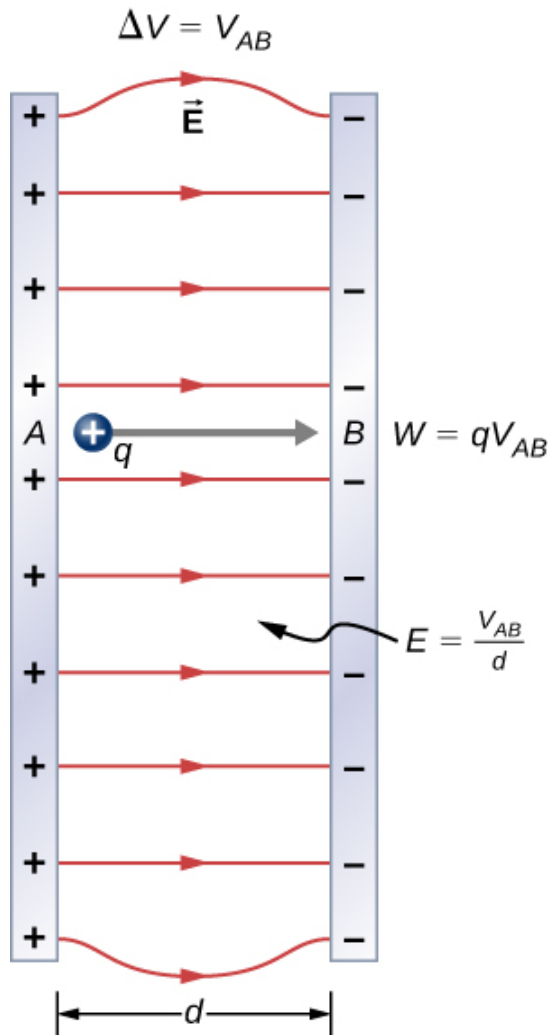
This result,

Equation:

$$V_r = \frac{kq}{r}$$

is the standard form of the potential of a point charge. This will be explored further in the next section.

To examine another interesting special case, suppose a uniform electric field $\vec{\mathbf{E}}$ is produced by placing a potential difference (or voltage) ΔV across two parallel metal plates, labeled A and B ([link](#)). Examining this situation will tell us what voltage is needed to produce a certain electric field strength. It will also reveal a more fundamental relationship between electric potential and electric field.



The relationship between V and E for parallel conducting plates is $E = V/d$. (Note that $\Delta V = V_{AB}$ in magnitude. For a charge that is moved from plate A at higher potential to plate B at lower potential, a minus sign needs to be included as follows:

$$-\Delta V = V_A - V_B = V_{AB}.)$$

From a physicist's point of view, either ΔV or \vec{E} can be used to describe any interaction between charges. However, ΔV is a scalar quantity and has no direction,

whereas $\vec{\mathbf{E}}$ is a vector quantity, having both magnitude and direction. (Note that the magnitude of the electric field, a scalar quantity, is represented by E .) The relationship between ΔV and $\vec{\mathbf{E}}$ is revealed by calculating the work done by the electric force in moving a charge from point A to point B . But, as noted earlier, arbitrary charge distributions require calculus. We therefore look at a uniform electric field as an interesting special case.

The work done by the electric field in [\[link\]](#) to move a positive charge q from A , the positive plate, higher potential, to B , the negative plate, lower potential, is

Equation:

$$W = -\Delta U = -q\Delta V.$$

The potential difference between points A and B is

Equation:

$$-\Delta V = -(V_B - V_A) = V_A - V_B = V_{AB}.$$

Entering this into the expression for work yields

Equation:

$$W = qV_{AB}.$$

Work is $W = \vec{\mathbf{F}} \cdot \vec{\mathbf{d}} = Fd \cos \theta$; here $\cos \theta = 1$, since the path is parallel to the field. Thus, $W = Fd$. Since $F = qE$, we see that $W = qEd$.

Substituting this expression for work into the previous equation gives

Equation:

$$qEd = qV_{AB}.$$

The charge cancels, so we obtain for the voltage between points A and B

Equation:

$$\left. \begin{array}{l} V_{AB} = Ed \\ E = \frac{V_{AB}}{d} \end{array} \right\} \text{(uniform } E\text{-field only)}$$

where d is the distance from A to B , or the distance between the plates in [\[link\]](#). Note that this equation implies that the units for electric field are volts per meter. We already know the units for electric field are newtons per coulomb; thus, the following relation among units is valid:

Equation:

$$1 \text{ N} / \text{C} = 1 \text{ V} / \text{m}.$$

Furthermore, we may extend this to the integral form. Substituting [\[link\]](#) into our definition for the potential difference between points A and B , we obtain

Equation:

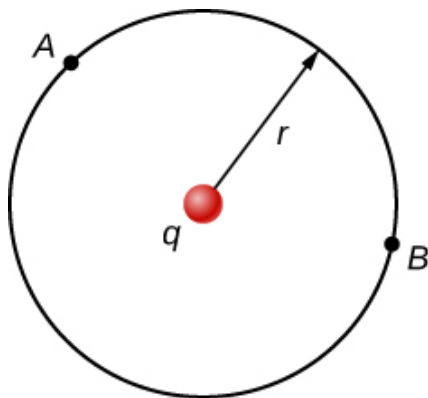
$$V_{BA} = V_B - V_A = - \int_R^B \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} + \int_R^A \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$$

which simplifies to

Equation:

$$V_B - V_A = - \int_A^B \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}.$$

As a demonstration, from this we may calculate the potential difference between two points (A and B) equidistant from a point charge q at the origin, as shown in [\[link\]](#).



The arc for calculating the potential difference between two points that

are equidistant from a point charge at the origin.

To do this, we integrate around an arc of the circle of constant radius r between A and B , which means we let $d\vec{l} = r\hat{\varphi}d\varphi$, while using $\vec{E} = \frac{kq}{r^2}\hat{r}$. Thus,

Note:

Equation:

$$\Delta V_{BA} = V_B - V_A = - \int_A^B \vec{E} \cdot d\vec{l}$$

for this system becomes

Equation:

$$V_B - V_A = - \int_A^B \frac{kq}{r^2} \hat{r} \cdot r\hat{\varphi}d\varphi.$$

However, $\hat{r} \cdot \hat{\varphi} = 0$ and therefore

Equation:

$$V_B - V_A = 0.$$

This result, that there is no difference in potential along a constant radius from a point charge, will come in handy when we map potentials.

Example:

What Is the Highest Voltage Possible between Two Plates?

Dry air can support a maximum electric field strength of about $3.0 \times 10^6 \text{ V/m}$. Above that value, the field creates enough ionization in the air to make the air a conductor. This allows a discharge or spark that reduces the field. What, then, is the maximum voltage between two parallel conducting plates separated by 2.5 cm of dry air?

Strategy

We are given the maximum electric field E between the plates and the distance d between them. We can use the equation $V_{AB} = Ed$ to calculate the maximum voltage.

Solution

The potential difference or voltage between the plates is

Equation:

$$V_{AB} = Ed.$$

Entering the given values for E and d gives

Equation:

$$V_{AB} = (3.0 \times 10^6 \text{ V/m})(0.025 \text{ m}) = 7.5 \times 10^4 \text{ V}$$

or

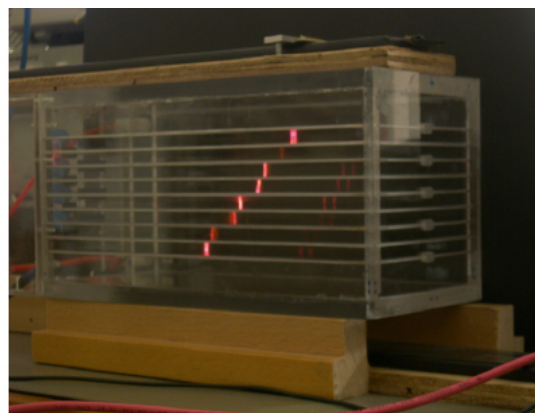
Equation:

$$V_{AB} = 75 \text{ kV}.$$

(The answer is quoted to only two digits, since the maximum field strength is approximate.)

Significance

One of the implications of this result is that it takes about 75 kV to make a spark jump across a 2.5-cm (1-in.) gap, or 150 kV for a 5-cm spark. This limits the voltages that can exist between conductors, perhaps on a power transmission line. A smaller voltage can cause a spark if there are spines on the surface, since sharp points have larger field strengths than smooth surfaces. Humid air breaks down at a lower field strength, meaning that a smaller voltage will make a spark jump through humid air. The largest voltages can be built up with static electricity on dry days ([link](#)).



A spark chamber is used to trace the paths of high-energy particles. Ionization created by the particles as they pass through the gas between the plates allows a spark to jump. The sparks are perpendicular to the plates, following electric field lines between them. The potential difference between adjacent plates is not high enough to cause sparks without the ionization produced by particles from accelerator experiments (or cosmic rays). This form of detector is now archaic and no longer in use except for demonstration purposes. (credit b: modification of work by Jack Collins)

Example:

Field and Force inside an Electron Gun

An electron gun ([link](#)) has parallel plates separated by 4.00 cm and gives electrons 25.0 keV of energy. (a) What is the electric field strength between the plates? (b) What force would this field exert on a piece of plastic with a $0.500\text{-}\mu\text{C}$ charge that gets between the plates?

Strategy

Since the voltage and plate separation are given, the electric field strength can be calculated directly from the expression $E = \frac{V_{AB}}{d}$. Once we know the electric field

strength, we can find the force on a charge by using $\vec{F} = q\vec{E}$. Since the electric field is in only one direction, we can write this equation in terms of the magnitudes, $F = qE$.

Solution

- The expression for the magnitude of the electric field between two uniform metal plates is

Equation:

$$E = \frac{V_{AB}}{d}.$$

Since the electron is a single charge and is given 25.0 keV of energy, the potential difference must be 25.0 kV. Entering this value for V_{AB} and the plate separation of 0.0400 m, we obtain

Equation:

$$E = \frac{25.0 \text{ kV}}{0.0400 \text{ m}} = 6.25 \times 10^5 \text{ V/m}.$$

- b. The magnitude of the force on a charge in an electric field is obtained from the equation

Equation:

$$F = qE.$$

Substituting known values gives

Equation:

$$F = (0.500 \times 10^{-6} \text{ C})(6.25 \times 10^5 \text{ V/m}) = 0.313 \text{ N}.$$

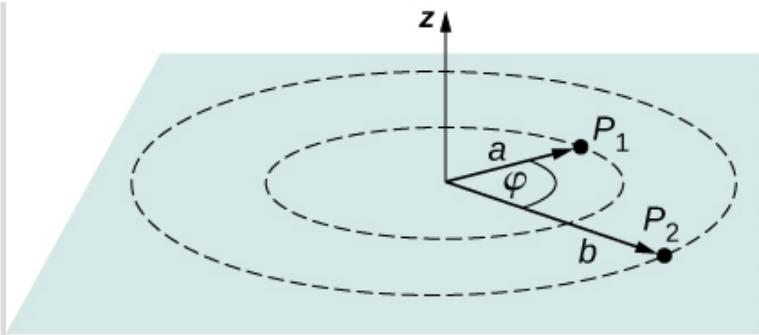
Significance

Note that the units are newtons, since $1 \text{ V/m} = 1 \text{ N/C}$. Because the electric field is uniform between the plates, the force on the charge is the same no matter where the charge is located between the plates.

Example:

Calculating Potential of a Point Charge

Given a point charge $q = +2.0 \text{ nC}$ at the origin, calculate the potential difference between point P_1 a distance $a = 4.0 \text{ cm}$ from q , and P_2 a distance $b = 12.0 \text{ cm}$ from q , where the two points have an angle of $\varphi = 24^\circ$ between them ([\[link\]](#)).



Find the difference in potential between P_1 and P_2 .

Strategy

Do this in two steps. The first step is to use $V_B - V_A = - \int_A^B \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$ and let $A = a = 4.0 \text{ cm}$ and $B = b = 12.0 \text{ cm}$, with $d\vec{\mathbf{l}} = d\vec{\mathbf{r}} = \hat{\mathbf{r}}dr$ and $\vec{\mathbf{E}} = \frac{kq}{r^2} \hat{\mathbf{r}}$. Then perform the integral. The second step is to integrate $V_B - V_A = - \int_A^B \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$ around an arc of constant radius r , which means we let $d\vec{\mathbf{l}} = r\hat{\boldsymbol{\phi}}d\varphi$ with limits $0 \leq \varphi \leq 24^\circ$, still using $\vec{\mathbf{E}} = \frac{kq}{r^2} \hat{\mathbf{r}}$. Then add the two results together.

Solution

For the first part, $V_B - V_A = - \int_A^B \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$ for this system becomes

$$V_b - V_a = - \int_a^b \frac{kq}{r^2} \hat{\mathbf{r}} \cdot \hat{\mathbf{r}} dr \text{ which computes to}$$

Equation:

$$\begin{aligned} \Delta V &= - \int_a^b \frac{kq}{r^2} dr = kq \left[\frac{1}{a} - \frac{1}{b} \right] \\ &= \left(8.99 \times 10^9 \text{ Nm}^2/\text{C}^2 \right) (2.0 \times 10^{-9} \text{ C}) \left[\frac{1}{0.040 \text{ m}} - \frac{1}{0.12 \text{ m}} \right] = 300 \text{ V}. \end{aligned}$$

For the second step, $V_B - V_A = - \int_A^B \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$ becomes

$$\Delta V = - \int_0^{24^\circ} \frac{kq}{r^2} \hat{\mathbf{r}} \cdot r\hat{\boldsymbol{\phi}} d\varphi, \text{ but } \hat{\mathbf{r}} \cdot \hat{\boldsymbol{\phi}} = 0 \text{ and therefore } \Delta V = 0. \text{ Adding the two}$$

parts together, we get 300 V.

Significance

We have demonstrated the use of the integral form of the potential difference to obtain a numerical result. Notice that, in this particular system, we could have also used the formula for the potential due to a point charge at the two points and simply taken the difference.

Note:

Exercise:

Problem:

Check Your Understanding From the examples, how does the energy of a lightning strike vary with the height of the clouds from the ground? Consider the cloud-ground system to be two parallel plates.

Solution:

Given a fixed maximum electric field strength, the potential at which a strike occurs increases with increasing height above the ground. Hence, each electron will carry more energy. Determining if there is an effect on the total number of electrons lies in the future.

Before presenting problems involving electrostatics, we suggest a problem-solving strategy to follow for this topic.

Note:

Electrostatics

1. Examine the situation to determine if static electricity is involved; this may concern separated stationary charges, the forces among them, and the electric fields they create.
2. Identify the system of interest. This includes noting the number, locations, and types of charges involved.
3. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Determine whether the Coulomb force is

to be considered directly—if so, it may be useful to draw a free-body diagram, using electric field lines.

4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). It is important to distinguish the Coulomb force F from the electric field E , for example.
5. Solve the appropriate equation for the quantity to be determined (the unknown) or draw the field lines as requested.
6. Examine the answer to see if it is reasonable: Does it make sense? Are units correct and the numbers involved reasonable?

Summary

- Electric potential is potential energy per unit charge.
- The potential difference between points A and B , $V_B - V_A$, that is, the change in potential of a charge q moved from A to B , is equal to the change in potential energy divided by the charge.
- Potential difference is commonly called voltage, represented by the symbol ΔV :
$$\Delta V = \frac{\Delta U}{q} \text{ or } \Delta U = q\Delta V.$$
- An electron-volt is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,
$$1 \text{ eV} = (1.60 \times 10^{-19} \text{ C})(1 \text{ V})$$
$$= (1.60 \times 10^{-19} \text{ C})(1 \text{ J/C}) = 1.60 \times 10^{-19} \text{ J}.$$

Conceptual Questions

Exercise:

Problem:

Discuss how potential difference and electric field strength are related. Give an example.

Exercise:

Problem:

What is the strength of the electric field in a region where the electric potential is constant?

Solution:

The electric field strength is zero because electric potential differences are directly related to the field strength. If the potential difference is zero, then the field strength must also be zero.

Exercise:**Problem:**

If a proton is released from rest in an electric field, will it move in the direction of increasing or decreasing potential? Also answer this question for an electron and a neutron. Explain why.

Exercise:**Problem:**

Voltage is the common word for potential difference. Which term is more descriptive, voltage or potential difference?

Solution:

Potential difference is more descriptive because it indicates that it is the difference between the electric potential of two points.

Exercise:**Problem:**

If the voltage between two points is zero, can a test charge be moved between them with zero net work being done? Can this necessarily be done without exerting a force? Explain.

Exercise:**Problem:**

What is the relationship between voltage and energy? More precisely, what is the relationship between potential difference and electric potential energy?

Solution:

They are very similar, but potential difference is a feature of the system; when a charge is introduced to the system, it will have a potential energy which may be calculated by multiplying the magnitude of the charge by the potential difference.

Exercise:

Problem: Voltages are always measured between two points. Why?

Exercise:**Problem:**

How are units of volts and electron-volts related? How do they differ?

Solution:

An electron-volt is a volt multiplied by the charge of an electron. Volts measure potential difference, electron-volts are a unit of energy.

Exercise:**Problem:**

Can a particle move in a direction of increasing electric potential, yet have its electric potential energy decrease? Explain

Problems**Exercise:****Problem:**

Find the ratio of speeds of an electron and a negative hydrogen ion (one having an extra electron) accelerated through the same voltage, assuming non-relativistic final speeds. Take the mass of the hydrogen ion to be 1.67×10^{-27} kg.

Solution:

$$\begin{aligned}\frac{1}{2}m_e v_e^2 &= qV, \quad \frac{1}{2}m_H v_H^2 = qV, \text{ so that} \\ \frac{m_e v_e^2}{m_H v_H^2} &= 1 \text{ or } \frac{v_e}{v_H} = 42.8\end{aligned}$$

Exercise:

Problem:

An evacuated tube uses an accelerating voltage of 40 kV to accelerate electrons to hit a copper plate and produce X-rays. Non-relativistically, what would be the maximum speed of these electrons?

Exercise:**Problem:**

Show that units of V/m and N/C for electric field strength are indeed equivalent.

Solution:

$$1 \text{ V} = 1 \text{ J/C}; 1 \text{ J} = 1 \text{ N} \cdot \text{m} \rightarrow 1 \text{ V/m} = 1 \text{ N/C}$$

Exercise:**Problem:**

What is the strength of the electric field between two parallel conducting plates separated by 1.00 cm and having a potential difference (voltage) between them of $1.50 \times 10^4 \text{ V}$?

Exercise:**Problem:**

The electric field strength between two parallel conducting plates separated by 4.00 cm is $7.50 \times 10^4 \text{ V/m}$. (a) What is the potential difference between the plates? (b) The plate with the lowest potential is taken to be zero volts. What is the potential 1.00 cm from that plate and 3.00 cm from the other?

Solution:

$$\text{a. } V_{AB} = 3.00 \text{ kV}; \text{ b. } V_{AB} = 750 \text{ V}$$

Exercise:**Problem:**

The voltage across a membrane forming a cell wall is 80.0 mV and the membrane is 9.00 nm thick. What is the electric field strength? (The value is surprisingly large, but correct.) You may assume a uniform electric field.

Exercise:

Problem:

Two parallel conducting plates are separated by 10.0 cm, and one of them is taken to be at zero volts. (a) What is the electric field strength between them, if the potential 8.00 cm from the zero volt plate (and 2.00 cm from the other) is 450 V? (b) What is the voltage between the plates?

Solution:

a. $V_{AB} = Ed \rightarrow E = 5.63 \text{ kV/m};$

b. $V_{AB} = 563 \text{ V}$

Exercise:**Problem:**

Find the maximum potential difference between two parallel conducting plates separated by 0.500 cm of air, given the maximum sustainable electric field strength in air to be $3.0 \times 10^6 \text{ V/m}$.

Exercise:**Problem:**

An electron is to be accelerated in a uniform electric field having a strength of $2.00 \times 10^6 \text{ V/m}$. (a) What energy in keV is given to the electron if it is accelerated through 0.400 m? (b) Over what distance would it have to be accelerated to increase its energy by 50.0 GeV?

Solution:

$\Delta K = q\Delta V$ and $V_{AB} = Ed$, so that

a. $\Delta K = 800 \text{ keV};$

b. $d = 25.0 \text{ km}$

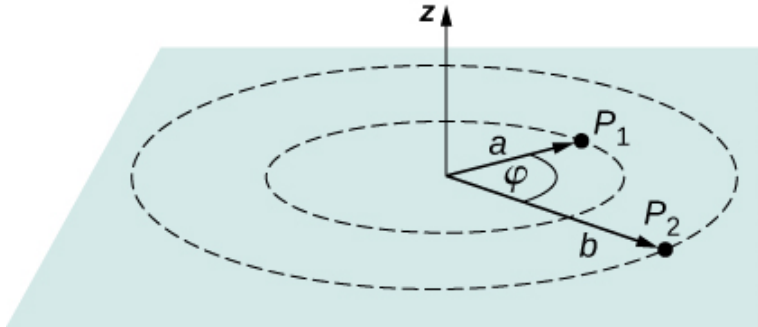
Exercise:**Problem:**

Use the definition of potential difference in terms of electric field to deduce the formula for potential difference between $r = r_a$ and $r = r_b$ for a point charge located at the origin. Here r is the spherical radial coordinate.

Exercise:

Problem:

The electric field in a region is pointed away from the z-axis and the magnitude depends upon the distance s from the axis. The magnitude of the electric field is given as $E = \frac{\alpha}{s}$ where α is a constant. Find the potential difference between points P_1 and P_2 , explicitly stating the path over which you conduct the integration for the line integral.



Solution:

One possibility is to stay at constant radius and go along the arc from P_1 to P_2 , which will have zero potential due to the path being perpendicular to the electric field. Then integrate from a to b : $V_{ab} = \alpha \ln \left(\frac{b}{a} \right)$

Exercise:**Problem:**

Singly charged gas ions are accelerated from rest through a voltage of 13.0 V. At what temperature will the average kinetic energy of gas molecules be the same as that given these ions?

Glossary

electric potential

potential energy per unit charge

electric potential difference

the change in potential energy of a charge q moved between two points, divided by the charge.

electron-volt

energy given to a fundamental charge accelerated through a potential difference of one volt

voltage

change in potential energy of a charge moved from one point to another, divided by the charge; units of potential difference are joules per coulomb, known as volt

Calculations of Electric Potential

By the end of this section, you will be able to:

- Calculate the potential due to a point charge
- Calculate the potential of a system of multiple point charges
- Describe an electric dipole
- Define dipole moment
- Calculate the potential of a continuous charge distribution

Point charges, such as electrons, are among the fundamental building blocks of matter. Furthermore, spherical charge distributions (such as charge on a metal sphere) create external electric fields exactly like a point charge. The electric potential due to a point charge is, thus, a case we need to consider.

We can use calculus to find the work needed to move a test charge q from a large distance away to a distance of r from a point charge q . Noting the connection between work and potential $W = -q\Delta V$, as in the last section, we can obtain the following result.

Note:**Electric Potential V of a Point Charge**

The electric potential V of a point charge is given by

Equation:

$$V = \frac{kq}{r} \text{ (point charge)}$$

where k is a constant equal to $8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$.

The potential at infinity is chosen to be zero. Thus, V for a point charge decreases with distance, whereas \vec{E} for a point charge decreases with distance squared:

Equation:

$$E = \frac{F}{q_t} = \frac{kq}{r^2}.$$

Recall that the electric potential V is a scalar and has no direction, whereas the electric field \vec{E} is a vector. To find the voltage due to a combination of point charges, you add the individual voltages as numbers. To find the total electric field, you must add the individual fields as vectors, taking magnitude and direction into account. This is consistent with the fact that V is closely associated with energy, a scalar, whereas \vec{E} is closely associated with force, a vector.

Example:**What Voltage Is Produced by a Small Charge on a Metal Sphere?**

Charges in static electricity are typically in the nanocoulomb (nC) to microcoulomb (μC) range. What is the voltage 5.00 cm away from the center of a 1-cm-diameter solid metal sphere that has a -3.00-nC static charge?

Strategy

As we discussed in [Electric Charges and Fields](#), charge on a metal sphere spreads out uniformly and produces a field like that of a point charge located at its center. Thus, we can find the voltage using the equation $V = \frac{kq}{r}$.

Solution

Entering known values into the expression for the potential of a point charge, we obtain

Equation:

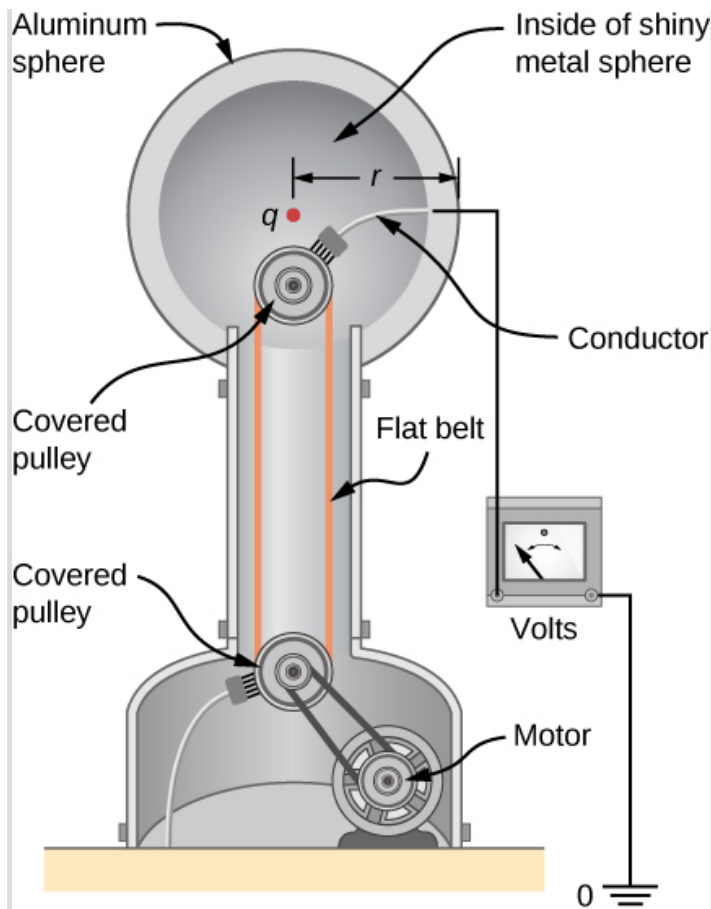
$$V = k\frac{q}{r} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2\right) \left(\frac{-3.00 \times 10^{-9} \text{ C}}{5.00 \times 10^{-2} \text{ m}}\right) = -539 \text{ V}.$$

Significance

The negative value for voltage means a positive charge would be attracted from a larger distance, since the potential is lower (more negative) than at larger distances. Conversely, a negative charge would be repelled, as expected.

Example:**What Is the Excess Charge on a Van de Graaff Generator?**

A demonstration Van de Graaff generator has a 25.0-cm-diameter metal sphere that produces a voltage of 100 kV near its surface ([link](#)). What excess charge resides on the sphere? (Assume that each numerical value here is shown with three significant figures.)



The voltage of this demonstration Van de Graaff generator is measured between the charged sphere and ground. Earth's potential is taken to be zero as a reference. The potential of the charged conducting sphere is the same as that of an equal point charge at its center.

Strategy

The potential on the surface is the same as that of a point charge at the center of the sphere, 12.5 cm away. (The radius of the sphere is 12.5 cm.) We can thus determine the excess charge using the equation

Equation:

$$V = \frac{kq}{r}.$$

Solution

Solving for q and entering known values gives

Equation:

$$q = \frac{rV}{k} = \frac{(0.125 \text{ m})(100 \times 10^3 \text{ V})}{8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2} = 1.39 \times 10^{-6} \text{ C} = 1.39 \mu\text{C}.$$

Significance

This is a relatively small charge, but it produces a rather large voltage. We have another indication here that it is difficult to store isolated charges.

Note:

Exercise:

Problem:

Check Your Understanding What is the potential inside the metal sphere in [\[link\]](#)?

Solution:

$V = k \frac{q}{r} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2\right) \left(\frac{-3.00 \times 10^{-9} \text{ C}}{5.00 \times 10^{-3} \text{ m}}\right) = -5390 \text{ V}$; recall that the electric field inside a conductor is zero. Hence, any path from a point on the surface to any point in the interior will have an integrand of zero when calculating the change in potential, and thus the potential in the interior of the sphere is identical to that on the surface.

The voltages in both of these examples could be measured with a meter that compares the measured potential with ground potential. Ground potential is often taken to be zero (instead of taking the potential at infinity to be zero). It is the potential difference between two points that is of importance, and very often there is a tacit assumption that some reference point, such as Earth or a very distant point, is at zero potential. As noted earlier, this is analogous to taking sea level as $h = 0$ when considering gravitational potential energy $U_g = mgh$.

Systems of Multiple Point Charges

Just as the electric field obeys a superposition principle, so does the electric potential. Consider a system consisting of N charges q_1, q_2, \dots, q_N . What is the net electric potential V at a space point P from these charges? Each of these charges is a source charge that produces its own electric potential at point P , independent of whatever other charges may be doing. Let V_1, V_2, \dots, V_N be the electric potentials at P produced by the charges q_1, q_2, \dots, q_N , respectively. Then, the net electric potential V_P at that point is equal to the sum of these individual electric potentials. You can easily show this by calculating the potential energy of a test charge when you bring the test charge from the reference point at infinity to point P :

Equation:

$$V_P = V_1 + V_2 + \cdots + V_N = \sum_1^N V_i.$$

Note that electric potential follows the same principle of superposition as electric field and electric potential energy. To show this more explicitly, note that a test charge q_i at the point P in space has distances of r_1, r_2, \dots, r_N from the N charges fixed in space above, as shown in [\[link\]](#). Using our formula for the potential of a point charge for each of these (assumed to be point) charges, we find that

Note:

Equation:

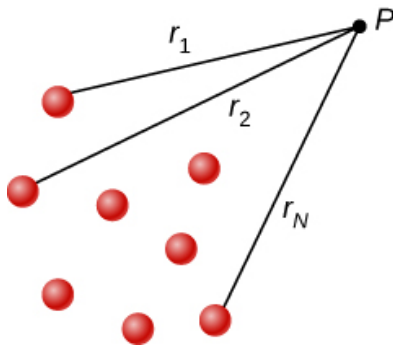
$$V_P = \sum_1^N k \frac{q_i}{r_i} = k \sum_1^N \frac{q_i}{r_i}.$$

Therefore, the electric potential energy of the test charge is

Equation:

$$U_P = q_t V_P = q_t k \sum_1^N \frac{q_i}{r_i},$$

which is the same as the work to bring the test charge into the system, as found in the first section of the chapter.



Notation for direct
distances from charges
to a space point P .

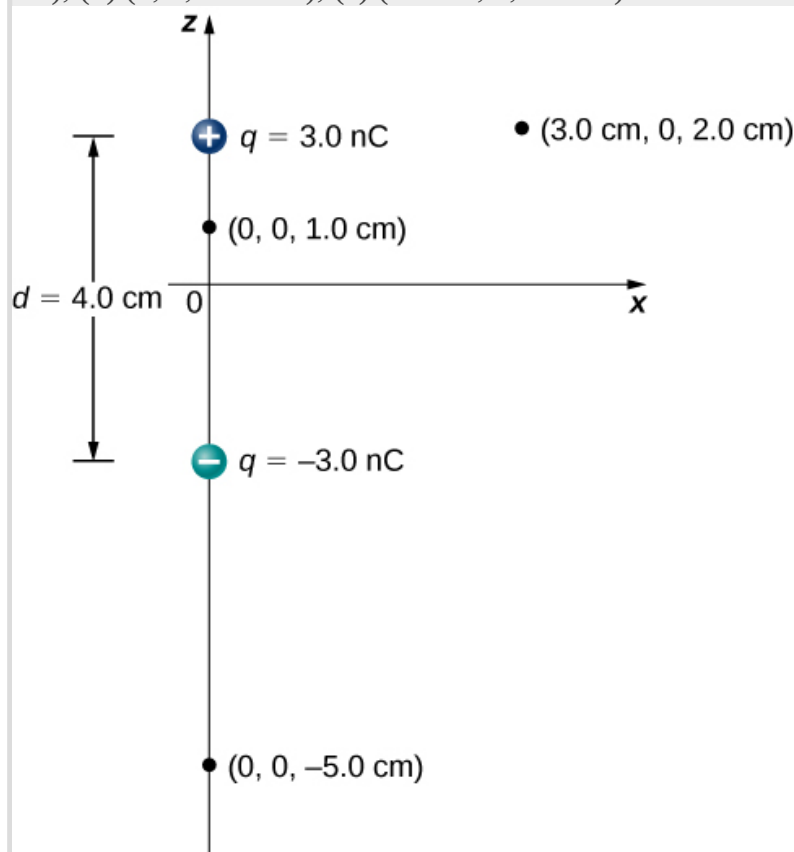
The Electric Dipole

An **electric dipole** is a system of two equal but opposite charges a fixed distance apart. This system is used to model many real-world systems, including atomic and molecular interactions. One of these systems is the water molecule, under certain circumstances. These circumstances are met inside a microwave oven, where electric fields with alternating directions make the water molecules change orientation. This vibration is the same as heat at the molecular level.

Example:

Electric Potential of a Dipole

Consider the dipole in [\[link\]](#) with the charge magnitude of $q = 3.0 \text{ nC}$ and separation distance $d = 4.0 \text{ cm}$. What is the potential at the following locations in space? (a) $(0, 0, 1.0 \text{ cm})$; (b) $(0, 0, -5.0 \text{ cm})$; (c) $(3.0 \text{ cm}, 0, 2.0 \text{ cm})$.



A general diagram of an electric dipole, and the notation for the distances from the individual charges

to a point P in space.

Strategy

Apply $V_P = k \sum_1^N \frac{q_i}{r_i}$ to each of these three points.

Solution

$$\text{a. } V_P = k \sum_1^N \frac{q_i}{r_i} = (9.0 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \left(\frac{3.0 \text{ nC}}{0.010 \text{ m}} - \frac{3.0 \text{ nC}}{0.030 \text{ m}} \right) = 1.8 \times 10^3 \text{ V}$$

$$\text{b. } V_P = k \sum_1^N \frac{q_i}{r_i} = (9.0 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \left(\frac{3.0 \text{ nC}}{0.070 \text{ m}} - \frac{3.0 \text{ nC}}{0.030 \text{ m}} \right) = -5.1 \times 10^2 \text{ V}$$

$$\text{c. } V_P = k \sum_1^N \frac{q_i}{r_i} = (9.0 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \left(\frac{3.0 \text{ nC}}{0.030 \text{ m}} - \frac{3.0 \text{ nC}}{0.050 \text{ m}} \right) = 3.6 \times 10^2 \text{ V}$$

Significance

Note that evaluating potential is significantly simpler than electric field, due to potential being a scalar instead of a vector.

Note:

Exercise:

Problem: Check Your Understanding What is the potential on the x -axis? The z -axis?

Solution:

The x -axis the potential is zero, due to the equal and opposite charges the same distance from it. On the z -axis, we may superimpose the two potentials; we will find that for $z \gg d$, again the potential goes to zero due to cancellation.

Now let us consider the special case when the distance of the point P from the dipole is much greater than the distance between the charges in the dipole, $r \gg d$; for example, when we are interested in the electric potential due to a polarized molecule such as a water molecule. This is not so far (infinity) that we can simply treat the potential as zero, but the distance is great enough that we can simplify our calculations relative to the previous example.

We start by noting that in [\[link\]](#) the potential is given by

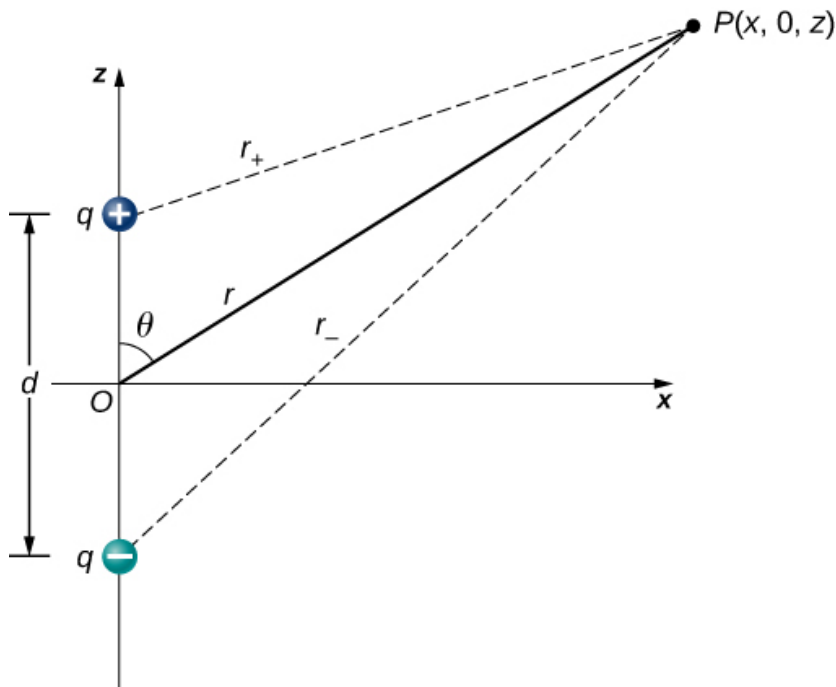
Equation:

$$V_P = V_+ + V_- = k \left(\frac{q}{r_+} - \frac{q}{r_-} \right)$$

where

Equation:

$$r_{\pm} = \sqrt{x^2 + \left(z \mp \frac{d}{2} \right)^2}.$$



A general diagram of an electric dipole, and the notation for the distances from the individual charges to a point P in space.

This is still the exact formula. To take advantage of the fact that $r \gg d$, we rewrite the radii in terms of polar coordinates, with $x = r \sin \theta$ and $z = r \cos \theta$. This gives us

Equation:

$$r_{\pm} = \sqrt{r^2 \sin^2 \theta + \left(r \cos \theta \mp \frac{d}{2} \right)^2}.$$

We can simplify this expression by pulling r out of the root,

Equation:

$$r_{\pm} = r \sqrt{\sin^2 \theta + \left(\cos \theta \mp \frac{d}{2r} \right)^2}$$

and then multiplying out the parentheses

Equation:

$$r_{\pm} = r \sqrt{\sin^2 \theta + \cos^2 \theta \mp \cos \theta \frac{d}{r} + \left(\frac{d}{2r} \right)^2} = r \sqrt{1 \mp \cos \theta \frac{d}{r} + \left(\frac{d}{2r} \right)^2}.$$

The last term in the root is small enough to be negligible (remember $r \gg d$, and hence $(d/r)^2$ is extremely small, effectively zero to the level we will probably be measuring), leaving us with

Equation:

$$r_{\pm} = r \sqrt{1 \mp \cos \theta \frac{d}{r}}.$$

Using the binomial approximation (a standard result from the mathematics of series, when α is small)

Equation:

$$\frac{1}{\sqrt{1 \mp \alpha}} \approx 1 \pm \frac{\alpha}{2}$$

and substituting this into our formula for V_P , we get

Equation:

$$V_P = k \left[\frac{q}{r} \left(1 + \frac{d \cos \theta}{2r} \right) - \frac{q}{r} \left(1 - \frac{d \cos \theta}{2r} \right) \right] = k \frac{qd \cos \theta}{r^2}.$$

This may be written more conveniently if we define a new quantity, the **electric dipole moment**,

Note:

Equation:

$$\vec{\mathbf{p}} = q\vec{\mathbf{d}},$$

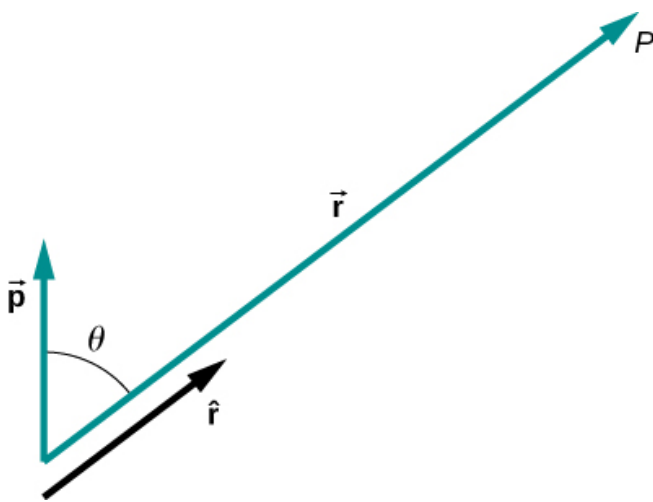
where these vectors point from the negative to the positive charge. Note that this has magnitude qd . This quantity allows us to write the potential at point P due to a dipole at the origin as

Note:

Equation:

$$V_P = k \frac{\vec{\mathbf{p}} \cdot \hat{\mathbf{r}}}{r^2}.$$

A diagram of the application of this formula is shown in [\[link\]](#).



The geometry for the application of the potential of a dipole.

There are also higher-order moments, for quadrupoles, octupoles, and so on. You will see these in future classes.

Potential of Continuous Charge Distributions

We have been working with point charges a great deal, but what about continuous charge distributions? Recall from [\[link\]](#) that

Equation:

$$V_P = k \sum \frac{q_i}{r_i}.$$

We may treat a continuous charge distribution as a collection of infinitesimally separated individual points. This yields the integral

Note:

Equation:

$$V_P = k \int \frac{dq}{r}$$

for the potential at a point P . Note that r is the distance from each individual point in the charge distribution to the point P . As we saw in [Electric Charges and Fields](#), the infinitesimal charges are given by

Equation:

$$dq = \begin{cases} \lambda dl & \text{(one dimension)} \\ \sigma dA & \text{(two dimensions)} \\ \rho dV & \text{(three dimensions)} \end{cases}$$

where λ is linear charge density, σ is the charge per unit area, and ρ is the charge per unit volume.

Example:

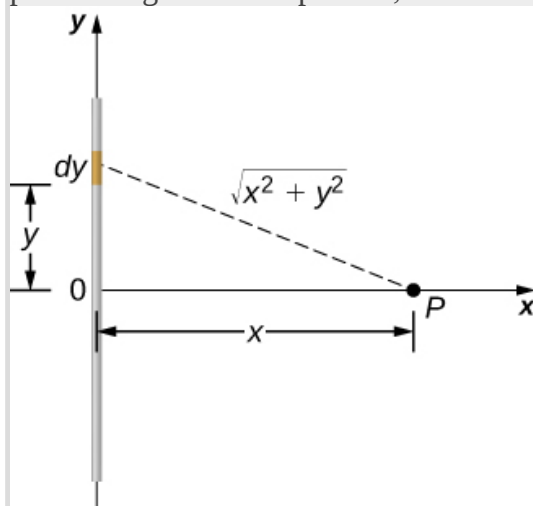
Potential of a Line of Charge

Find the electric potential of a uniformly charged, nonconducting wire with linear density λ (coulomb/meter) and length L at a point that lies on a line that divides the wire into two equal parts.

Strategy

To set up the problem, we choose Cartesian coordinates in such a way as to exploit the symmetry in the problem as much as possible. We place the origin at the center of the wire

and orient the y -axis along the wire so that the ends of the wire are at $y = \pm L/2$. The field point P is in the xy -plane and since the choice of axes is up to us, we choose the x -axis to pass through the field point P , as shown in [\[link\]](#).



We want to calculate the electric potential due to a line of charge.

Solution

Consider a small element of the charge distribution between y and $y + dy$. The charge in this cell is $dq = \lambda dy$ and the distance from the cell to the field point P is $\sqrt{x^2 + y^2}$. Therefore, the potential becomes

Equation:

$$\begin{aligned} V_P &= k \int \frac{dq}{r} = k \int_{-L/2}^{L/2} \frac{\lambda dy}{\sqrt{x^2 + y^2}} = k\lambda \left[\ln \left(y + \sqrt{y^2 + x^2} \right) \right]_{-L/2}^{L/2} \\ &= k\lambda \left[\ln \left(\left(\frac{L}{2} \right) + \sqrt{\left(\frac{L}{2} \right)^2 + x^2} \right) - \ln \left(\left(-\frac{L}{2} \right) + \sqrt{\left(-\frac{L}{2} \right)^2 + x^2} \right) \right] \\ &= k\lambda \ln \left[\frac{L + \sqrt{L^2 + 4x^2}}{-L + \sqrt{L^2 + 4x^2}} \right]. \end{aligned}$$

Significance

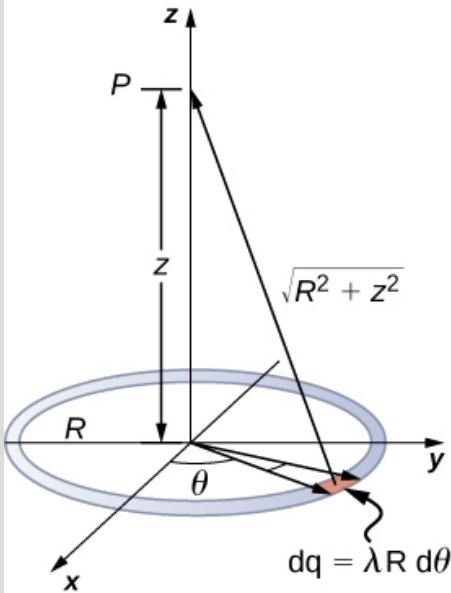
Note that this was simpler than the equivalent problem for electric field, due to the use of scalar quantities. Recall that we expect the zero level of the potential to be at infinity, when we have a finite charge. To examine this, we take the limit of the above potential as x approaches infinity; in this case, the terms inside the natural log approach one, and hence the potential approaches zero in this limit. Note that we could have done this problem equivalently in cylindrical coordinates; the only effect would be to substitute r for x and z for y .

Example:**Potential Due to a Ring of Charge**

A ring has a uniform charge density λ , with units of coulomb per unit meter of arc. Find the electric potential at a point on the axis passing through the center of the ring.

Strategy

We use the same procedure as for the charged wire. The difference here is that the charge is distributed on a circle. We divide the circle into infinitesimal elements shaped as arcs on the circle and use cylindrical coordinates shown in [\[link\]](#).



We want to calculate the electric potential due to a ring of charge.

Solution

A general element of the arc between θ and $\theta + d\theta$ is of length $Rd\theta$ and therefore contains a charge equal to $\lambda R d\theta$. The element is at a distance of $\sqrt{z^2 + R^2}$ from P , and therefore the potential is

Equation:

$$V_P = k \int \frac{dq}{r} = k \int_0^{2\pi} \frac{\lambda R d\theta}{\sqrt{z^2 + R^2}} = \frac{k\lambda R}{\sqrt{z^2 + R^2}} \int_0^{2\pi} d\theta = \frac{2\pi k\lambda R}{\sqrt{z^2 + R^2}} = k \frac{q_{\text{tot}}}{\sqrt{z^2 + R^2}}.$$

Significance

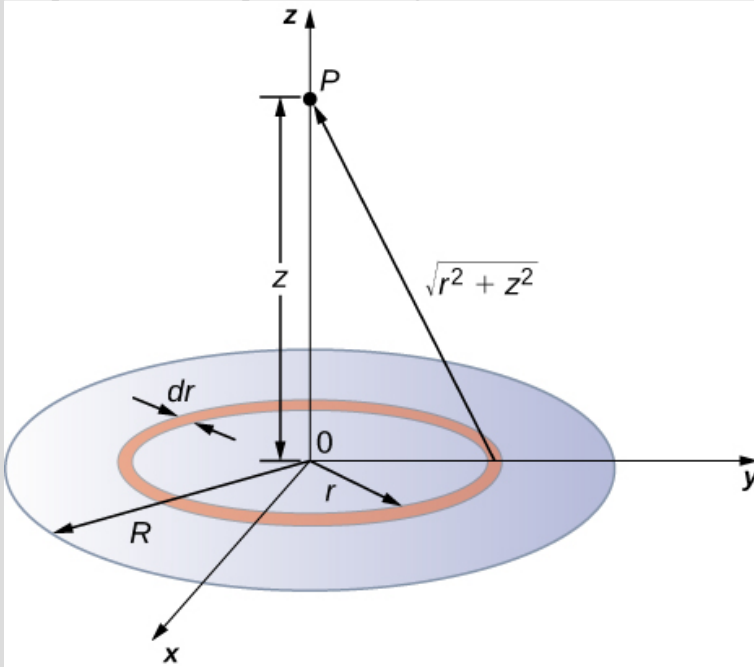
This result is expected because every element of the ring is at the same distance from point P . The net potential at P is that of the total charge placed at the common distance, $\sqrt{z^2 + R^2}$.

Example:**Potential Due to a Uniform Disk of Charge**

A disk of radius R has a uniform charge density σ , with units of coulomb meter squared. Find the electric potential at any point on the axis passing through the center of the disk.

Strategy

We divide the disk into ring-shaped cells, and make use of the result for a ring worked out in the previous example, then integrate over r in addition to θ . This is shown in [\[link\]](#).



We want to calculate the electric potential due to a disk of charge.

Solution

An infinitesimal width cell between cylindrical coordinates r and $r + dr$ shown in [\[link\]](#) will be a ring of charges whose electric potential dV_P at the field point has the following expression

Equation:

$$dV_P = k \frac{dq}{\sqrt{z^2 + r^2}}$$

where

Equation:

$$dq = \sigma 2\pi r dr.$$

The superposition of potential of all the infinitesimal rings that make up the disk gives the net potential at point P . This is accomplished by integrating from $r = 0$ to $r = R$:

Equation:

$$\begin{aligned}
 V_P &= \int dV_P = k2\pi\sigma \int_0^R \frac{r \, dr}{\sqrt{z^2 + r^2}}, \\
 &= k2\pi\sigma \left(\sqrt{z^2 + R^2} - \sqrt{z^2} \right).
 \end{aligned}$$

Significance

The basic procedure for a disk is to first integrate around θ and then over r . This has been demonstrated for uniform (constant) charge density. Often, the charge density will vary with r , and then the last integral will give different results.

Example:**Potential Due to an Infinite Charged Wire**

Find the electric potential due to an infinitely long uniformly charged wire.

Strategy

Since we have already worked out the potential of a finite wire of length L in [\[link\]](#), we might wonder if taking $L \rightarrow \infty$ in our previous result will work:

Equation:

$$V_P = \lim_{L \rightarrow \infty} k\lambda \ln \left(\frac{L + \sqrt{L^2 + 4x^2}}{-L + \sqrt{L^2 + 4x^2}} \right).$$

However, this limit does not exist because the argument of the logarithm becomes $[2/0]$ as $L \rightarrow \infty$, so this way of finding V of an infinite wire does not work. The reason for this problem may be traced to the fact that the charges are not localized in some space but continue to infinity in the direction of the wire. Hence, our (unspoken) assumption that zero potential must be an infinite distance from the wire is no longer valid.

To avoid this difficulty in calculating limits, let us use the definition of potential by integrating over the electric field from the previous section, and the value of the electric field from this charge configuration from the previous chapter.

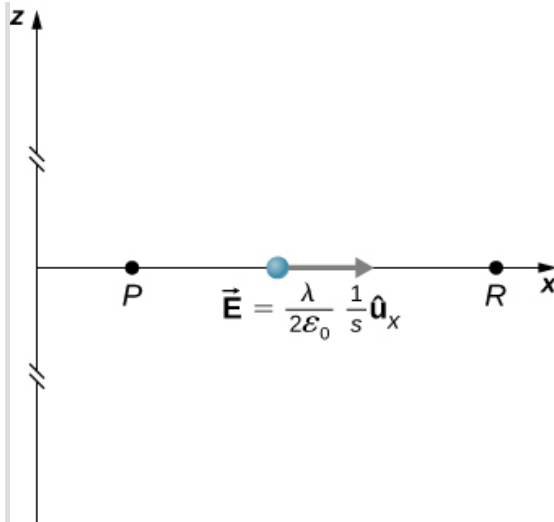
Solution

We use the integral

Equation:

$$V_P = - \int_R^P \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$$

where R is a finite distance from the line of charge, as shown in [\[link\]](#).



Points of interest for calculating the potential of an infinite line of charge.

With this setup, we use $\vec{E}_P = 2k\lambda \frac{1}{s} \hat{s}$ and $d\vec{l} = d\vec{s}$ to obtain

Equation:

$$V_P - V_R = - \int_R^P 2k\lambda \frac{1}{s} ds = -2k\lambda \ln \frac{s_P}{s_R}.$$

Now, if we define the reference potential $V_R = 0$ at $s_R = 1$ m, this simplifies to

Equation:

$$V_P = -2k\lambda \ln s_P.$$

Note that this form of the potential is quite usable; it is 0 at 1 m and is undefined at infinity, which is why we could not use the latter as a reference.

Significance

Although calculating potential directly can be quite convenient, we just found a system for which this strategy does not work well. In such cases, going back to the definition of potential in terms of the electric field may offer a way forward.

Note:

Exercise:

Problem:

Check Your Understanding What is the potential on the axis of a nonuniform ring of charge, where the charge density is $\lambda(\theta) = \lambda \cos \theta$?

Solution:

It will be zero, as at all points on the axis, there are equal and opposite charges equidistant from the point of interest. Note that this distribution will, in fact, have a dipole moment.

Summary

- Electric potential is a scalar whereas electric field is a vector.
- Addition of voltages as numbers gives the voltage due to a combination of point charges, allowing us to use the principle of superposition: $V_P = k \sum_1^N \frac{q_i}{r_i}$.
- An electric dipole consists of two equal and opposite charges a fixed distance apart, with a dipole moment $\vec{p} = q\vec{d}$.
- Continuous charge distributions may be calculated with $V_P = k \int \frac{dq}{r}$.

Conceptual Questions**Exercise:****Problem:**

Compare the electric dipole moments of charges $\pm Q$ separated by a distance d and charges $\pm Q/2$ separated by a distance $d/2$.

Solution:

The second has 1/4 the dipole moment of the first.

Exercise:**Problem:**

Would Gauss's law be helpful for determining the electric field of a dipole? Why?

Exercise:

Problem:

In what region of space is the potential due to a uniformly charged sphere the same as that of a point charge? In what region does it differ from that of a point charge?

Solution:

The region outside of the sphere will have a potential indistinguishable from a point charge; the interior of the sphere will have a different potential.

Exercise:**Problem:**

Can the potential of a nonuniformly charged sphere be the same as that of a point charge? Explain.

Problems**Exercise:****Problem:**

A 0.500-cm-diameter plastic sphere, used in a static electricity demonstration, has a uniformly distributed 40.0-pC charge on its surface. What is the potential near its surface?

Solution:

$$V = 144 \text{ V}$$

Exercise:**Problem:**

How far from a 1.00- μC point charge is the potential 100 V? At what distance is it $2.00 \times 10^2 \text{ V}$?

Exercise:**Problem:**

If the potential due to a point charge is $5.00 \times 10^2 \text{ V}$ at a distance of 15.0 m, what are the sign and magnitude of the charge?

Solution:

$$V = \frac{kQ}{r} \rightarrow Q = 8.33 \times 10^{-7} \text{ C};$$

The charge is positive because the potential is positive.

Exercise:**Problem:**

In nuclear fission, a nucleus splits roughly in half. (a) What is the potential 2.00×10^{-14} m from a fragment that has 46 protons in it? (b) What is the potential energy in MeV of a similarly charged fragment at this distance?

Exercise:**Problem:**

A research Van de Graaff generator has a 2.00-m-diameter metal sphere with a charge of 5.00 mC on it. Assume the potential energy is zero at a reference point infinitely far away from the Van de Graaff. (a) What is the potential near its surface? (b) At what distance from its center is the potential 1.00 MV? (c) An oxygen atom with three missing electrons is released near the Van de Graaff generator. What is its kinetic energy in MeV when the atom is at the distance found in part b?

Solution:

- a. $V = 45.0$ MV;
- b. $V = \frac{kQ}{r} \rightarrow r = 45.0$ m;
- c. $\Delta U = 132$ MeV

Exercise:**Problem:**

An electrostatic paint sprayer has a 0.200-m-diameter metal sphere at a potential of 25.0 kV that repels paint droplets onto a grounded object.

(a) What charge is on the sphere? (b) What charge must a 0.100-mg drop of paint have to arrive at the object with a speed of 10.0 m/s?

Exercise:**Problem:**

(a) What is the potential between two points situated 10 cm and 20 cm from a $3.0\text{-}\mu\text{C}$ point charge? (b) To what location should the point at 20 cm be moved to increase this potential difference by a factor of two?

Solution:

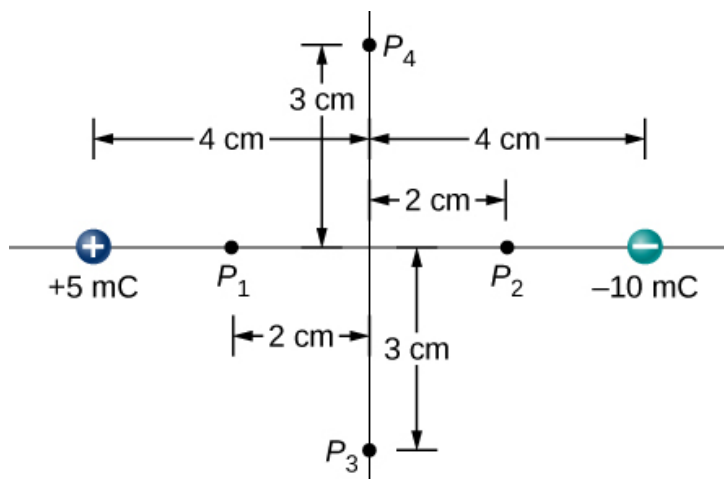
$V = kQ/r$; a. Relative to origin, find the potential at each point and then calculate the difference.

$$\Delta V = 135 \times 10^3 \text{ V};$$

b. To double the potential difference, move the point from 20 cm to infinity; the potential at 20 cm is halfway between zero and that at 10 cm.

Exercise:**Problem:**

Find the potential at points P_1 , P_2 , P_3 , and P_4 in the diagram due to the two given charges.

**Exercise:****Problem:**

Two charges $-2.0 \mu\text{C}$ and $+2.0 \mu\text{C}$ are separated by 4.0 cm on the z -axis symmetrically about origin, with the positive one uppermost. Two space points of interest P_1 and P_2 are located 3.0 cm and 30 cm from origin at an angle 30° with respect to the z -axis. Evaluate electric potentials at P_1 and P_2 in two ways: (a) Using the exact formula for point charges, and (b) using the approximate dipole potential formula.

Solution:

- a. $V_{P1} = 7.4 \times 10^5 \text{ V}$
and $V_{P2} = 6.9 \times 10^3 \text{ V}$;
b. $V_{P1} = 6.9 \times 10^5 \text{ V}$ and $V_{P2} = 6.9 \times 10^3 \text{ V}$

Exercise:**Problem:**

(a) Plot the potential of a uniformly charged 1-m rod with 1 C/m charge as a function of the perpendicular distance from the center. Draw your graph from $s = 0.1 \text{ m}$ to $s = 1.0 \text{ m}$. (b) On the same graph, plot the potential of a point charge with a 1-C charge at the origin. (c) Which potential is stronger near the rod? (d) What happens to the difference as the distance increases? Interpret your result.

Glossary

electric dipole

system of two equal but opposite charges a fixed distance apart

electric dipole moment

quantity defined as $\vec{\mathbf{p}} = q\vec{\mathbf{d}}$ for all dipoles, where the vector points from the negative to positive charge

Determining Field from Potential

By the end of this section, you will be able to:

- Explain how to calculate the electric field in a system from the given potential
- Calculate the electric field in a given direction from a given potential
- Calculate the electric field throughout space from a given potential

Recall that we were able, in certain systems, to calculate the potential by integrating over the electric field. As you may already suspect, this means that we may calculate the electric field by taking derivatives of the potential, although going from a scalar to a vector quantity introduces some interesting wrinkles. We frequently need \vec{E} to calculate the force in a system; since it is often simpler to calculate the potential directly, there are systems in which it is useful to calculate V and then derive \vec{E} from it.

In general, regardless of whether the electric field is uniform, it points in the direction of decreasing potential, because the force on a positive charge is in the direction of \vec{E} and also in the direction of lower potential V . Furthermore, the magnitude of \vec{E} equals the rate of decrease of V with distance. The faster V decreases over distance, the greater the electric field. This gives us the following result.

Note:

Relationship between Voltage and Uniform Electric Field

In equation form, the relationship between voltage and uniform electric field is

Equation:

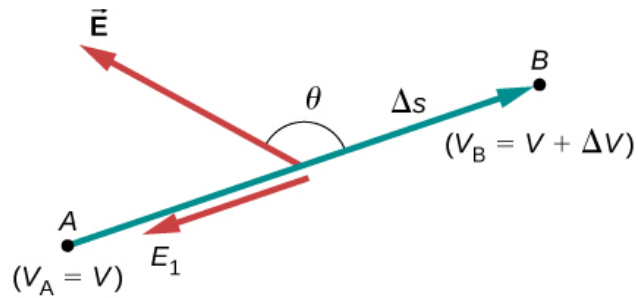
$$E = -\frac{\Delta V}{\Delta s}$$

where Δs is the distance over which the change in potential ΔV takes place. The minus sign tells us that E points in the direction of decreasing potential. The electric field is said to be the gradient (as in grade or slope) of the electric potential.

For continually changing potentials, ΔV and Δs become infinitesimals, and we need differential calculus to determine the electric field. As shown in [\[link\]](#), if we treat the distance Δs as very small so that the electric field is essentially constant over it, we find that

Equation:

$$E_s = -\frac{dV}{ds}.$$



The electric field component, E_1 , along the displacement Δs is given by $E = -\frac{\Delta V}{\Delta s}$. Note that A and B are assumed to be so close together that the field is constant along Δs .

Therefore, the electric field components in the Cartesian directions are given by

Note:

Equation:

$$E_x = -\frac{\partial V}{\partial x}, E_y = -\frac{\partial V}{\partial y}, E_z = -\frac{\partial V}{\partial z}.$$

This allows us to define the “grad” or “del” vector operator, which allows us to compute the gradient in one step. In Cartesian coordinates, it takes the form

Note:

Equation:

$$\vec{\nabla} = \hat{i}\frac{\partial}{\partial x} + \hat{j}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z}.$$

With this notation, we can calculate the electric field from the potential with

Note:

Equation:

$$\vec{E} = -\vec{\nabla}V,$$

a process we call calculating the gradient of the potential.

If we have a system with either cylindrical or spherical symmetry, we only need to use the del operator in the appropriate coordinates:

Note:

Equation:

$$\text{Cylindrical: } \vec{\nabla} = \hat{\mathbf{r}} \frac{\partial}{\partial r} + \hat{\boldsymbol{\varphi}} \frac{1}{r} \frac{\partial}{\partial \varphi} + \hat{\mathbf{z}} \frac{\partial}{\partial z}$$

Note:

Equation:

$$\text{Spherical: } \vec{\nabla} = \hat{\mathbf{r}} \frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{\boldsymbol{\varphi}} \frac{1}{r \sin \theta} \frac{\partial}{\partial \varphi}$$

Example:

Electric Field of a Point Charge

Calculate the electric field of a point charge from the potential.

Strategy

The potential is known to be $V = k \frac{q}{r}$, which has a spherical symmetry. Therefore, we use the spherical del operator in the formula $\vec{\mathbf{E}} = -\vec{\nabla} V$.

Solution

Performing this calculation gives us

Equation:

$$\vec{\mathbf{E}} = - \left(\hat{\mathbf{r}} \frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{\boldsymbol{\varphi}} \frac{1}{r \sin \theta} \frac{\partial}{\partial \varphi} \right) k \frac{q}{r} = -kq \left(\hat{\mathbf{r}} \frac{\partial}{\partial r} \frac{1}{r} + \hat{\boldsymbol{\theta}} \frac{1}{r} \frac{\partial}{\partial \theta} \frac{1}{r} + \hat{\boldsymbol{\varphi}} \frac{1}{r \sin \theta} \frac{\partial}{\partial \varphi} \frac{1}{r} \right).$$

This equation simplifies to

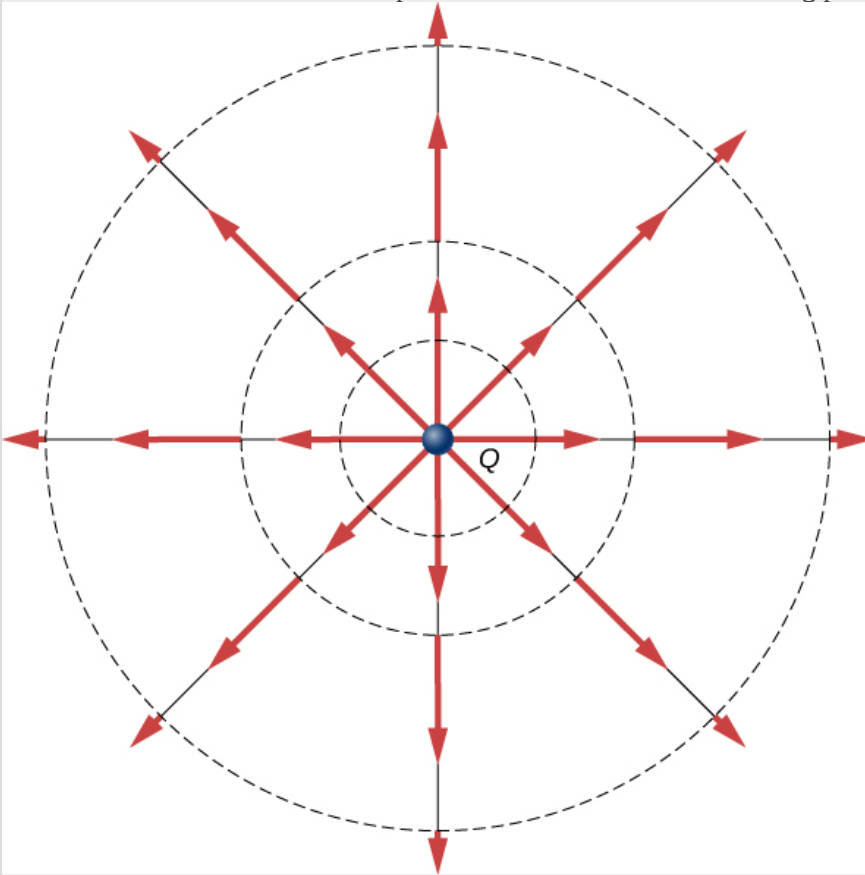
Equation:

$$\vec{\mathbf{E}} = -kq \left(\hat{\mathbf{r}} \frac{-1}{r^2} + \hat{\boldsymbol{\theta}} 0 + \hat{\boldsymbol{\varphi}} 0 \right) = k \frac{q}{r^2} \hat{\mathbf{r}}$$

as expected.

Significance

We not only obtained the equation for the electric field of a point particle that we've seen before, we also have a demonstration that \vec{E} points in the direction of decreasing potential, as shown in [\[link\]](#).

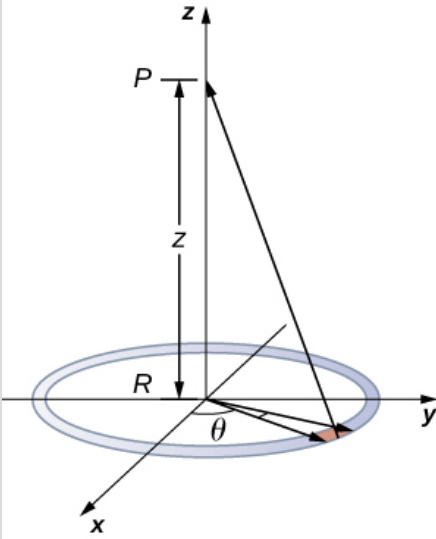


Electric field vectors inside and outside a uniformly charged sphere.

Example:

Electric Field of a Ring of Charge

Use the potential found in [\[link\]](#) to calculate the electric field along the axis of a ring of charge ([\[link\]](#)).



We want to calculate the electric field from the electric potential due to a ring charge.

Strategy

In this case, we are only interested in one dimension, the z -axis. Therefore, we use $E_z = -\frac{\partial V}{\partial z}$ with the potential $V = k \frac{q_{\text{tot}}}{\sqrt{z^2 + R^2}}$ found previously.

Solution

Taking the derivative of the potential yields

Equation:

$$E_z = -\frac{\partial}{\partial z} \frac{kq_{\text{tot}}}{\sqrt{z^2 + R^2}} = k \frac{q_{\text{tot}} z}{(z^2 + R^2)^{3/2}}.$$

Significance

Again, this matches the equation for the electric field found previously. It also demonstrates a system in which using the full del operator is not necessary.

Note:

Exercise:

Problem:

Check Your Understanding Which coordinate system would you use to calculate the electric field of a dipole?

Solution:

Any, but cylindrical is closest to the symmetry of a dipole.

Summary

- Just as we may integrate over the electric field to calculate the potential, we may take the derivative of the potential to calculate the electric field.
- This may be done for individual components of the electric field, or we may calculate the entire electric field vector with the gradient operator.

Conceptual Questions

Exercise:

Problem:

If the electric field is zero throughout a region, must the electric potential also be zero in that region?

Solution:

No. It will be constant, but not necessarily zero.

Exercise:

Problem:

Explain why knowledge of $\vec{E}(x, y, z)$ is not sufficient to determine $V(x, y, z)$. What about the other way around?

Problems

Exercise:

Problem:

Throughout a region, equipotential surfaces are given by $z = \text{constant}$. The surfaces are equally spaced with $V = 100 \text{ V}$ for $z = 0.00 \text{ m}$, $V = 200 \text{ V}$ for $z = 0.50 \text{ m}$, $V = 300 \text{ V}$ for $z = 1.00 \text{ m}$. What is the electric field in this region?

Solution:

The problem is describing a uniform field, so $E = 200 \text{ V/m}$ in the $-z$ -direction.

Exercise:

Problem:

In a particular region, the electric potential is given by $V = -xy^2z + 4xy$. What is the electric field in this region?

Exercise:

Problem: Calculate the electric field of an infinite line charge, throughout space.

Solution:

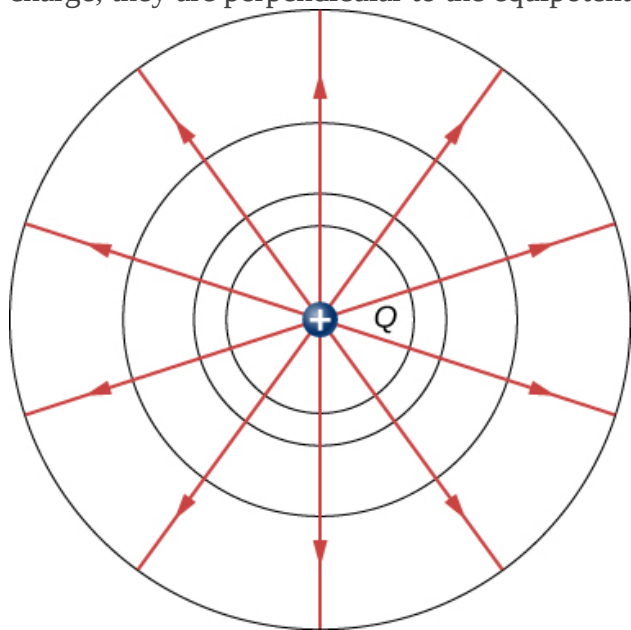
Apply $\vec{\mathbf{E}} = -\vec{\nabla}V$ with $\vec{\nabla} = \hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\varphi}\frac{1}{r}\frac{\partial}{\partial\varphi} + \hat{\mathbf{z}}\frac{\partial}{\partial z}$ to the potential calculated earlier,
 $V = -2k\lambda\ln s$: $\vec{\mathbf{E}} = 2k\lambda\frac{1}{r}\hat{\mathbf{r}}$ as expected.

Equipotential Surfaces and Conductors

By the end of this section, you will be able to:

- Define equipotential surfaces and equipotential lines
- Explain the relationship between equipotential lines and electric field lines
- Map equipotential lines for one or two point charges
- Describe the potential of a conductor
- Compare and contrast equipotential lines and elevation lines on topographic maps

We can represent electric potentials (voltages) pictorially, just as we drew pictures to illustrate electric fields. This is not surprising, since the two concepts are related. Consider [\[link\]](#), which shows an isolated positive point charge and its electric field lines, which radiate out from a positive charge and terminate on negative charges. We use red arrows to represent the magnitude and direction of the electric field, and we use black lines to represent places where the electric potential is constant. These are called **equipotential surfaces** in three dimensions, or **equipotential lines** in two dimensions. The term *equipotential* is also used as a noun, referring to an equipotential line or surface. The potential for a point charge is the same anywhere on an imaginary sphere of radius r surrounding the charge. This is true because the potential for a point charge is given by $V = kq/r$ and thus has the same value at any point that is a given distance r from the charge. An equipotential sphere is a circle in the two-dimensional view of [\[link\]](#). Because the electric field lines point radially away from the charge, they are perpendicular to the equipotential lines.



An isolated point charge Q with its electric field lines in red and equipotential lines in black. The potential is the same along each equipotential line, meaning that no work is required to move a charge anywhere along one of those lines. Work

any where along one of those lines. From
 is needed to move a charge from one
 equipotential line to another.
 Equipotential lines are perpendicular to
 electric field lines in every case. For a
 three-dimensional version, explore the
 first media link.

It is important to note that *equipotential lines are always perpendicular to electric field lines*. No work is required to move a charge along an equipotential, since $\Delta V = 0$. Thus, the work is

Equation:

$$W = -\Delta U = -q\Delta V = 0.$$

Work is zero if the direction of the force is perpendicular to the displacement. Force is in the same direction as E , so motion along an equipotential must be perpendicular to E . More precisely, work is related to the electric field by

Equation:

$$W = \vec{\mathbf{F}} \cdot \vec{\mathbf{d}} = q\vec{\mathbf{E}} \cdot \vec{\mathbf{d}} = qEd \cos \theta = 0.$$

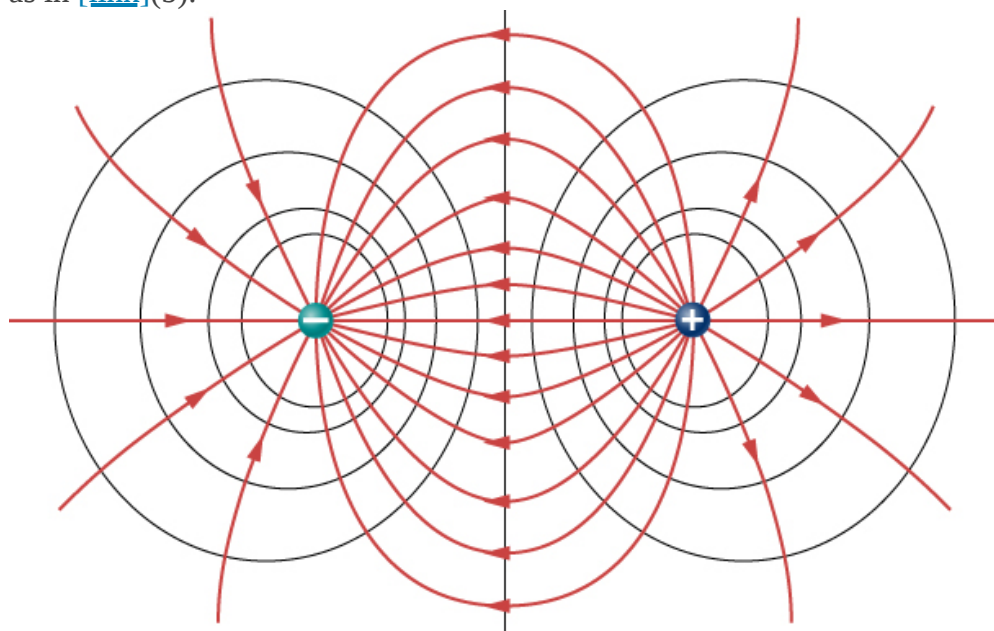
Note that in this equation, E and F symbolize the magnitudes of the electric field and force, respectively. Neither q nor E is zero; d is also not zero. So $\cos \theta$ must be 0, meaning θ must be 90° . In other words, motion along an equipotential is perpendicular to E .

One of the rules for static electric fields and conductors is that the electric field must be perpendicular to the surface of any conductor. This implies that a *conductor is an equipotential surface in static situations*. There can be no voltage difference across the surface of a conductor, or charges will flow. One of the uses of this fact is that a conductor can be fixed at what we consider zero volts by connecting it to the earth with a good conductor—a process called **grounding**. Grounding can be a useful safety tool. For example, grounding the metal case of an electrical appliance ensures that it is at zero volts relative to Earth.

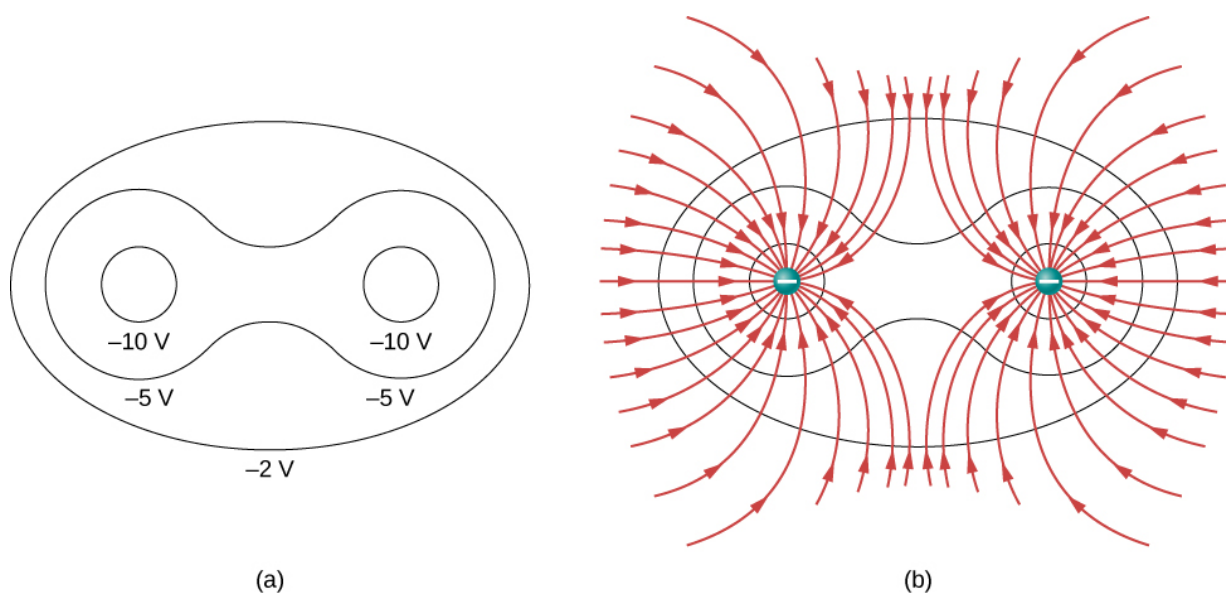
Because a conductor is an equipotential, it can replace any equipotential surface. For example, in [\[link\]](#), a charged spherical conductor can replace the point charge, and the electric field and potential surfaces outside of it will be unchanged, confirming the contention that a spherical charge distribution is equivalent to a point charge at its center.

[\[link\]](#) shows the electric field and equipotential lines for two equal and opposite charges. Given the electric field lines, the equipotential lines can be drawn simply by making them perpendicular to the electric field lines. Conversely, given the equipotential lines, as in [\[link\]](#)

(a), the electric field lines can be drawn by making them perpendicular to the equipotentials, as in [link](#)(b).



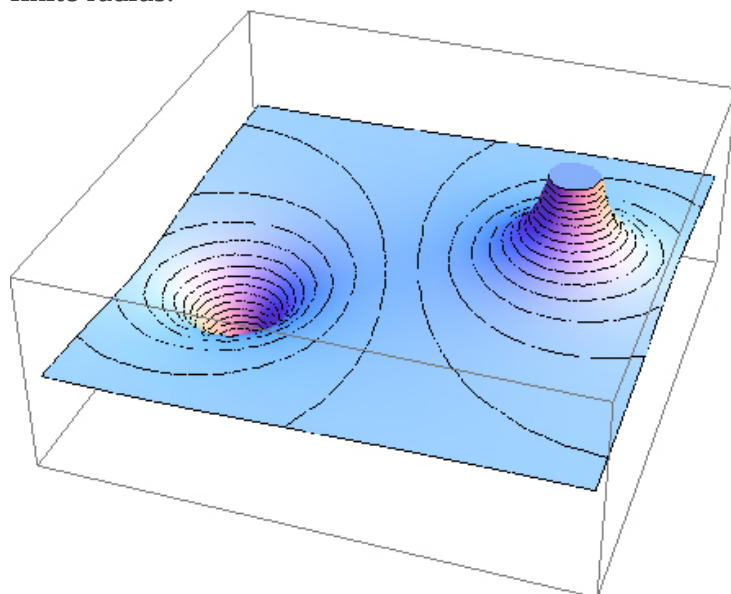
The electric field lines and equipotential lines for two equal but opposite charges. The equipotential lines can be drawn by making them perpendicular to the electric field lines, if those are known. Note that the potential is greatest (most positive) near the positive charge and least (most negative) near the negative charge. For a three-dimensional version, explore the first media link.



(a) These equipotential lines might be measured with a voltmeter in a laboratory

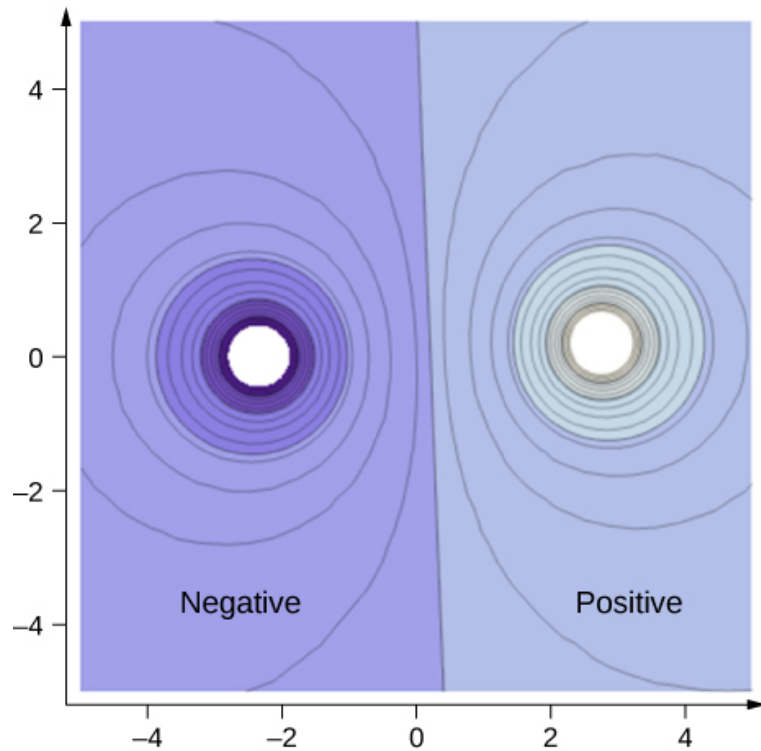
experiment. (b) The corresponding electric field lines are found by drawing them perpendicular to the equipotentials. Note that these fields are consistent with two equal negative charges. For a three-dimensional version, play with the first media link.

To improve your intuition, we show a three-dimensional variant of the potential in a system with two opposing charges. [\[link\]](#) displays a three-dimensional map of electric potential, where lines on the map are for equipotential surfaces. The hill is at the positive charge, and the trough is at the negative charge. The potential is zero far away from the charges. Note that the cut off at a particular potential implies that the charges are on conducting spheres with a finite radius.



Electric potential map of two opposite charges of equal magnitude on conducting spheres. The potential is negative near the negative charge and positive near the positive charge.

A two-dimensional map of the cross-sectional plane that contains both charges is shown in [\[link\]](#). The line that is equidistant from the two opposite charges corresponds to zero potential, since at the points on the line, the positive potential from the positive charge cancels the negative potential from the negative charge. Equipotential lines in the cross-sectional plane are closed loops, which are not necessarily circles, since at each point, the net potential is the sum of the potentials from each charge.

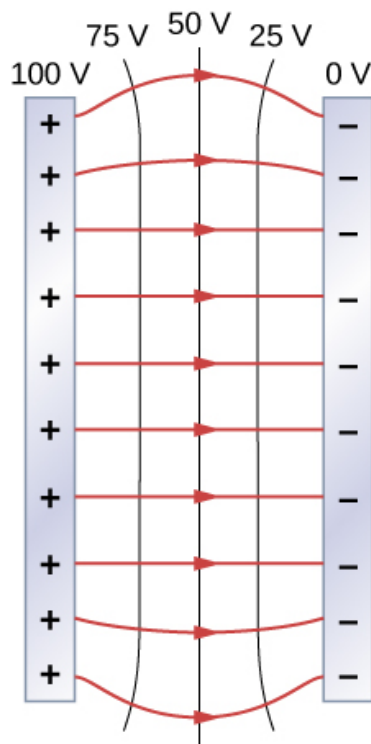


A cross-section of the electric potential map of two opposite charges of equal magnitude. The potential is negative near the negative charge and positive near the positive charge.

Note:

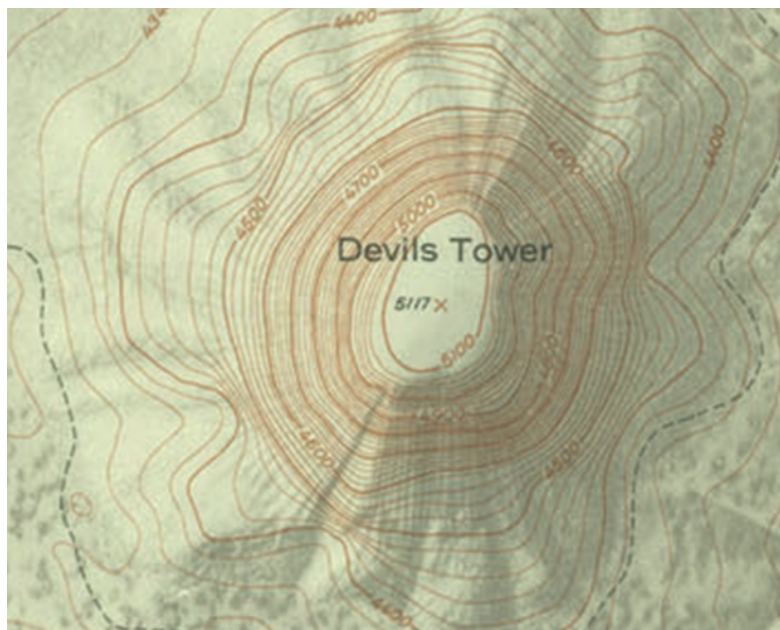
View this [simulation](#) to observe and modify the equipotential surfaces and electric fields for many standard charge configurations. There's a lot to explore.

One of the most important cases is that of the familiar parallel conducting plates shown in [\[link\]](#). Between the plates, the equipotentials are evenly spaced and parallel. The same field could be maintained by placing conducting plates at the equipotential lines at the potentials shown.



The electric field and equipotential lines between two metal plates. Note that the electric field is perpendicular to the equipotentials and hence normal to the plates at their surface as well as in the center of the region between them.

Consider the parallel plates in [\[link\]](#). These have equipotential lines that are parallel to the plates in the space between and evenly spaced. An example of this (with sample values) is given in [\[link\]](#). We could draw a similar set of equipotential isolines for gravity on the hill shown in [\[link\]](#). If the hill has any extent at the same slope, the isolines along that extent would be parallel to each other. Furthermore, in regions of constant slope, the isolines would be evenly spaced. An example of real topographic lines is shown in [\[link\]](#).



(a)



(b)

A topographical map along a ridge has roughly parallel elevation lines, similar to the equipotential lines in [\[link\]](#). (a) A topographical map of Devil's Tower, Wyoming. Lines that are close together indicate very steep terrain. (b) A perspective photo of Devil's Tower shows just how steep its sides are. Notice the top of the tower has the same shape as the center of the topographical map.

Example:

Calculating Equipotential Lines

You have seen the equipotential lines of a point charge in [\[link\]](#). How do we calculate them? For example, if we have a $+10\text{-nC}$ charge at the origin, what are the equipotential surfaces at which the potential is (a) 100 V, (b) 50 V, (c) 20 V, and (d) 10 V?

Strategy

Set the equation for the potential of a point charge equal to a constant and solve for the remaining variable(s). Then calculate values as needed.

Solution

In $V = k\frac{q}{r}$, let V be a constant. The only remaining variable is r ; hence,

$r = k\frac{q}{V} = \text{constant}$. Thus, the equipotential surfaces are spheres about the origin. Their locations are:

$$\text{a. } r = k\frac{q}{V} = \left(8.99 \times 10^9 \text{ Nm}^2/\text{C}^2\right) \frac{(10 \times 10^{-9} \text{ C})}{100 \text{ V}} = 0.90 \text{ m};$$

$$\text{b. } r = k\frac{q}{V} = \left(8.99 \times 10^9 \text{ Nm}^2/\text{C}^2\right) \frac{(10 \times 10^{-9} \text{ C})}{50 \text{ V}} = 1.8 \text{ m};$$

$$\text{c. } r = k \frac{q}{V} = \left(8.99 \times 10^9 \text{ Nm}^2/\text{C}^2 \right) \frac{(10 \times 10^{-9} \text{ C})}{20 \text{ V}} = 4.5 \text{ m};$$

$$\text{d. } r = k \frac{q}{V} = \left(8.99 \times 10^9 \text{ Nm}^2/\text{C}^2 \right) \frac{(10 \times 10^{-9} \text{ C})}{10 \text{ V}} = 9.0 \text{ m}.$$

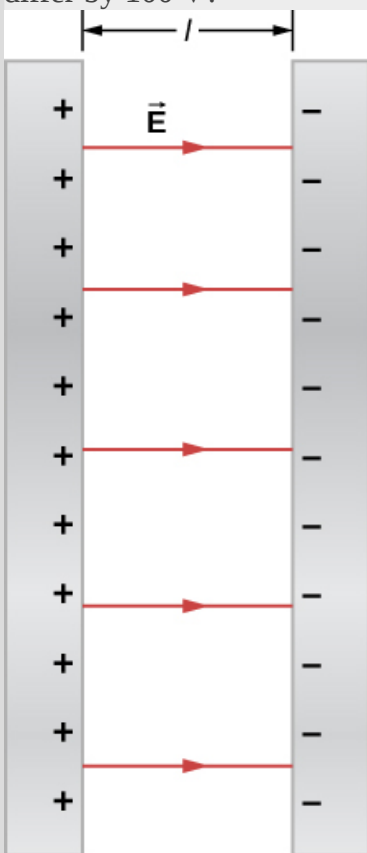
Significance

This means that equipotential surfaces around a point charge are spheres of constant radius, as shown earlier, with well-defined locations.

Example:

Potential Difference between Oppositely Charged Parallel Plates

Two large conducting plates carry equal and opposite charges, with a surface charge density σ of magnitude $6.81 \times 10^{-7} \text{ C/m}^2$, as shown in [\[link\]](#). The separation between the plates is $l = 6.50 \text{ mm}$. (a) What is the electric field between the plates? (b) What is the potential difference between the plates? (c) What is the distance between equipotential planes which differ by 100 V?



The electric field
between oppositely
charged parallel plates.

A portion is released
at the positive plate.

Strategy

(a) Since the plates are described as “large” and the distance between them is not, we will approximate each of them as an infinite plane, and apply the result from Gauss’s law in the previous chapter.

(b) Use $\Delta V_{AB} = - \int_A^B \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$.

(c) Since the electric field is constant, find the ratio of 100 V to the total potential difference; then calculate this fraction of the distance.

Solution

- a. The electric field is directed from the positive to the negative plate as shown in the figure, and its magnitude is given by

Equation:

$$E = \frac{\sigma}{\epsilon_0} = \frac{6.81 \times 10^{-7} \text{ C/m}^2}{8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2} = 7.69 \times 10^4 \text{ V/m}.$$

- b. To find the potential difference ΔV between the plates, we use a path from the negative to the positive plate that is directed against the field. The displacement vector $d\vec{\mathbf{l}}$ and the electric field $\vec{\mathbf{E}}$ are antiparallel so $\vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} = -E dl$. The potential difference between the positive plate and the negative plate is then

Equation:

$$\Delta V = - \int E \cdot dl = E \int dl = El = (7.69 \times 10^4 \text{ V/m})(6.50 \times 10^{-3} \text{ m}) = 500 \text{ V}.$$

- c. The total potential difference is 500 V, so 1/5 of the distance between the plates will be the distance between 100-V potential differences. The distance between the plates is 6.5 mm, so there will be 1.3 mm between 100-V potential differences.

Significance

You have now seen a numerical calculation of the locations of equipotentials between two charged parallel plates.

Note:

Exercise:

Problem:

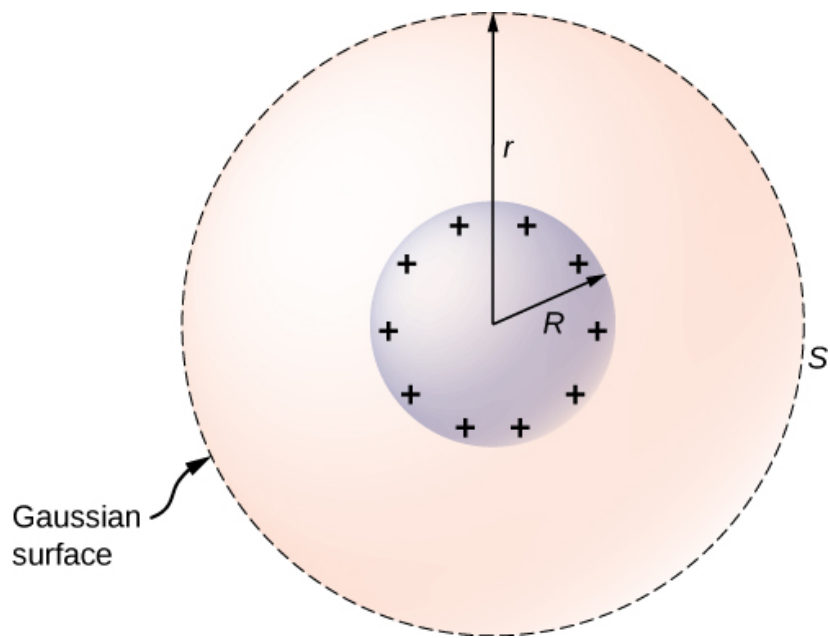
Check Your Understanding What are the equipotential surfaces for an infinite line charge?

Solution:

infinite cylinders of constant radius, with the line charge as the axis

Distribution of Charges on Conductors

In [\[link\]](#) with a point charge, we found that the equipotential surfaces were in the form of spheres, with the point charge at the center. Given that a conducting sphere in electrostatic equilibrium is a spherical equipotential surface, we should expect that we could replace one of the surfaces in [\[link\]](#) with a conducting sphere and have an identical solution outside the sphere. Inside will be rather different, however.



An isolated conducting sphere.

To investigate this, consider the isolated conducting sphere of [\[link\]](#) that has a radius R and an excess charge q . To find the electric field both inside and outside the sphere, note that the sphere is isolated, so its surface charge distribution and the electric field of that distribution

are spherically symmetric. We can therefore represent the field as $\vec{\mathbf{E}} = E(r)\hat{\mathbf{r}}$. To calculate $E(r)$, we apply Gauss's law over a closed spherical surface S of radius r that is concentric with the conducting sphere. Since r is constant and $\hat{\mathbf{n}} = \hat{\mathbf{r}}$ on the sphere,

Equation:

$$\oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} da = E(r) \oint da = E(r) 4\pi r^2.$$

For $r < R$, S is within the conductor, so recall from our previous study of Gauss's law that $q_{\text{enc}} = 0$ and Gauss's law gives $E(r) = 0$, as expected inside a conductor at equilibrium. If $r > R$, S encloses the conductor so $q_{\text{enc}} = q$. From Gauss's law,

Equation:

$$E(r) 4\pi r^2 = \frac{q}{\epsilon_0}.$$

The electric field of the sphere may therefore be written as

Equation:

$$\begin{aligned} E &= 0 & (r < R), \\ E &= \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \hat{\mathbf{r}} & (r \geq R). \end{aligned}$$

As expected, in the region $r \geq R$, the electric field due to a charge q placed on an isolated conducting sphere of radius R is identical to the electric field of a point charge q located at the center of the sphere.

To find the electric potential inside and outside the sphere, note that for $r \geq R$, the potential must be the same as that of an isolated point charge q located at $r = 0$,

Equation:

$$V(r) = \frac{1}{4\pi\epsilon_0} \frac{q}{r} \quad (r \geq R)$$

simply due to the similarity of the electric field.

For $r < R$, $E = 0$, so $V(r)$ is constant in this region. Since $V(R) = q/4\pi\epsilon_0 R$,

Equation:

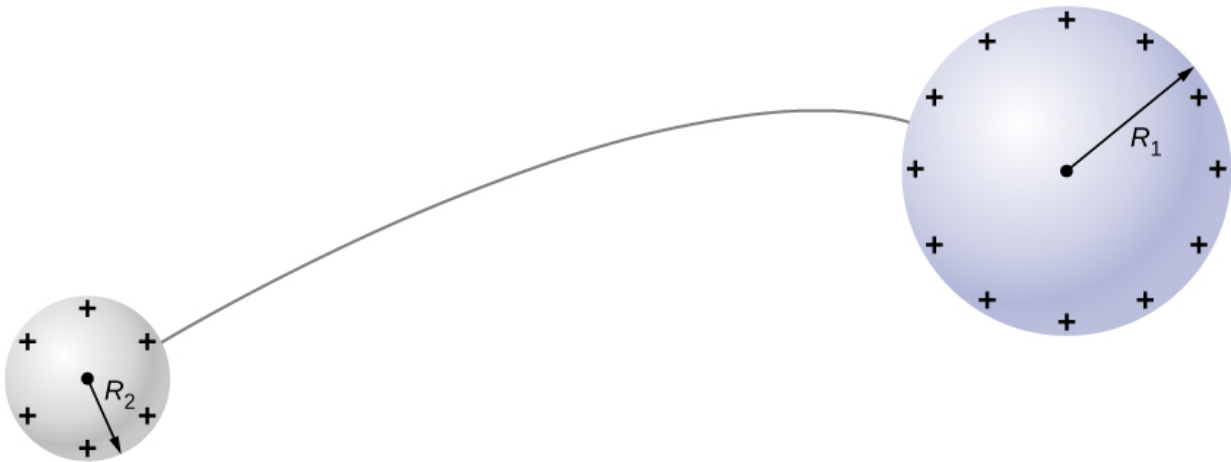
$$V(r) = \frac{1}{4\pi\epsilon_0} \frac{q}{R} \quad (r < R).$$

We will use this result to show that

Equation:

$$\sigma_1 R_1 = \sigma_2 R_2,$$

for two conducting spheres of radii R_1 and R_2 , with surface charge densities σ_1 and σ_2 respectively, that are connected by a thin wire, as shown in [\[link\]](#). The spheres are sufficiently separated so that each can be treated as if it were isolated (aside from the wire). Note that the connection by the wire means that this entire system must be an equipotential.



Two conducting spheres are connected by a thin conducting wire.

We have just seen that the electrical potential at the surface of an isolated, charged conducting sphere of radius R is

Equation:

$$V = \frac{1}{4\pi\epsilon_0} \frac{q}{R}.$$

Now, the spheres are connected by a conductor and are therefore at the same potential; hence

Equation:

$$\frac{1}{4\pi\epsilon_0} \frac{q_1}{R_1} = \frac{1}{4\pi\epsilon_0} \frac{q_2}{R_2},$$

and

Equation:

$$\frac{q_1}{R_1} = \frac{q_2}{R_2}.$$

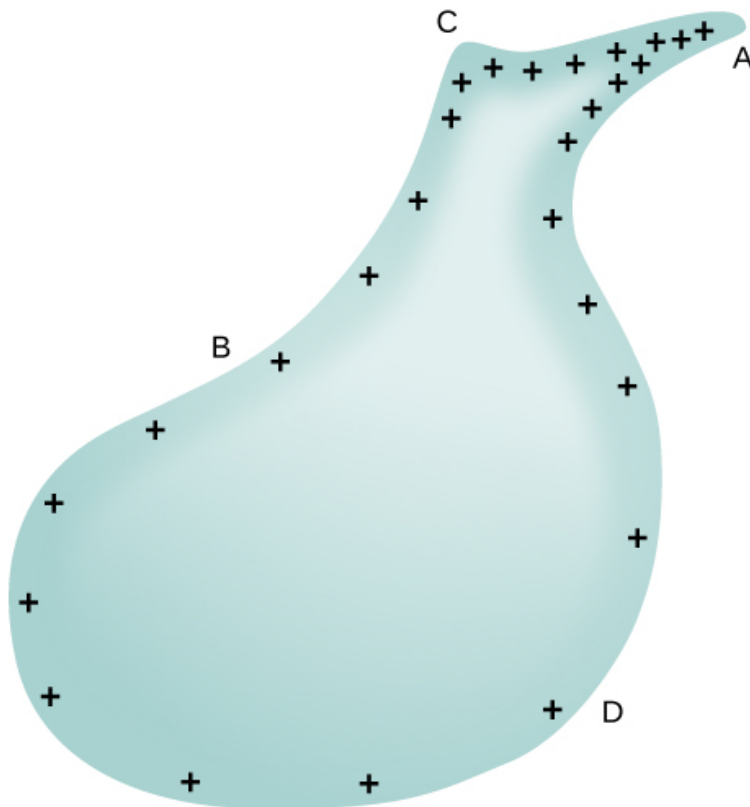
The net charge on a conducting sphere and its surface charge density are related by $q = \sigma(4\pi R^2)$. Substituting this equation into the previous one, we find

Equation:

$$\sigma_1 R_1 = \sigma_2 R_2.$$

Obviously, two spheres connected by a thin wire do not constitute a typical conductor with a variable radius of curvature. Nevertheless, this result does at least provide a qualitative idea of how charge density varies over the surface of a conductor. The equation indicates that where the radius of curvature is large (points *B* and *D* in [\[link\]](#)), σ and E are small.

Similarly, the charges tend to be denser where the curvature of the surface is greater, as demonstrated by the charge distribution on oddly shaped metal ([\[link\]](#)). The surface charge density is higher at locations with a small radius of curvature than at locations with a large radius of curvature.



The surface charge density and the electric field of a conductor are greater at regions with smaller radii of curvature.

A practical application of this phenomenon is the lightning rod, which is simply a grounded metal rod with a sharp end pointing upward. As positive charge accumulates in the ground due to a negatively charged cloud overhead, the electric field around the sharp point gets very large. When the field reaches a value of approximately $3.0 \times 10^6 \text{ N/C}$ (the *dielectric strength* of the air), the free ions in the air are accelerated to such high energies that their collisions with air molecules actually ionize the molecules. The resulting free electrons in the air then flow through the rod to Earth, thereby neutralizing some of the positive charge. This keeps the electric field between the cloud and the ground from getting large enough to produce a lightning bolt in the region around the rod.

An important application of electric fields and equipotential lines involves the heart. The heart relies on electrical signals to maintain its rhythm. The movement of electrical signals causes the chambers of the heart to contract and relax. When a person has a heart attack, the movement of these electrical signals may be disturbed. An artificial pacemaker and a defibrillator can be used to initiate the rhythm of electrical signals. The equipotential lines around the heart, the thoracic region, and the axis of the heart are useful ways of monitoring the structure and functions of the heart. An electrocardiogram (ECG) measures the small electric signals being generated during the activity of the heart.

Note:

Play with the simulation below to move point charges around on the playing field and then view the electric field, voltages, equipotential lines, and more.

<https://openstax.org/l/21fieldlindrpr>

Summary

- An equipotential surface is the collection of points in space that are all at the same potential. Equipotential lines are the two-dimensional representation of equipotential surfaces.
- Equipotential surfaces are always perpendicular to electric field lines.
- Conductors in static equilibrium are equipotential surfaces.
- Topographic maps may be thought of as showing gravitational equipotential lines.

Conceptual Questions

Exercise:

Problem:

If two points are at the same potential, are there any electric field lines connecting them?

Solution:

no

Exercise:**Problem:**

Suppose you have a map of equipotential surfaces spaced 1.0 V apart. What do the distances between the surfaces in a particular region tell you about the strength of the \vec{E} in that region?

Exercise:

Problem: Is the electric potential necessarily constant over the surface of a conductor?

Solution:

No; it might not be at electrostatic equilibrium.

Exercise:**Problem:**

Under electrostatic conditions, the excess charge on a conductor resides on its surface. Does this mean that all of the conduction electrons in a conductor are on the surface?

Exercise:

Problem: Can a positively charged conductor be at a negative potential? Explain.

Solution:

Yes. It depends on where the zero reference for potential is. (Though this might be unusual.)

Exercise:

Problem: Can equipotential surfaces intersect?

Problems**Exercise:**

Problem:

Two very large metal plates are placed 2.0 cm apart, with a potential difference of 12 V between them. Consider one plate to be at 12 V, and the other at 0 V. (a) Sketch the equipotential surfaces for 0, 4, 8, and 12 V. (b) Next sketch in some electric field lines, and confirm that they are perpendicular to the equipotential lines.

Exercise:**Problem:**

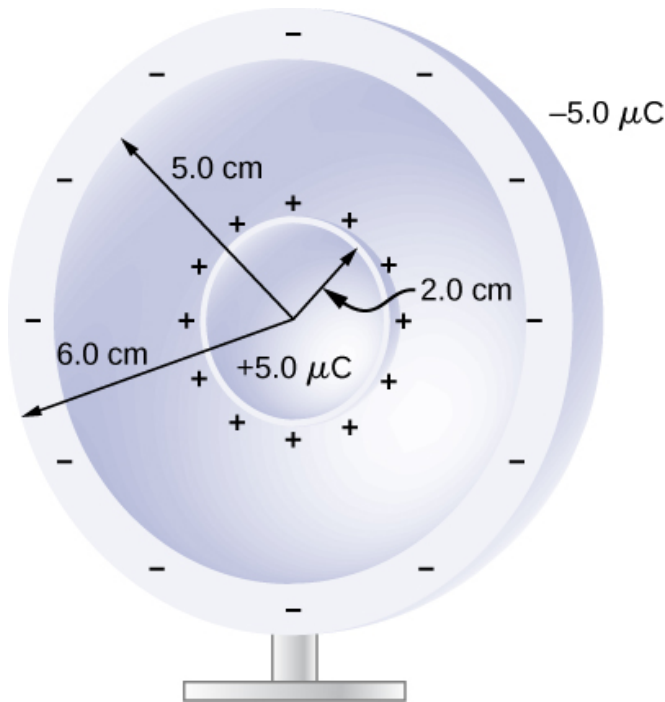
A very large sheet of insulating material has had an excess of electrons placed on it to a surface charge density of -3.00 nC/m^2 . (a) As the distance from the sheet increases, does the potential increase or decrease? Can you explain why without any calculations? Does the location of your reference point matter? (b) What is the shape of the equipotential surfaces? (c) What is the spacing between surfaces that differ by 1.00 V?

Solution:

a. increases; the constant (negative) electric field has this effect, the reference point only matters for magnitude; b. they are planes parallel to the sheet; c. 0.006 m/V

Exercise:**Problem:**

A metallic sphere of radius 2.0 cm is charged with $+5.0\text{-}\mu\text{C}$ charge, which spreads on the surface of the sphere uniformly. The metallic sphere stands on an insulated stand and is surrounded by a larger metallic spherical shell, of inner radius 5.0 cm and outer radius 6.0 cm. Now, a charge of $-5.0\text{-}\mu\text{C}$ is placed on the inside of the spherical shell, which spreads out uniformly on the inside surface of the shell. If potential is zero at infinity, what is the potential of (a) the spherical shell, (b) the sphere, (c) the space between the two, (d) inside the sphere, and (e) outside the shell?



Exercise:

Problem:

Two large charged plates of charge density $\pm 30 \mu\text{C}/\text{m}^2$ face each other at a separation of 5.0 mm. (a) Find the electric potential everywhere. (b) An electron is released from rest at the negative plate; with what speed will it strike the positive plate?

Solution:

- a. from the previous chapter, the electric field has magnitude $\frac{\sigma}{\epsilon_0}$ in the region between the plates and zero outside; defining the negatively charged plate to be at the origin and zero potential, with the positively charged plate located at +5 mm in the z-direction, $V = 1.7 \times 10^4 \text{ V}$ so the potential is 0 for $z < 0$, $1.7 \times 10^4 \text{ V} \left(\frac{z}{5 \text{ mm}} \right)$ for $0 \leq z \leq 5 \text{ mm}$, $1.7 \times 10^4 \text{ V}$ for $z > 5 \text{ mm}$;
- b. $qV = \frac{1}{2}mv^2 \rightarrow v = 7.7 \times 10^7 \text{ m/s}$

Exercise:

Problem:

A long cylinder of aluminum of radius R meters is charged so that it has a uniform charge per unit length on its surface of λ .

- (a) Find the electric field inside and outside the cylinder. (b) Find the electric potential inside and outside the cylinder. (c) Plot electric field and electric potential as a function of distance from the center of the rod.

Exercise:

Problem:

Two parallel plates 10 cm on a side are given equal and opposite charges of magnitude 5.0×10^{-9} C. The plates are 1.5 mm apart. What is the potential difference between the plates?

Solution:

$$V = 85 \text{ V}$$

Exercise:

Problem:

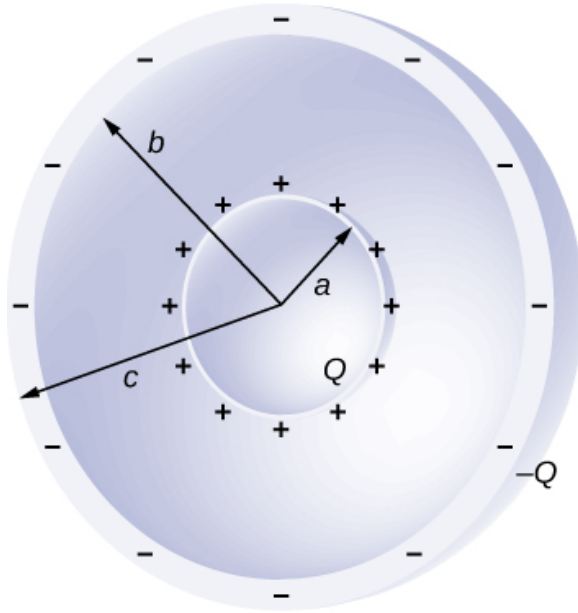
The surface charge density on a long straight metallic pipe is σ . What is the electric potential outside and inside the pipe? Assume the pipe has a diameter of $2a$.



Exercise:

Problem:

Concentric conducting spherical shells carry charges Q and $-Q$, respectively. The inner shell has negligible thickness. What is the potential difference between the shells?

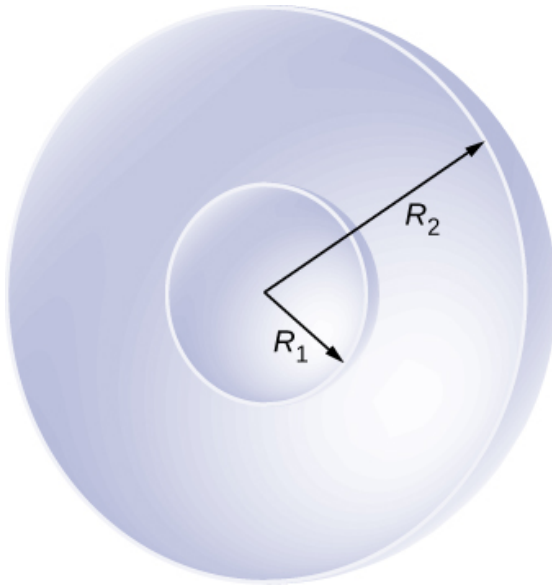


Solution:

In the region $a \leq r \leq b$, $\vec{E} = \frac{kQ}{r^2} \hat{r}$, and E is zero elsewhere; hence, the potential difference is $V = kQ \left(\frac{1}{a} - \frac{1}{b} \right)$.

Exercise:**Problem:**

Shown below are two concentric spherical shells of negligible thicknesses and radii R_1 and R_2 . The inner and outer shell carry net charges q_1 and q_2 , respectively, where both q_1 and q_2 are positive. What is the electric potential in the regions (a) $r < R_1$, (b) $R_1 < r < R_2$, and (c) $r > R_2$?



Exercise:

Problem:

A solid cylindrical conductor of radius a is surrounded by a concentric cylindrical shell of inner radius b . The solid cylinder and the shell carry charges Q and $-Q$, respectively. Assuming that the length L of both conductors is much greater than a or b , what is the potential difference between the two conductors?

Solution:

From previous results $V_P - V_R = -2k\lambda \ln \frac{s_P}{s_R}$, note that b is a very convenient location to define the zero level of potential: $\Delta V = -2k \frac{Q}{L} \ln \frac{a}{b}$.

Glossary

equipotential line

two-dimensional representation of an equipotential surface

equipotential surface

surface (usually in three dimensions) on which all points are at the same potential

grounding

process of attaching a conductor to the earth to ensure that there is no potential difference between it and Earth

Applications of Electrostatics

By the end of this section, you will be able to:

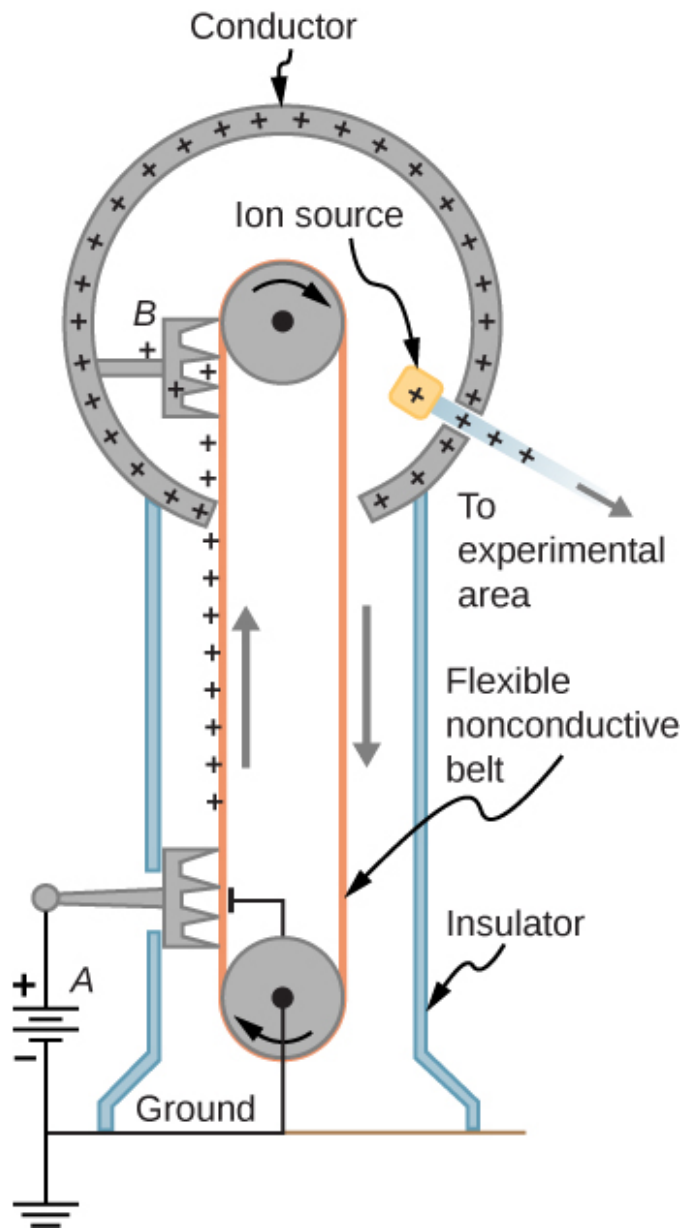
- Describe some of the many practical applications of electrostatics, including several printing technologies
- Relate these applications to Newton's second law and the electric force

The study of electrostatics has proven useful in many areas. This module covers just a few of the many applications of electrostatics.

The Van de Graaff Generator

Van de Graaff generators (or Van de Graaffs) are not only spectacular devices used to demonstrate high voltage due to static electricity—they are also used for serious research. The first was built by Robert Van de Graaff in 1931 (based on original suggestions by Lord Kelvin) for use in nuclear physics research. [\[link\]](#) shows a schematic of a large research version. Van de Graaffs use both smooth and pointed surfaces, and conductors and insulators to generate large static charges and, hence, large voltages.

A very large excess charge can be deposited on the sphere because it moves quickly to the outer surface. Practical limits arise because the large electric fields polarize and eventually ionize surrounding materials, creating free charges that neutralize excess charge or allow it to escape. Nevertheless, voltages of 15 million volts are well within practical limits.

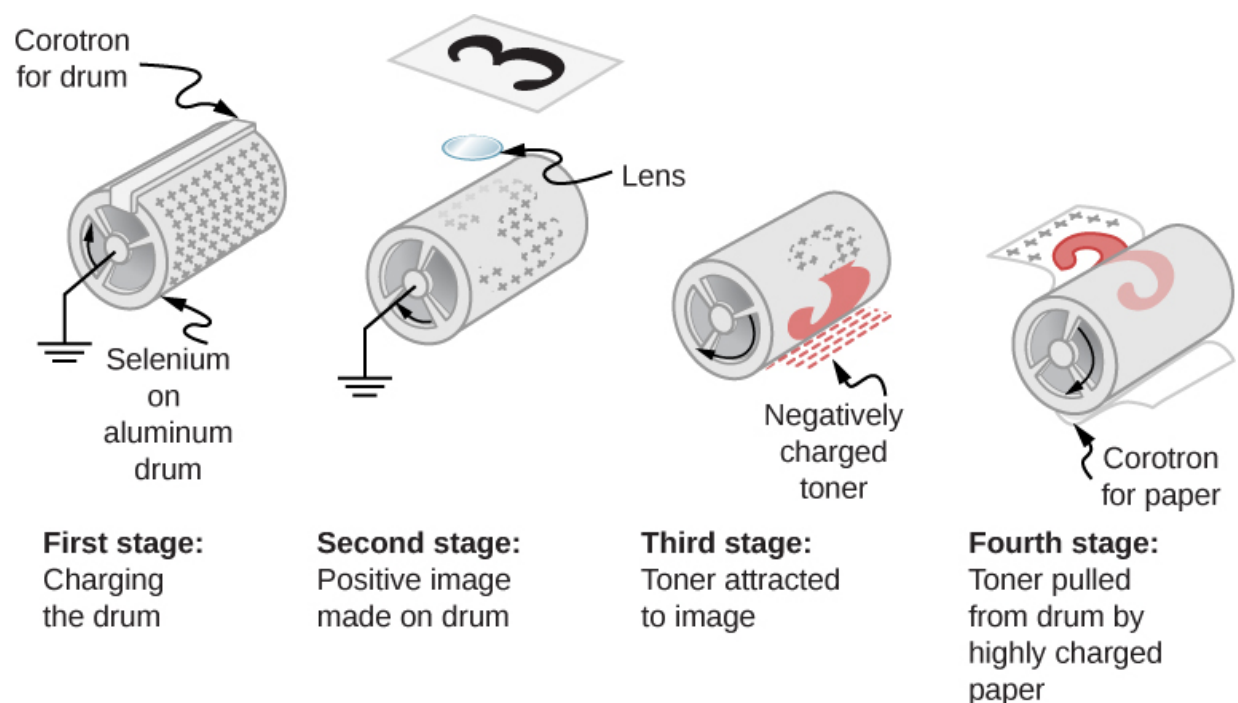


Schematic of Van de Graaff generator. A battery (A) supplies excess positive charge to a pointed conductor, the points of which spray the charge onto a moving insulating belt near the bottom. The pointed conductor (B) on top in the large sphere picks up the charge. (The induced electric field at the points is so large that it removes the charge

from the belt.) This can be done because the charge does not remain inside the conducting sphere but moves to its outside surface. An ion source inside the sphere produces positive ions, which are accelerated away from the positive sphere to high velocities.

Xerography

Most copy machines use an electrostatic process called **xerography**—a word coined from the Greek words *xeros* for dry and *graphos* for writing. The heart of the process is shown in simplified form in [\[link\]](#).



Xerography is a dry copying process based on electrostatics. The major steps in the process are the charging of the photoconducting drum, transfer of an image, creating a positive charge duplicate, attraction of

toner to the charged parts of the drum, and transfer of toner to the paper. Not shown are heat treatment of the paper and cleansing of the drum for the next copy.

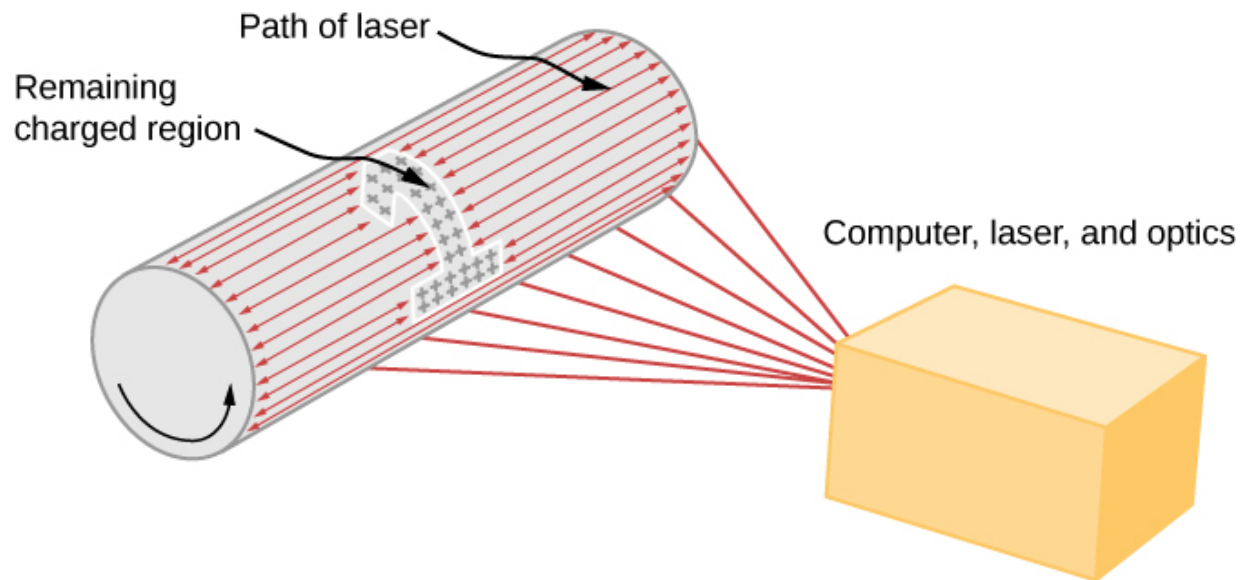
A selenium-coated aluminum drum is sprayed with positive charge from points on a device called a corotron. Selenium is a substance with an interesting property—it is a **photoconductor**. That is, selenium is an insulator when in the dark and a conductor when exposed to light.

In the first stage of the xerography process, the conducting aluminum drum is grounded so that a negative charge is induced under the thin layer of uniformly positively charged selenium. In the second stage, the surface of the drum is exposed to the image of whatever is to be copied. In locations where the image is light, the selenium becomes conducting, and the positive charge is neutralized. In dark areas, the positive charge remains, so the image has been transferred to the drum.

The third stage takes a dry black powder, called toner, and sprays it with a negative charge so that it is attracted to the positive regions of the drum. Next, a blank piece of paper is given a greater positive charge than on the drum so that it will pull the toner from the drum. Finally, the paper and electrostatically held toner are passed through heated pressure rollers, which melt and permanently adhere the toner to the fibers of the paper.

Laser Printers

Laser printers use the xerographic process to make high-quality images on paper, employing a laser to produce an image on the photoconducting drum as shown in [\[link\]](#). In its most common application, the laser printer receives output from a computer, and it can achieve high-quality output because of the precision with which laser light can be controlled. Many laser printers do significant information processing, such as making sophisticated letters or fonts, and in the past may have contained a computer more powerful than the one giving them the raw data to be printed.

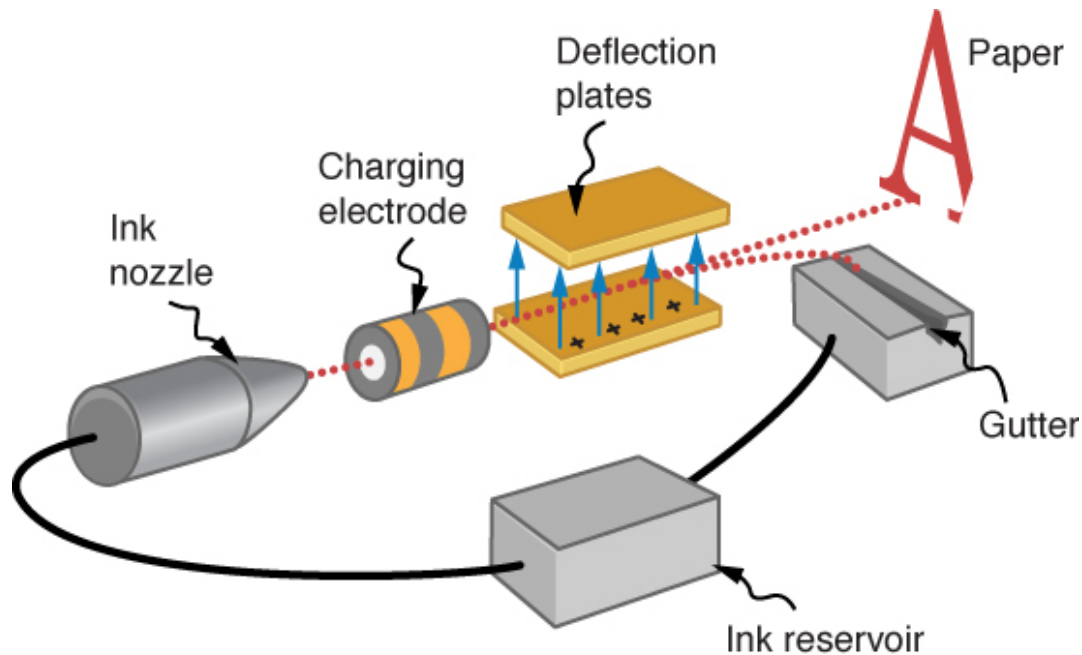


In a laser printer, a laser beam is scanned across a photoconducting drum, leaving a positively charged image. The other steps for charging the drum and transferring the image to paper are the same as in xerography. Laser light can be very precisely controlled, enabling laser printers to produce high-quality images.

Ink Jet Printers and Electrostatic Painting

The **ink jet printer**, commonly used to print computer-generated text and graphics, also employs electrostatics. A nozzle makes a fine spray of tiny ink droplets, which are then given an electrostatic charge ([link](#)).

Once charged, the droplets can be directed, using pairs of charged plates, with great precision to form letters and images on paper. Ink jet printers can produce color images by using a black jet and three other jets with primary colors, usually cyan, magenta, and yellow, much as a color television produces color. (This is more difficult with xerography, requiring multiple drums and toners.)



The nozzle of an ink-jet printer produces small ink droplets, which are sprayed with electrostatic charge. Various computer-driven devices are then used to direct the droplets to the correct positions on a page.

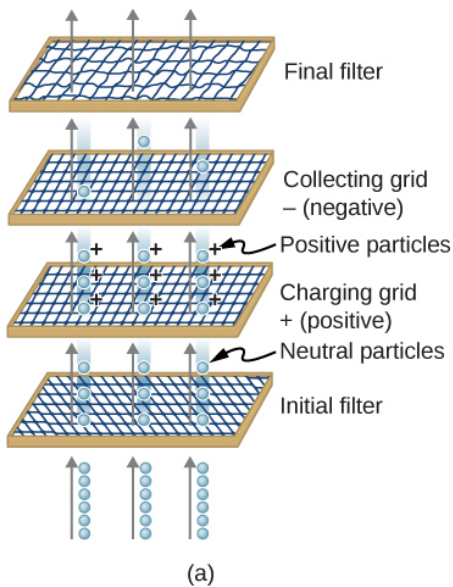
Electrostatic painting employs electrostatic charge to spray paint onto oddly shaped surfaces. Mutual repulsion of like charges causes the paint to fly away from its source. Surface tension forms drops, which are then attracted by unlike charges to the surface to be painted. Electrostatic painting can reach hard-to-get-to places, applying an even coat in a controlled manner. If the object is a conductor, the electric field is perpendicular to the surface, tending to bring the drops in perpendicularly. Corners and points on conductors will receive extra paint. Felt can similarly be applied.

Smoke Precipitators and Electrostatic Air Cleaning

Another important application of electrostatics is found in air cleaners, both large and small. The electrostatic part of the process places excess (usually positive) charge on smoke, dust, pollen, and other particles in the air and then

passes the air through an oppositely charged grid that attracts and retains the charged particles ([\[link\]](#))

Large **electrostatic precipitators** are used industrially to remove over 99 % of the particles from stack gas emissions associated with the burning of coal and oil. Home precipitators, often in conjunction with the home heating and air conditioning system, are very effective in removing polluting particles, irritants, and allergens.



(a) Schematic of an electrostatic precipitator. Air is passed through grids of opposite charge. The first grid charges airborne particles, while the second attracts and collects them. (b) The dramatic effect of electrostatic precipitators is seen by the absence of smoke from this power plant. (credit b: modification of work by “Cmdalgleish”/Wikimedia Commons)

Summary

- Electrostatics is the study of electric fields in static equilibrium.
- In addition to research using equipment such as a Van de Graaff generator, many practical applications of electrostatics exist, including

photocopiers, laser printers, ink jet printers, and electrostatic air filters.

Key Equations

Potential energy of a two-charge system	$U(r) = k \frac{qQ}{r}$
Work done to assemble a system of charges	$W_{12 \dots N} = \frac{k}{2} \sum_i^N \sum_j^N \frac{q_i q_j}{r_{ij}} \text{ for } i \neq j$
Potential difference	$\Delta V = \frac{\Delta U}{q} \text{ or } \Delta U = q \Delta V$
Electric potential	$V = \frac{U}{q} = - \int_R^P \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$
Potential difference between two points	$\Delta V_{BA} = V_B - V_A = - \int_A^B \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}$
Electric potential of a point charge	$V = \frac{kq}{r}$
Electric potential of a system of point charges	$V_P = k \sum_1^N \frac{q_i}{r_i}$
Electric dipole moment	$\vec{\mathbf{p}} = q\vec{\mathbf{d}}$
Electric potential due to a dipole	$V_P = k \frac{\vec{\mathbf{p}} \cdot \hat{\mathbf{r}}}{r^2}$
Electric potential of a	

continuous charge distribution	$V_P = k \int \frac{dq}{r}$
Electric field components	$E_x = -\frac{\partial V}{\partial x}, E_y = -\frac{\partial V}{\partial y}, E_z = -\frac{\partial V}{\partial z}$
Del operator in Cartesian coordinates	$\vec{\nabla} = \hat{\mathbf{i}} \frac{\partial}{\partial x} + \hat{\mathbf{j}} \frac{\partial}{\partial y} + \hat{\mathbf{k}} \frac{\partial}{\partial z}$
Electric field as gradient of potential	$\vec{\mathbf{E}} = -\vec{\nabla} V$
Del operator in cylindrical coordinates	$\vec{\nabla} = \hat{\mathbf{r}} \frac{\partial}{\partial r} + \hat{\boldsymbol{\varphi}} \frac{1}{r} \frac{\partial}{\partial \varphi} + \hat{\mathbf{z}} \frac{\partial}{\partial z}$
Del operator in spherical coordinates	$\vec{\nabla} = \hat{\mathbf{r}} \frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}} \frac{1}{r} \frac{\partial}{\partial \theta} + \hat{\boldsymbol{\varphi}} \frac{1}{r \sin \theta} \frac{\partial}{\partial \varphi}$

Conceptual Questions

Exercise:

Problem:

Why are the metal support rods for satellite network dishes generally grounded?

Solution:

So that lightning striking them goes into the ground instead of the television equipment.

Exercise:

Problem:

(a) Why are fish reasonably safe in an electrical storm? (b) Why are swimmers nonetheless ordered to get out of the water in the same circumstance?

Exercise:

Problem:

What are the similarities and differences between the processes in a photocopier and an electrostatic precipitator?

Solution:

They both make use of static electricity to stick small particles to another surface. However, the precipitator has to charge a wide variety of particles, and is not designed to make sure they land in a particular place.

Exercise:**Problem:**

About what magnitude of potential is used to charge the drum of a photocopy machine? A web search for “xerography” may be of use.

Problems**Exercise:****Problem:**

(a) What is the electric field 5.00 m from the center of the terminal of a Van de Graaff with a 3.00-mC charge, noting that the field is equivalent to that of a point charge at the center of the terminal? (b) At this distance, what force does the field exert on a 2.00- μC charge on the Van de Graaff's belt?

Exercise:**Problem:**

(a) What is the direction and magnitude of an electric field that supports the weight of a free electron near the surface of Earth? (b) Discuss what the small value for this field implies regarding the relative strength of the gravitational and electrostatic forces.

Solution:

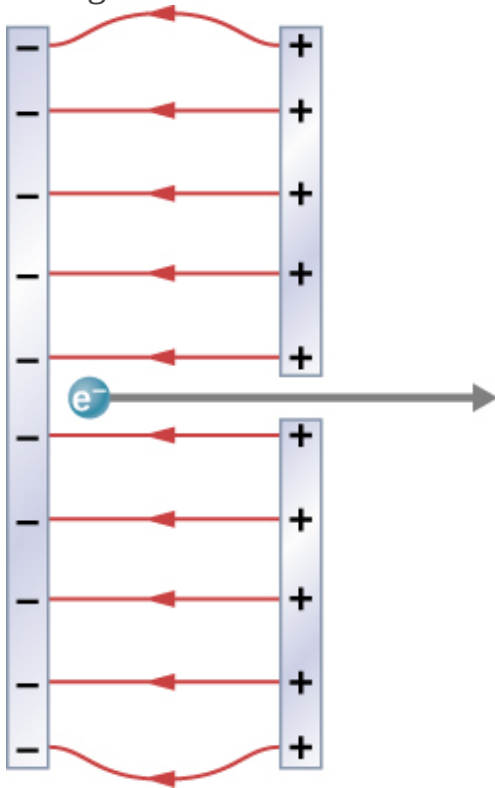
a. $F = 5.58 \times 10^{-11} \text{ N/C}$;

The electric field is towards the surface of Earth. b. The coulomb force is much stronger than gravity.

Exercise:

Problem:

A simple and common technique for accelerating electrons is shown in [\[link\]](#), where there is a uniform electric field between two plates. Electrons are released, usually from a hot filament, near the negative plate, and there is a small hole in the positive plate that allows the electrons to continue moving. (a) Calculate the acceleration of the electron if the field strength is $2.50 \times 10^4 \text{ N/C}$. (b) Explain why the electron will not be pulled back to the positive plate once it moves through the hole.



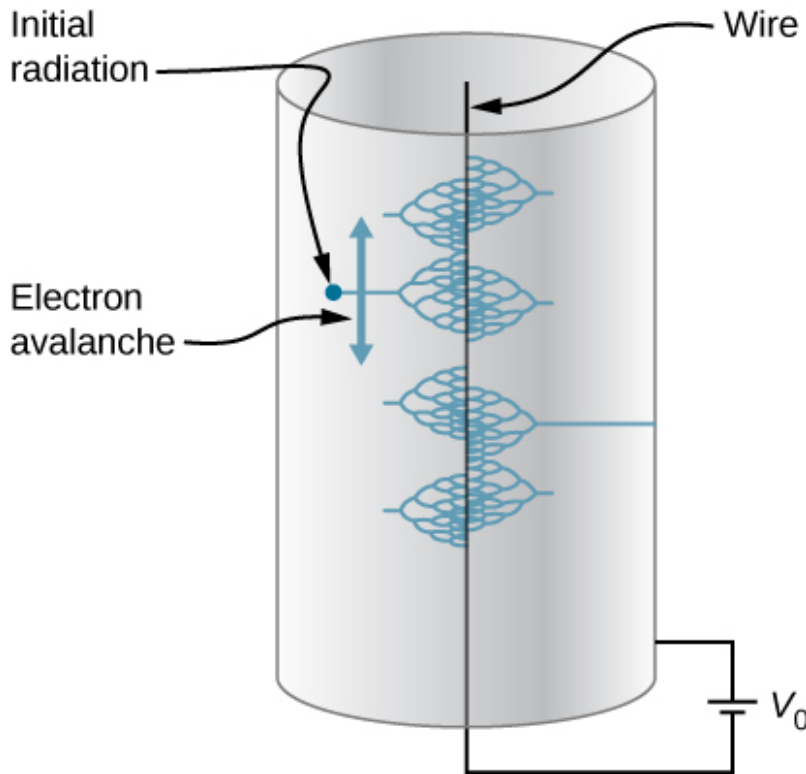
Parallel conducting plates with opposite charges on them create a relatively uniform electric field used to accelerate electrons to

to accelerate electrons to the right. Those that go through the hole can be used to make a TV or computer screen glow or to produce X- rays.

Exercise:

Problem:

In a Geiger counter, a thin metallic wire at the center of a metallic tube is kept at a high voltage with respect to the metal tube. Ionizing radiation entering the tube knocks electrons off gas molecules or sides of the tube that then accelerate towards the center wire, knocking off even more electrons. This process eventually leads to an avalanche that is detectable as a current. A particular Geiger counter has a tube of radius R and the inner wire of radius a is at a potential of V_0 volts with respect to the outer metal tube. Consider a point P at a distance s from the center wire and far away from the ends. (a) Find a formula for the electric field at a point P inside using the infinite wire approximation. (b) Find a formula for the electric potential at a point P inside. (c) Use $V_0 = 900 \text{ V}$, $a = 3.00 \text{ mm}$, $R = 2.00 \text{ cm}$, and find the value of the electric field at a point 1.00 cm from the center.



Solution:

We know from the Gauss's law chapter that the electric field for an infinite line charge is $\vec{E}_P = 2k\lambda \frac{1}{s} \hat{s}$, and from earlier in this chapter that the potential of a wire-cylinder system of this sort is $V_P = -2k\lambda \ln \frac{s_P}{R}$ by integration. We are not given λ , but we are given a fixed V_0 ; thus, we know that $V_0 = -2k\lambda \ln \frac{a}{R}$ and hence $\lambda = -\frac{V_0}{2k \ln(\frac{a}{R})}$. We may

substitute this back in to find a. $\vec{E}_P = -\frac{V_0}{\ln(\frac{a}{R})} \frac{1}{s} \hat{s}$; b. $V_P = V_0 \frac{\ln(\frac{s_P}{R})}{\ln(\frac{a}{R})}$; c. $4.74 \times 10^4 \text{ N/C}$

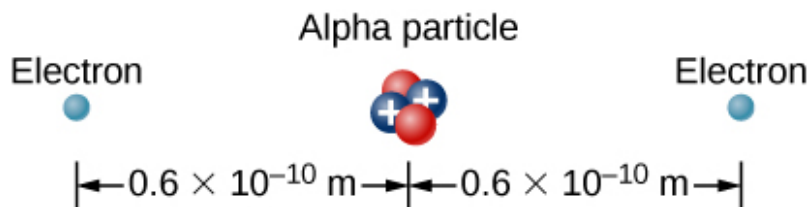
Exercise:

Problem:

The practical limit to an electric field in air is about $3.00 \times 10^6 \text{ N/C}$. Above this strength, sparking takes place because air begins to ionize. (a) At this electric field strength, how far would a proton travel before hitting the speed of light (ignore relativistic effects)? (b) Is it practical to leave air in particle accelerators?

Exercise:**Problem:**

To form a helium atom, an alpha particle that contains two protons and two neutrons is fixed at one location, and two electrons are brought in from far away, one at a time. The first electron is placed at $0.600 \times 10^{-10} \text{ m}$ from the alpha particle and held there while the second electron is brought to $0.600 \times 10^{-10} \text{ m}$ from the alpha particle on the other side from the first electron. See the final configuration below. (a) How much work is done in each step? (b) What is the electrostatic energy of the alpha particle and two electrons in the final configuration?

**Solution:**

a. $U_1 = 7.68 \times 10^{-18} \text{ J}$;
 $U_2 = 5.76 \times 10^{-18} \text{ J}$;
 b. $U_1 + U_2 = -1.34 \times 10^{-17} \text{ J}$

Exercise:**Problem:**

Find the electrostatic energy of eight equal charges ($+3 \mu\text{C}$) each fixed at the corners of a cube of side 2 cm.

Exercise:**Problem:**

The probability of fusion occurring is greatly enhanced when appropriate nuclei are brought close together, but mutual Coulomb repulsion must be overcome. This can be done using the kinetic energy of high-temperature gas ions or by accelerating the nuclei toward one another. (a) Calculate the potential energy of two singly charged nuclei separated by 1.00×10^{-12} m. (b) At what temperature will atoms of a gas have an average kinetic energy equal to this needed electrical potential energy?

Solution:

a. $U = 2.30 \times 10^{-16}$ J;

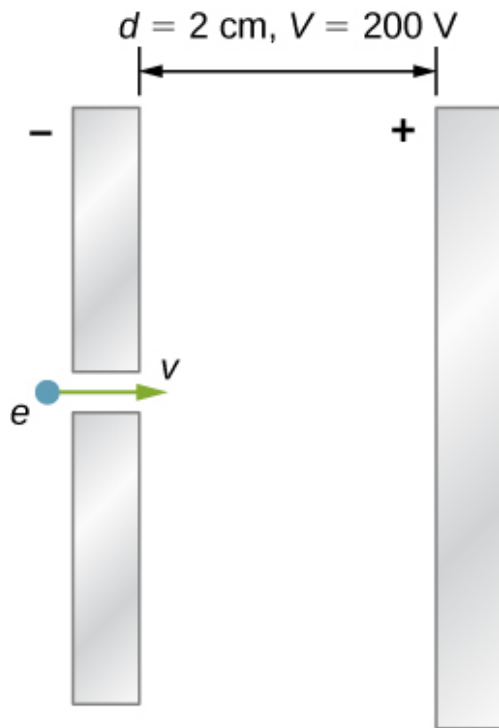
b. $\bar{K} = \frac{3}{2}kT \rightarrow T = 1.11 \times 10^7$ K

Exercise:**Problem:**

A bare helium nucleus has two positive charges and a mass of 6.64×10^{-27} kg. (a) Calculate its kinetic energy in joules at 2.00 % of the speed of light. (b) What is this in electron-volts? (c) What voltage would be needed to obtain this energy?

Exercise:**Problem:**

An electron enters a region between two large parallel plates made of aluminum separated by a distance of 2.0 cm and kept at a potential difference of 200 V. The electron enters through a small hole in the negative plate and moves toward the positive plate. At the time the electron is near the negative plate, its speed is 4.0×10^5 m/s. Assume the electric field between the plates to be uniform, and find the speed of electron at (a) 0.10 cm, (b) 0.50 cm, (c) 1.0 cm, and (d) 1.5 cm from the negative plate, and (e) immediately before it hits the positive plate.



Solution:

a. $1.9 \times 10^6 \text{ m/s}$; b. $4.2 \times 10^6 \text{ m/s}$; c. $5.9 \times 10^6 \text{ m/s}$; d. $7.3 \times 10^6 \text{ m/s}$; e. $8.4 \times 10^6 \text{ m/s}$

Exercise:

Problem:

How far apart are two conducting plates that have an electric field strength of $4.50 \times 10^3 \text{ V/m}$ between them, if their potential difference is 15.0 kV?

Exercise:

Problem:

(a) Will the electric field strength between two parallel conducting plates exceed the breakdown strength of dry air, which is $3.00 \times 10^6 \text{ V/m}$, if the plates are separated by 2.00 mm and a potential difference of $5.0 \times 10^3 \text{ V}$ is applied? (b) How close together can the plates be with this applied voltage?

Solution:

a.

$$E = 2.5 \times 10^6 \text{ V/m} < 3 \times 10^6 \text{ V/m}$$

No, the field strength is smaller than the breakdown strength for air.

;

b. $d = 1.7 \text{ mm}$

Exercise:**Problem:**

Membrane walls of living cells have surprisingly large electric fields across them due to separation of ions. What is the voltage across an 8.00-nm-thick membrane if the electric field strength across it is 5.50 MV/m? You may assume a uniform electric field.

Exercise:**Problem:**

A double charged ion is accelerated to an energy of 32.0 keV by the electric field between two parallel conducting plates separated by 2.00 cm. What is the electric field strength between the plates?

Solution:

$$\begin{aligned} K_f &= qV_{AB} = qEd \rightarrow \\ E &= 8.00 \times 10^5 \text{ V/m} \end{aligned}$$

Exercise:**Problem:**

The temperature near the center of the Sun is thought to be 15 million degrees Celsius ($1.5 \times 10^7 \text{ }^\circ\text{C}$) (or kelvin). Through what voltage must a singly charged ion be accelerated to have the same energy as the average kinetic energy of ions at this temperature?

Exercise:

Problem:

A lightning bolt strikes a tree, moving 20.0 C of charge through a potential difference of 1.00×10^2 MV. (a) What energy was dissipated? (b) What mass of water could be raised from 15 °C to the boiling point and then boiled by this energy? (c) Discuss the damage that could be caused to the tree by the expansion of the boiling steam.

Solution:

- a. Energy = 2.00×10^9 J;
- b. $Q = m(c\Delta T + L_v)$;
 $m = 766$ kg
- c. The expansion of the steam upon boiling can literally blow the tree apart.

Exercise:**Problem:**

What is the potential 0.530×10^{-10} m from a proton (the average distance between the proton and electron in a hydrogen atom)?

Exercise:**Problem:**

(a) A sphere has a surface uniformly charged with 1.00 C. At what distance from its center is the potential 5.00 MV? (b) What does your answer imply about the practical aspect of isolating such a large charge?

Solution:

- a. $V = \frac{kQ}{r} \rightarrow r = 1.80$ km; b. A 1-C charge is a very large amount of charge; a sphere of 1.80 km is impractical.

Exercise:

Problem:

What are the sign and magnitude of a point charge that produces a potential of -2.00 V at a distance of 1.00 mm ?

Exercise:**Problem:**

In one of the classic nuclear physics experiments at the beginning of the twentieth century, an alpha particle was accelerated toward a gold nucleus, and its path was substantially deflected by the Coulomb interaction. If the energy of the doubly charged alpha nucleus was 5.00 MeV , how close to the gold nucleus (79 protons) could it come before being deflected?

Solution:

The alpha particle approaches the gold nucleus until its original energy is converted to potential energy. $5.00\text{ MeV} = 8.00 \times 10^{-13}\text{ J}$, so

$$E_0 = \frac{qkQ}{r} \rightarrow$$

$$r = 4.54 \times 10^{-14}\text{ m}$$

(Size of gold nucleus is about $7 \times 10^{-15}\text{ m}$).

Additional Problems**Exercise:****Problem:**

A 12.0-V battery-operated bottle warmer heats 50.0 g of glass, $2.50 \times 10^2\text{ g}$ of baby formula, and $2.00 \times 10^2\text{ g}$ of aluminum from 20.0°C to 90.0°C . (a) How much charge is moved by the battery? (b) How many electrons per second flow if it takes 5.00 min to warm the formula? (*Hint:* Assume that the specific heat of baby formula is about the same as the specific heat of water.)

Exercise:

Problem:

A battery-operated car uses a 12.0-V system. Find the charge the batteries must be able to move in order to accelerate the 750 kg car from rest to 25.0 m/s, make it climb a 2.00×10^2 -m high hill, and finally cause it to travel at a constant 25.0 m/s while climbing with 5.00×10^2 -N force for an hour.

Solution:

$$E_{\text{tot}} = 4.67 \times 10^7 \text{ J}$$

$$E_{\text{tot}} = qV \rightarrow q = \frac{E_{\text{tot}}}{V} = 3.89 \times 10^6 \text{ C}$$

Exercise:**Problem:**

(a) Find the voltage near a 10.0 cm diameter metal sphere that has 8.00 C of excess positive charge on it. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:**Problem:**

A uniformly charged half-ring of radius 10 cm is placed on a nonconducting table. It is found that 3.0 cm above the center of the half-ring the potential is -3.0 V with respect to zero potential at infinity. How much charge is in the half-ring?

Solution:

$$V_P = k \frac{q_{\text{tot}}}{\sqrt{z^2 + R^2}} \rightarrow q_{\text{tot}} = -3.5 \times 10^{-11} \text{ C}$$

Exercise:

Problem:

A glass ring of radius 5.0 cm is painted with a charged paint such that the charge density around the ring varies continuously given by the following function of the polar angle θ , $\lambda = (3.0 \times 10^{-6} \text{ C/m}) \cos^2 \theta$. Find the potential at a point 15 cm above the center.

Exercise:**Problem:**

A CD disk of radius ($R = 3.0 \text{ cm}$) is sprayed with a charged paint so that the charge varies continually with radial distance r from the center in the following manner: $\sigma = -(6.0 \text{ C/m})r/R$.

Find the potential at a point 4 cm above the center.

Solution:

$$V_P = -2.2 \text{ GV}$$

Exercise:**Problem:**

(a) What is the final speed of an electron accelerated from rest through a voltage of 25.0 MV by a negatively charged Van de Graff terminal? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:**Problem:**

A large metal plate is charged uniformly to a density of $\sigma = 2.0 \times 10^{-9} \text{ C/m}^2$. How far apart are the equipotential surfaces that represent a potential difference of 25 V?

Solution:

Recall from the previous chapter that the electric field $E_P = \frac{\sigma_0}{2\epsilon_0}$ is uniform throughout space, and that for uniform fields we have

$E = -\frac{\Delta V}{\Delta z}$ for the relation. Thus, we get $\frac{\sigma}{2\epsilon_0} = \frac{\Delta V}{\Delta z} \rightarrow \Delta z = 0.22 \text{ m}$ for the distance between 25-V equipotentials.

Exercise:

Problem:

Your friend gets really excited by the idea of making a lightning rod or maybe just a sparking toy by connecting two spheres as shown in [\[link\]](#), and making R_2 so small that the electric field is greater than the dielectric strength of air, just from the usual 150 V/m electric field near the surface of the Earth. If R_1 is 10 cm, how small does R_2 need to be, and does this seem practical? (*Hint: recall the calculation for electric field at the surface of a conductor from [Gauss's Law](#).*)

Exercise:

Problem:

(a) Find $x \gg L$ limit of the potential of a finite uniformly charged rod and show that it coincides with that of a point charge formula. (b) Why would you expect this result?

Solution:

a. Take the result from [\[link\]](#), divide both the numerator and the denominator by x , take the limit of that, and then apply a Taylor expansion to the resulting log to get: $V_P \approx k\lambda \frac{L}{x}$; b. which is the result we expect, because at great distances, this should look like a point charge of $q = \lambda L$

Exercise:

Problem:

A small spherical pith ball of radius 0.50 cm is painted with a silver paint and then $-10 \mu\text{C}$ of charge is placed on it. The charged pith ball is put at the center of a gold spherical shell of inner radius 2.0 cm and outer radius 2.2 cm. (a) Find the electric potential of the gold shell with respect to zero potential at infinity. (b) How much charge should you put on the gold shell if you want to make its potential 100 V?

Exercise:**Problem:**

Two parallel conducting plates, each of cross-sectional area 400 cm^2 , are 2.0 cm apart and uncharged. If 1.0×10^{12} electrons are transferred from one plate to the other, (a) what is the potential difference between the plates? (b) What is the potential difference between the positive plate and a point 1.25 cm from it that is between the plates?

Solution:

$$\text{a. } V = 9.0 \times 10^3 \text{ V; b. } -9.0 \times 10^3 \text{ V} \left(\frac{1.25 \text{ cm}}{2.0 \text{ cm}} \right) = -5.7 \times 10^3 \text{ V}$$

Exercise:**Problem:**

A point charge of $q = 5.0 \times 10^{-8} \text{ C}$ is placed at the center of an uncharged spherical conducting shell of inner radius 6.0 cm and outer radius 9.0 cm . Find the electric potential at (a) $r = 4.0 \text{ cm}$, (b) $r = 8.0 \text{ cm}$, (c) $r = 12.0 \text{ cm}$.

Exercise:**Problem:**

Earth has a net charge that produces an electric field of approximately 150 N/C downward at its surface. (a) What is the magnitude and sign of the excess charge, noting the electric field of a conducting sphere is equivalent to a point charge at its center? (b) What acceleration will the field produce on a free electron near Earth's surface? (c) What mass object with a single extra electron will have its weight supported by this field?

Solution:

$$\begin{aligned} \text{a. } E &= \frac{kQ}{r^2} \rightarrow Q = -6.76 \times 10^5 \text{ C;} \\ F &= ma = qE \rightarrow \\ \text{b. } a &= \frac{qE}{m} = 2.63 \times 10^{13} \text{ m/s}^2 \text{ (upwards);} \end{aligned}$$

$$c. F = -mg = qE \rightarrow m = \frac{-qE}{g} = 2.45 \times 10^{-18} \text{ kg}$$

Exercise:

Problem:

Point charges of $25.0 \mu\text{C}$ and $45.0 \mu\text{C}$ are placed 0.500 m apart.

(a) At what point along the line between them is the electric field zero?

(b) What is the electric field halfway between them?

Exercise:

Problem:

What can you say about two charges q_1 and q_2 , if the electric field one-fourth of the way from q_1 to q_2 is zero?

Solution:

If the electric field is zero $\frac{1}{4}$ from the way of q_1 and q_2 , then we know from

$E = k \frac{Q}{r^2}$ that $|E_1| = |E_2| \rightarrow \frac{Kq_1}{x^2} = \frac{Kq_2}{(3x)^2}$ so that $\frac{q_2}{q_1} = \frac{(3x)^2}{x^2} = 9$;
the charge q_2 is 9 times larger than q_1 .

Exercise:

Problem:

Calculate the angular velocity ω of an electron orbiting a proton in the hydrogen atom, given the radius of the orbit is $0.530 \times 10^{-10} \text{ m}$. You may assume that the proton is stationary and the centripetal force is supplied by Coulomb attraction.

Exercise:

Problem:

An electron has an initial velocity of $5.00 \times 10^6 \text{ m/s}$ in a uniform $2.00 \times 10^5 \text{ N/C}$ electric field. The field accelerates the electron in the direction opposite to its initial velocity. (a) What is the direction of the electric field? (b) How far does the electron travel before coming to rest? (c) How long does it take the electron to come to rest? (d) What is the electron's velocity when it returns to its starting point?

Solution:

a. The field is in the direction of the electron's initial velocity.

b.

$$v^2 = v_0^2 + 2ax \rightarrow x = -\frac{v_0^2}{2a} (v = 0). \text{ Also, } F = ma = qE \rightarrow a = \frac{qE}{m},$$

$$x = 3.56 \times 10^{-4} \text{ m};$$

$$v_2 = v_0 + at \rightarrow t = -\frac{v_0 m}{qE} (v = 0),$$

c. $\therefore t = 1.42 \times 10^{-10} \text{ s};$

d. $v = -\left(\frac{2qEx}{m}\right)^{1/2} = -5.00 \times 10^6 \text{ m/s}$ (opposite its initial velocity)

Challenge Problems**Exercise:****Problem:**

Three Na^+ and three Cl^- ions are placed alternately and equally spaced around a circle of radius 50 nm. Find the electrostatic energy stored.

Exercise:**Problem:**

Look up (presumably online, or by dismantling an old device and making measurements) the magnitude of the potential deflection plates (and the space between them) in an ink jet printer. Then look up the speed with which the ink comes out the nozzle. Can you calculate the typical mass of an ink drop?

Solution:

Answers will vary. This appears to be proprietary information, and ridiculously difficult to find. Speeds will be 20 m/s or less, and there are claims of $\sim 10^{-7}$ grams for the mass of a drop.

Exercise:**Problem:**

Use the electric field of a finite sphere with constant volume charge density to calculate the electric potential, throughout space. Then check your results by calculating the electric field from the potential.

Exercise:**Problem:**

Calculate the electric field of a dipole throughout space from the potential.

Solution:

Apply $\vec{\mathbf{E}} = -\vec{\nabla}V$ with $\vec{\nabla} = \hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\theta}\frac{1}{r}\frac{\partial}{\partial\theta} + \hat{\varphi}\frac{1}{r\sin\theta}\frac{\partial}{\partial\varphi}$ to the potential calculated earlier, $V_P = k\frac{\vec{\mathbf{p}}\cdot\hat{\mathbf{r}}}{r^2}$ with $\vec{\mathbf{p}} = q\vec{\mathbf{d}}$, and assume that the axis of the dipole is aligned with the z-axis of the coordinate system.

Thus, the potential is $V_P = k\frac{q\vec{\mathbf{d}}\cdot\hat{\mathbf{r}}}{r^2} = k\frac{qd\cos\theta}{r^2}$.

$$\vec{\mathbf{E}} = 2kqd\left(\frac{\cos\theta}{r^3}\right)\hat{\mathbf{r}} + kqd\left(\frac{\sin\theta}{r^3}\right)\hat{\theta}$$

Glossary

electrostatic precipitators

filters that apply charges to particles in the air, then attract those charges to a filter, removing them from the airstream

ink jet printer

small ink droplets sprayed with an electric charge are controlled by electrostatic plates to create images on paper

photoconductor

substance that is an insulator until it is exposed to light, when it becomes a conductor

Van de Graaff generator

machine that produces a large amount of excess charge, used for experiments with high voltage

xerography

dry copying process based on electrostatics

Introduction

class="introduction"

The tree-like branch patterns in this clear Plexiglas® block are known as a Lichtenberg figure, named for the German physicist Georg Christof Lichtenberg (1742–1799), who was the first to study these patterns.

The “branches” are created by the dielectric breakdown produced by a strong electric field.

(credit: modification of work

by Bert
Hickman)



Capacitors are important components of electrical circuits in many electronic devices, including pacemakers, cell phones, and computers. In this chapter, we study their properties, and, over the next few chapters, we examine their function in combination with other circuit elements. By themselves, capacitors are often used to store electrical energy and release it when needed; with other circuit components, capacitors often act as part of a filter that allows some electrical signals to pass while blocking others. You can see why capacitors are considered one of the fundamental components of electrical circuits.

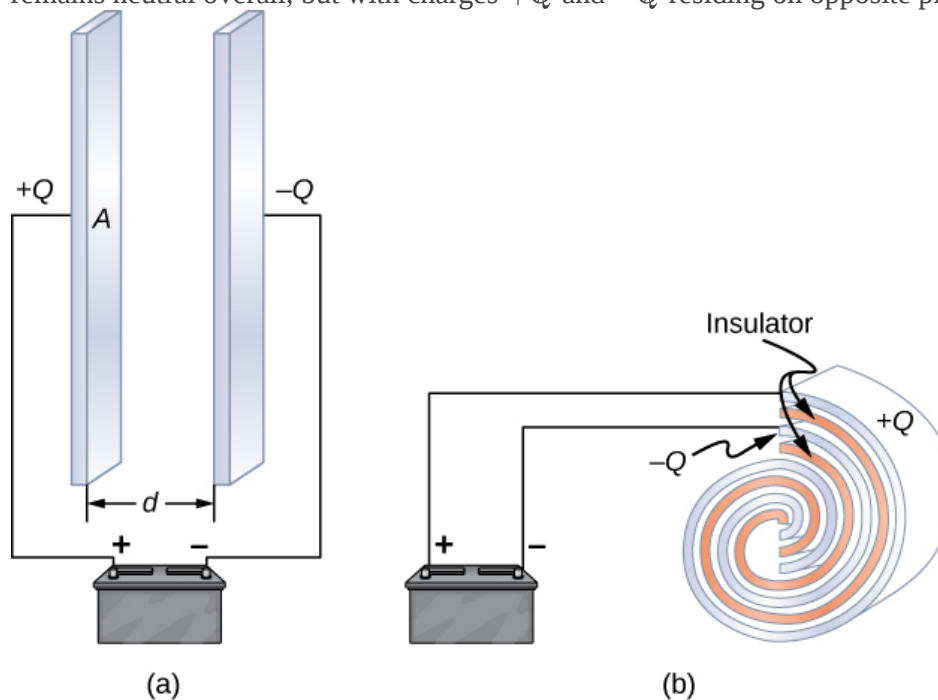
Capacitors and Capacitance

By the end of this section, you will be able to:

- Explain the concepts of a capacitor and its capacitance
- Describe how to evaluate the capacitance of a system of conductors

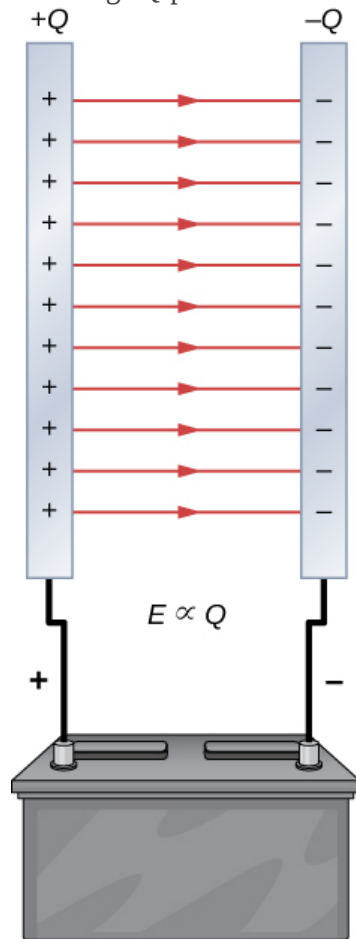
A **capacitor** is a device used to store electrical charge and electrical energy. Capacitors are generally with two electrical conductors separated by a distance. (Note that such electrical conductors are sometimes referred to as “electrodes,” but more correctly, they are “capacitor plates.”) The space between capacitors may simply be a vacuum, and, in that case, a capacitor is then known as a “vacuum capacitor.” However, the space is usually filled with an insulating material known as a **dielectric**. (You will learn more about dielectrics in the sections on dielectrics later in this chapter.) The amount of storage in a capacitor is determined by a property called *capacitance*, which you will learn more about a bit later in this section.

Capacitors have applications ranging from filtering static from radio reception to energy storage in heart defibrillators. Typically, commercial capacitors have two conducting parts close to one another but not touching, such as those in [\[link\]](#). Most of the time, a dielectric is used between the two plates. When battery terminals are connected to an initially uncharged capacitor, the battery potential moves a small amount of charge of magnitude Q from the positive plate to the negative plate. The capacitor remains neutral overall, but with charges $+Q$ and $-Q$ residing on opposite plates.



Both capacitors shown here were initially uncharged before being connected to a battery. They now have charges of $+Q$ and $-Q$ (respectively) on their plates. (a) A parallel-plate capacitor consists of two plates of opposite charge with area A separated by distance d . (b) A rolled capacitor has a dielectric material between its two conducting sheets (plates).

A system composed of two identical parallel-conducting plates separated by a distance is called a **parallel-plate capacitor** ([link](#)). The magnitude of the electrical field in the space between the parallel plates is $E = \sigma/\epsilon_0$, where σ denotes the surface charge density on one plate (recall that σ is the charge Q per the surface area A). Thus, the magnitude of the field is directly proportional to Q .



The charge separation in a capacitor shows that the charges remain on the surfaces of the capacitor plates. Electrical field lines in a parallel-plate capacitor begin with positive charges and end with negative charges. The magnitude of the electrical field in the space between the plates is in direct proportion to the

amount of charge on
the capacitor.

Capacitors with different physical characteristics (such as shape and size of their plates) store different amounts of charge for the same applied voltage V across their plates. The **capacitance** C of a capacitor is defined as the ratio of the maximum charge Q that can be stored in a capacitor to the applied voltage V across its plates. In other words, capacitance is the largest amount of charge per volt that can be stored on the device:

Note:

Equation:

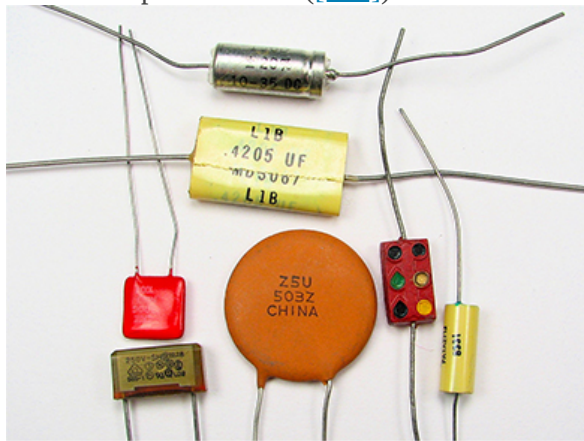
$$C = \frac{Q}{V}.$$

The SI unit of capacitance is the farad (F), named after Michael Faraday (1791–1867). Since capacitance is the charge per unit voltage, one farad is one coulomb per one volt, or

Equation:

$$1 \text{ F} = \frac{1 \text{ C}}{1 \text{ V}}.$$

By definition, a 1.0-F capacitor is able to store 1.0 C of charge (a very large amount of charge) when the potential difference between its plates is only 1.0 V. One farad is therefore a very large capacitance. Typical capacitance values range from picofarads ($1 \text{ pF} = 10^{-12} \text{ F}$) to millifarads ($1 \text{ mF} = 10^{-3} \text{ F}$), which also includes microfarads ($1 \mu\text{F} = 10^{-6} \text{ F}$). Capacitors can be produced in various shapes and sizes ([link](#)).



These are some typical capacitors used in electronic devices. A capacitor's size is not

necessarily related to its capacitance value.
(credit: Windell Oskay)

Calculation of Capacitance

We can calculate the capacitance of a pair of conductors with the standard approach that follows.

Note:

Calculating Capacitance

1. Assume that the capacitor has a charge Q .
2. Determine the electrical field $\vec{\mathbf{E}}$ between the conductors. If symmetry is present in the arrangement of conductors, you may be able to use Gauss's law for this calculation.
3. Find the potential difference between the conductors from

Equation:

$$V_B - V_A = - \int_A^B \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}},$$

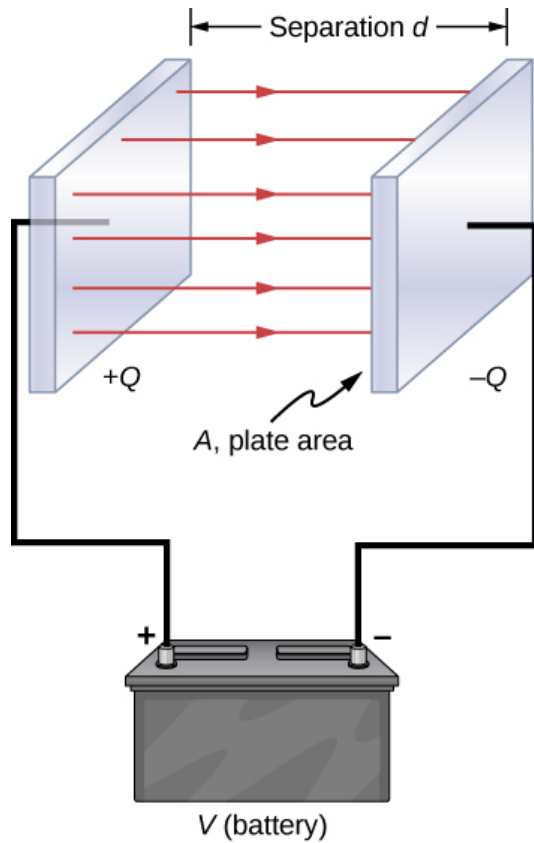
where the path of integration leads from one conductor to the other. The magnitude of the potential difference is then $V = |V_B - V_A|$.

4. With V known, obtain the capacitance directly from [\[link\]](#).

To show how this procedure works, we now calculate the capacitances of parallel-plate, spherical, and cylindrical capacitors. In all cases, we assume vacuum capacitors (empty capacitors) with no dielectric substance in the space between conductors.

Parallel-Plate Capacitor

The parallel-plate capacitor ([\[link\]](#)) has two identical conducting plates, each having a surface area A , separated by a distance d . When a voltage V is applied to the capacitor, it stores a charge Q , as shown. We can see how its capacitance may depend on A and d by considering characteristics of the Coulomb force. We know that force between the charges increases with charge values and decreases with the distance between them. We should expect that the bigger the plates are, the more charge they can store. Thus, C should be greater for a larger value of A . Similarly, the closer the plates are together, the greater the attraction of the opposite charges on them. Therefore, C should be greater for a smaller d .



In a parallel-plate capacitor with plates separated by a distance d , each plate has the same surface area A .

We define the surface charge density σ on the plates as
Equation:

$$\sigma = \frac{Q}{A}.$$

We know from previous chapters that when d is small, the electrical field between the plates is fairly uniform (ignoring edge effects) and that its magnitude is given by

Equation:

$$E = \frac{\sigma}{\epsilon_0},$$

where the constant ϵ_0 is the permittivity of free space, $\epsilon_0 = 8.85 \times 10^{-12} \text{ F/m}$. The SI unit of F/m is equivalent to $\text{C}^2/\text{N} \cdot \text{m}^2$. Since the electrical field \vec{E} between the plates is uniform, the potential difference between the plates is

Equation:

$$V = Ed = \frac{\sigma d}{\epsilon_0} = \frac{Qd}{\epsilon_0 A}.$$

Therefore [\[link\]](#) gives the capacitance of a parallel-plate capacitor as

Note:

Equation:

$$C = \frac{Q}{V} = \frac{Q}{Qd/\epsilon_0 A} = \epsilon_0 \frac{A}{d}.$$

Notice from this equation that capacitance is a function *only of the geometry* and what material fills the space between the plates (in this case, vacuum) of this capacitor. In fact, this is true not only for a parallel-plate capacitor, but for all capacitors: The capacitance is independent of Q or V . If the charge changes, the potential changes correspondingly so that Q/V remains constant.

Example:

Capacitance and Charge Stored in a Parallel-Plate Capacitor

(a) What is the capacitance of an empty parallel-plate capacitor with metal plates that each have an area of 1.00 m^2 , separated by 1.00 mm ? (b) How much charge is stored in this capacitor if a voltage of $3.00 \times 10^3 \text{ V}$ is applied to it?

Strategy

Finding the capacitance C is a straightforward application of [\[link\]](#). Once we find C , we can find the charge stored by using [\[link\]](#).

Solution

- a. Entering the given values into [\[link\]](#) yields

Equation:

$$C = \epsilon_0 \frac{A}{d} = \left(8.85 \times 10^{-12} \frac{\text{F}}{\text{m}} \right) \frac{1.00 \text{ m}^2}{1.00 \times 10^{-3} \text{ m}} = 8.85 \times 10^{-9} \text{ F} = 8.85 \text{ nF}.$$

This small capacitance value indicates how difficult it is to make a device with a large capacitance.

- b. Inverting [\[link\]](#) and entering the known values into this equation gives

Equation:

$$Q = CV = (8.85 \times 10^{-9} \text{ F})(3.00 \times 10^3 \text{ V}) = 26.6 \mu\text{C}.$$

Significance

This charge is only slightly greater than those found in typical static electricity applications. Since air breaks down (becomes conductive) at an electrical field strength of about 3.0 MV/m , no more charge

can be stored on this capacitor by increasing the voltage.

Example:

A 1-F Parallel-Plate Capacitor

Suppose you wish to construct a parallel-plate capacitor with a capacitance of 1.0 F. What area must you use for each plate if the plates are separated by 1.0 mm?

Solution

Rearranging [\[link\]](#), we obtain

Equation:

$$A = \frac{Cd}{\epsilon_0} = \frac{(1.0 \text{ F})(1.0 \times 10^{-3} \text{ m})}{8.85 \times 10^{-12} \text{ F/m}} = 1.1 \times 10^8 \text{ m}^2.$$

Each square plate would have to be 10 km across. It used to be a common prank to ask a student to go to the laboratory stockroom and request a 1-F parallel-plate capacitor, until stockroom attendants got tired of the joke.

Note:

Exercise:

Problem:

Check Your Understanding The capacitance of a parallel-plate capacitor is 2.0 pF. If the area of each plate is 2.4 cm^2 , what is the plate separation?

Solution:

$$1.1 \times 10^{-3} \text{ m}$$

Note:

Exercise:

Problem: Check Your Understanding Verify that σ/V and ϵ_0/d have the same physical units.

Spherical Capacitor

A spherical capacitor is another set of conductors whose capacitance can be easily determined ([\[link\]](#)). It consists of two concentric conducting spherical shells of radii R_1 (inner shell) and R_2 (outer shell). The shells are given equal and opposite charges $+Q$ and $-Q$, respectively. From symmetry, the electrical field between the shells is directed radially outward. We can obtain the magnitude of the field by applying Gauss's law over a spherical Gaussian surface of radius r concentric with the shells. The enclosed charge is $+Q$; therefore we have

Equation:

$$\oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = E(4\pi r^2) = \frac{Q}{\varepsilon_0}.$$

Thus, the electrical field between the conductors is

Equation:

$$\vec{\mathbf{E}} = \frac{1}{4\pi\varepsilon_0} \frac{Q}{r^2} \hat{\mathbf{r}}.$$

We substitute this $\vec{\mathbf{E}}$ into [\[link\]](#) and integrate along a radial path between the shells:

Equation:

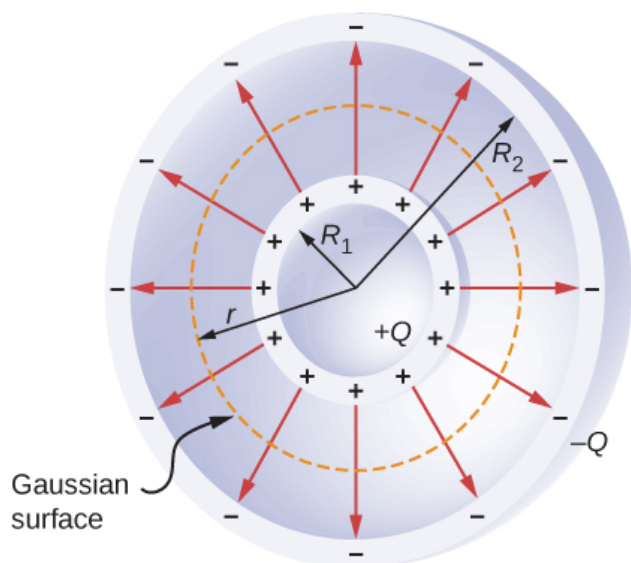
$$V = \int_{R_1}^{R_2} \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} = \int_{R_1}^{R_2} \left(\frac{1}{4\pi\varepsilon_0} \frac{Q}{r^2} \hat{\mathbf{r}} \right) \cdot (\hat{\mathbf{r}} dr) = \frac{Q}{4\pi\varepsilon_0} \int_{R_1}^{R_2} \frac{dr}{r^2} = \frac{Q}{4\pi\varepsilon_0} \left(\frac{1}{R_1} - \frac{1}{R_2} \right).$$

In this equation, the potential difference between the plates is $V = -(V_2 - V_1) = V_1 - V_2$. We substitute this result into [\[link\]](#) to find the capacitance of a spherical capacitor:

Note:

Equation:

$$C = \frac{Q}{V} = 4\pi\varepsilon_0 \frac{R_1 R_2}{R_2 - R_1}.$$



A spherical capacitor consists of two concentric conducting spheres. Note that the charges on a conductor reside on its surface.

Example:

Capacitance of an Isolated Sphere

Calculate the capacitance of a single isolated conducting sphere of radius R_1 and compare it with [\[link\]](#) in the limit as $R_2 \rightarrow \infty$.

Strategy

We assume that the charge on the sphere is Q , and so we follow the four steps outlined earlier. We also assume the other conductor to be a concentric hollow sphere of infinite radius.

Solution

On the outside of an isolated conducting sphere, the electrical field is given by [\[link\]](#). The magnitude of the potential difference between the surface of an isolated sphere and infinity is

Equation:

$$V = \int_{R_1}^{+\infty} \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} = \frac{Q}{4\pi\epsilon_0} \int_{R_1}^{+\infty} \frac{1}{r^2} \hat{\mathbf{r}} \cdot (\hat{\mathbf{r}} dr) = \frac{Q}{4\pi\epsilon_0} \int_{R_1}^{+\infty} \frac{dr}{r^2} = \frac{1}{4\pi\epsilon_0} \frac{Q}{R_1}.$$

The capacitance of an isolated sphere is therefore

Equation:

$$C = \frac{Q}{V} = Q \frac{4\pi\epsilon_0 R_1}{Q} = 4\pi\epsilon_0 R_1.$$

Significance

The same result can be obtained by taking the limit of [\[link\]](#) as $R_2 \rightarrow \infty$. A single isolated sphere is therefore equivalent to a spherical capacitor whose outer shell has an infinitely large radius.

Note:

Exercise:

Problem:

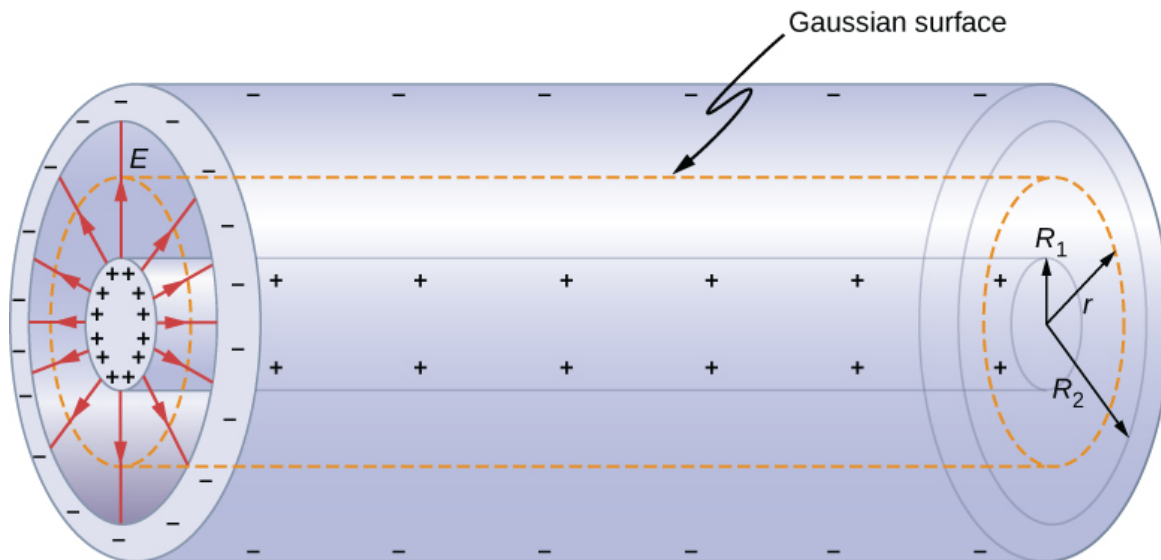
Check Your Understanding The radius of the outer sphere of a spherical capacitor is five times the radius of its inner shell. What are the dimensions of this capacitor if its capacitance is 5.00 pF?

Solution:

3.59 cm, 17.98 cm

Cylindrical Capacitor

A cylindrical capacitor consists of two concentric, conducting cylinders ([\[link\]](#)). The inner cylinder, of radius R_1 , may either be a shell or be completely solid. The outer cylinder is a shell of inner radius R_2 . We assume that the length of each cylinder is l and that the excess charges $+Q$ and $-Q$ reside on the inner and outer cylinders, respectively.



A cylindrical capacitor consists of two concentric, conducting cylinders. Here, the charge on the outer surface of the inner cylinder is positive (indicated by $+$) and the charge on the inner surface of the outer cylinder is negative (indicated by $-$).

With edge effects ignored, the electrical field between the conductors is directed radially outward from the common axis of the cylinders. Using the Gaussian surface shown in [\[link\]](#), we have

Equation:

$$\oint_S \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} dA = E(2\pi r l) = \frac{Q}{\epsilon_0}.$$

Therefore, the electrical field between the cylinders is

Equation:

$$\vec{\mathbf{E}} = \frac{1}{2\pi\epsilon_0} \frac{Q}{r l} \hat{\mathbf{r}}.$$

Here $\hat{\mathbf{r}}$ is the unit radial vector along the radius of the cylinder. We can substitute into [\[link\]](#) and find the potential difference between the cylinders:

Equation:

$$V = \int_{R_1}^{R_2} \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}_p = \frac{Q}{2\pi\epsilon_0 l} \int_{R_1}^{R_2} \frac{1}{r} \hat{\mathbf{r}} \cdot (\hat{\mathbf{r}} dr) = \frac{Q}{2\pi\epsilon_0 l} \int_{R_1}^{R_2} \frac{dr}{r} = \frac{Q}{2\pi\epsilon_0 l} \ln r \Big|_{R_1}^{R_2} = \frac{Q}{2\pi\epsilon_0 l} \ln \frac{R_2}{R_1}.$$

Thus, the capacitance of a cylindrical capacitor is

Note:

Equation:

$$C = \frac{Q}{V} = \frac{2\pi\epsilon_0 l}{\ln(R_2/R_1)}.$$

As in other cases, this capacitance depends only on the geometry of the conductor arrangement. An important application of [\[link\]](#) is the determination of the capacitance per unit length of a *coaxial cable*, which is commonly used to transmit time-varying electrical signals. A coaxial cable consists of two concentric, cylindrical conductors separated by an insulating material. (Here, we assume a vacuum between the conductors, but the physics is qualitatively almost the same when the space between the conductors is filled by a dielectric.) This configuration shields the electrical signal propagating down the inner conductor from stray electrical fields external to the cable. Current flows in opposite directions in the inner and the outer conductors, with the outer conductor usually grounded. Now, from [\[link\]](#), the capacitance per unit length of the coaxial cable is given by

Equation:

$$\frac{C}{l} = \frac{2\pi\epsilon_0}{\ln(R_2/R_1)}.$$

In practical applications, it is important to select specific values of C/l . This can be accomplished with appropriate choices of radii of the conductors and of the insulating material between them.

Note:**Exercise:****Problem:**

Check Your Understanding When a cylindrical capacitor is given a charge of 0.500 nC , a potential difference of 20.0 V is measured between the cylinders. (a) What is the capacitance of this system? (b) If the cylinders are 1.0 m long, what is the ratio of their radii?

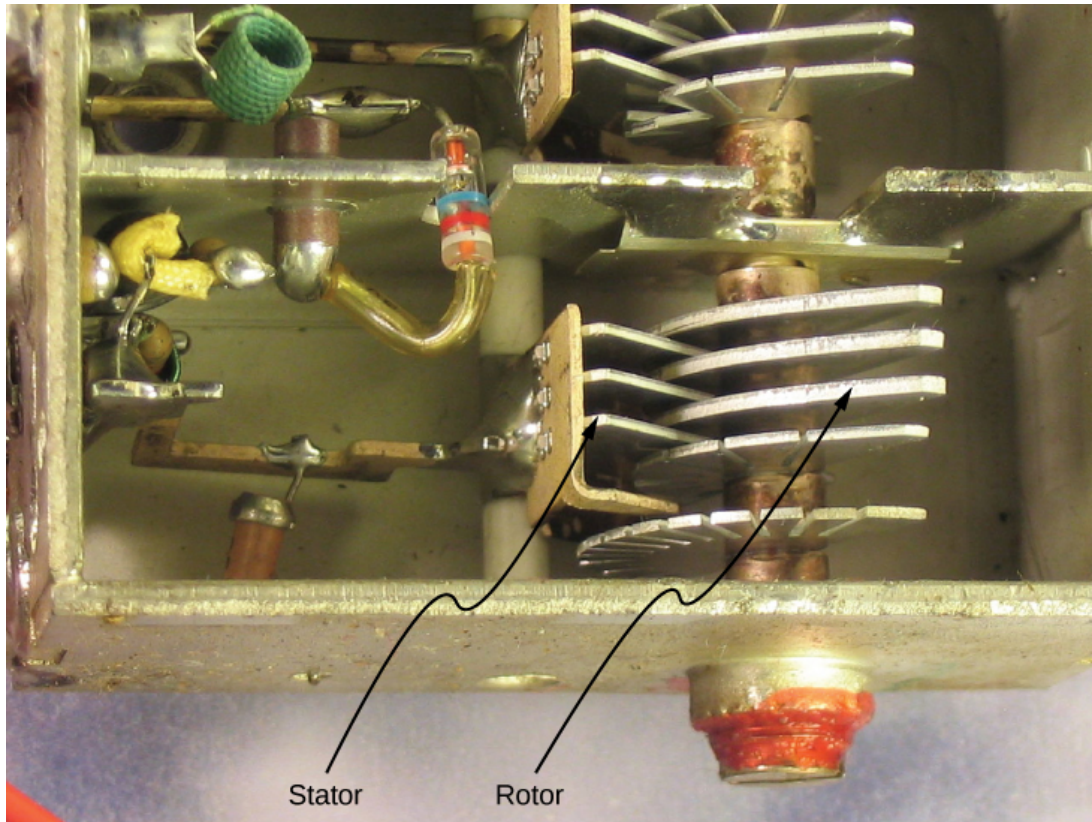
Solution:

a. 25.0 pF ; b. 9.2

Several types of practical capacitors are shown in [\[link\]](#). Common capacitors are often made of two small pieces of metal foil separated by two small pieces of insulation (see [\[link\]\(b\)](#)). The metal foil and insulation are encased in a protective coating, and two metal leads are used for connecting the foils to an external circuit. Some common insulating materials are mica, ceramic, paper, and Teflon™ non-stick coating.

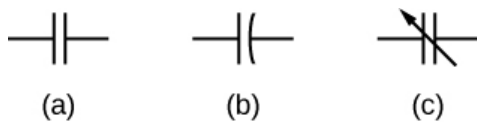
Another popular type of capacitor is an electrolytic capacitor. It consists of an oxidized metal in a conducting paste. The main advantage of an electrolytic capacitor is its high capacitance relative to other common types of capacitors. For example, capacitance of one type of aluminum electrolytic capacitor can be as high as 1.0 F . However, you must be careful when using an electrolytic capacitor in a circuit, because it only functions correctly when the metal foil is at a higher potential than the conducting paste. When reverse polarization occurs, electrolytic action destroys the oxide film. This type of capacitor cannot be connected across an alternating current source, because half of the time, ac voltage would have the wrong polarity, as an alternating current reverses its polarity (see [Alternating-Current Circuits](#) on alternating-current circuits).

A variable air capacitor ([\[link\]](#)) has two sets of parallel plates. One set of plates is fixed (indicated as “stator”), and the other set of plates is attached to a shaft that can be rotated (indicated as “rotor”). By turning the shaft, the cross-sectional area in the overlap of the plates can be changed; therefore, the capacitance of this system can be tuned to a desired value. Capacitor tuning has applications in any type of radio transmission and in receiving radio signals from electronic devices. Any time you tune your car radio to your favorite station, think of capacitance.



In a variable air capacitor, capacitance can be tuned by changing the effective area of the plates. (credit: modification of work by Robbie Sproule)

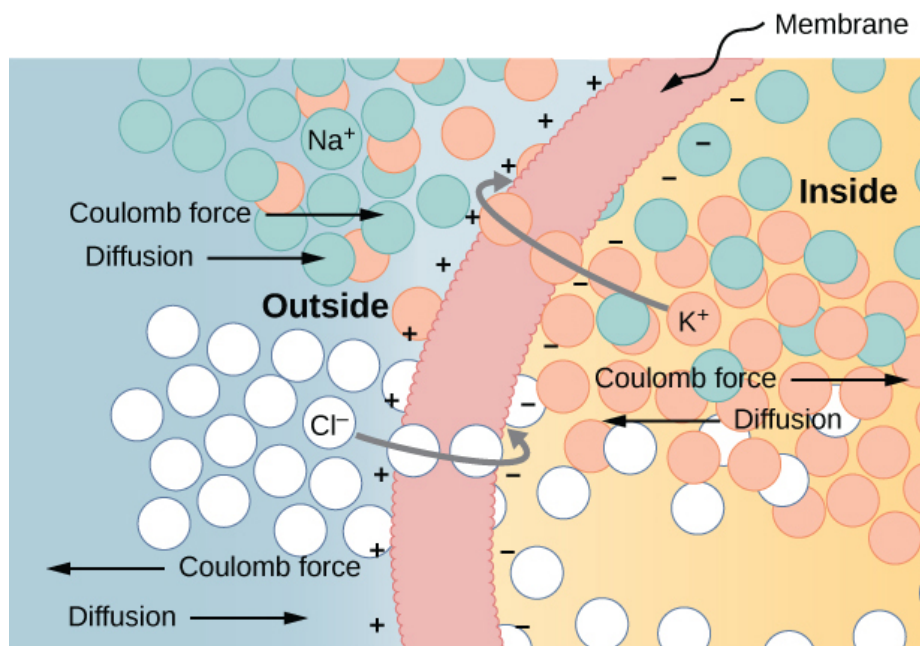
The symbols shown in [\[link\]](#) are circuit representations of various types of capacitors. We generally use the symbol shown in [\[link\]\(a\)](#). The symbol in [\[link\]\(c\)](#) represents a variable-capacitance capacitor. Notice the similarity of these symbols to the symmetry of a parallel-plate capacitor. An electrolytic capacitor is represented by the symbol in part [\[link\]\(b\)](#), where the curved plate indicates the negative terminal.



This shows three different circuit representations of capacitors. The symbol in (a) is the most commonly used one. The symbol in (b) represents an electrolytic capacitor. The symbol in (c) represents a variable-capacitance capacitor.

An interesting applied example of a capacitor model comes from cell biology and deals with the electrical potential in the plasma membrane of a living cell ([link](#)). Cell membranes separate cells from their surroundings but allow some selected ions to pass in or out of the cell. The potential difference across a membrane is about 70 mV. The cell membrane may be 7 to 10 nm thick. Treating the cell membrane as a nano-sized capacitor, the estimate of the smallest electrical field strength across its 'plates' yields the value $E = \frac{V}{d} = \frac{70 \times 10^{-3} \text{V}}{10 \times 10^{-9} \text{m}} = 7 \times 10^6 \text{ V/m} > 3 \text{ MV/m}$.

This magnitude of electrical field is great enough to create an electrical spark in the air.



The semipermeable membrane of a biological cell has different concentrations of ions on its interior surface than on its exterior. Diffusion moves the K^+ (potassium) and Cl^- (chloride) ions in the directions shown, until the Coulomb force halts further transfer. In this way, the exterior of the membrane acquires a positive charge and its interior surface acquires a negative charge, creating a potential difference across the membrane. The membrane is normally impermeable to Na^+ (sodium ions).

Note:

Visit the [PhET Explorations: Capacitor Lab](#) to explore how a capacitor works. Change the size of the plates and add a dielectric to see the effect on capacitance. Change the voltage and see charges built

up on the plates. Observe the electrical field in the capacitor. Measure the voltage and the electrical field.

Summary

- A capacitor is a device that stores an electrical charge and electrical energy. The amount of charge a vacuum capacitor can store depends on two major factors: the voltage applied and the capacitor's physical characteristics, such as its size and geometry.
- The capacitance of a capacitor is a parameter that tells us how much charge can be stored in the capacitor per unit potential difference between its plates. Capacitance of a system of conductors depends only on the geometry of their arrangement and physical properties of the insulating material that fills the space between the conductors. The unit of capacitance is the farad, where $1 \text{ F} = 1 \text{ C}/1 \text{ V}$.

Conceptual Questions

Exercise:

Problem:

Does the capacitance of a device depend on the applied voltage? Does the capacitance of a device depend on the charge residing on it?

Solution:

no; yes

Exercise:

Problem:

Would you place the plates of a parallel-plate capacitor closer together or farther apart to increase their capacitance?

Exercise:

Problem: The value of the capacitance is zero if the plates are not charged. True or false?

Solution:

false

Exercise:

Problem:

If the plates of a capacitor have different areas, will they acquire the same charge when the capacitor is connected across a battery?

Exercise:

Problem:

Does the capacitance of a spherical capacitor depend on which sphere is charged positively or negatively?

Solution:

no

Problems**Exercise:**

Problem: What charge is stored in a $180.0\text{-}\mu\text{F}$ capacitor when 120.0 V is applied to it?

Solution:

21.6 mC

Exercise:

Problem: Find the charge stored when 5.50 V is applied to an 8.00-pF capacitor.

Exercise:

Problem: Calculate the voltage applied to a $2.00\text{-}\mu\text{F}$ capacitor when it holds $3.10\text{ }\mu\text{C}$ of charge.

Solution:

1.55 V

Exercise:

Problem: What voltage must be applied to an 8.00-nF capacitor to store 0.160 mC of charge?

Exercise:

Problem: What capacitance is needed to store $3.00\text{ }\mu\text{C}$ of charge at a voltage of 120 V ?

Solution:

25.0 nF

Exercise:**Problem:**

What is the capacitance of a large Van de Graaff generator's terminal, given that it stores 8.00 mC of charge at a voltage of 12.0 MV ?

Exercise:

Problem:

The plates of an empty parallel-plate capacitor of capacitance 5.0 pF are 2.0 mm apart. What is the area of each plate?

Solution:

$$1.1 \times 10^{-3} \text{m}^2$$

Exercise:**Problem:**

A 60.0-pF vacuum capacitor has a plate area of 0.010 m^2 . What is the separation between its plates?

Exercise:**Problem:**

A set of parallel plates has a capacitance of $5.0 \mu\text{F}$. How much charge must be added to the plates to increase the potential difference between them by 100 V?

Solution:

$$500 \mu\text{C}$$

Exercise:**Problem:**

Consider Earth to be a spherical conductor of radius 6400 km and calculate its capacitance.

Exercise:**Problem:**

If the capacitance per unit length of a cylindrical capacitor is 20 pF/m, what is the ratio of the radii of the two cylinders?

Solution:

$$1:16$$

Exercise:**Problem:**

An empty parallel-plate capacitor has a capacitance of $20 \mu\text{F}$. How much charge must leak off its plates before the voltage across them is reduced by 100 V?

Glossary

capacitance

amount of charge stored per unit volt

capacitor

device that stores electrical charge and electrical energy

dielectric

insulating material used to fill the space between two plates

parallel-plate capacitor

system of two identical parallel conducting plates separated by a distance

Capacitors in Series and in Parallel

By the end of this section, you will be able to:

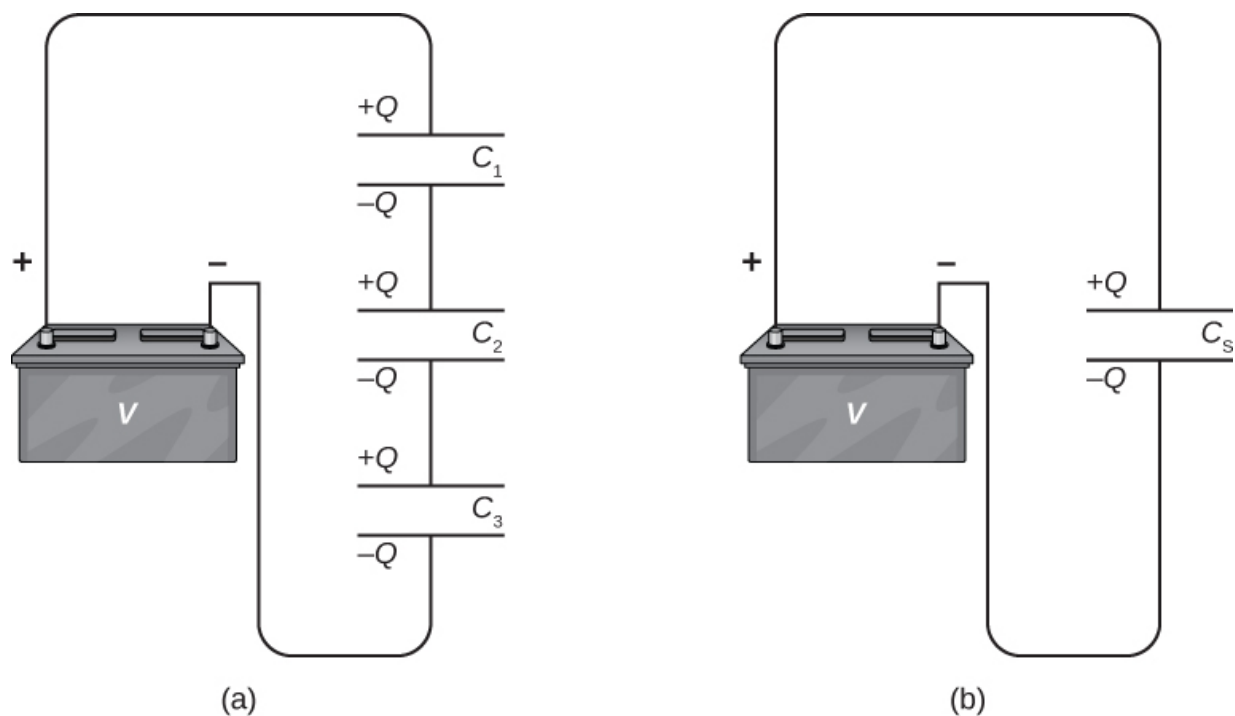
- Explain how to determine the equivalent capacitance of capacitors in series and in parallel combinations
- Compute the potential difference across the plates and the charge on the plates for a capacitor in a network and determine the net capacitance of a network of capacitors

Several capacitors can be connected together to be used in a variety of applications. Multiple connections of capacitors behave as a single equivalent capacitor. The total capacitance of this equivalent single capacitor depends both on the individual capacitors and how they are connected. Capacitors can be arranged in two simple and common types of connections, known as *series* and *parallel*, for which we can easily calculate the total capacitance. These two basic combinations, series and parallel, can also be used as part of more complex connections.

The Series Combination of Capacitors

[\[link\]](#) illustrates a series combination of three capacitors, arranged in a row within the circuit. As for any capacitor, the capacitance of the combination is related to the charge and voltage by using [\[link\]](#). When this series combination is connected to a battery with voltage V , each of the capacitors acquires an identical charge Q . To explain, first note that the charge on the plate connected to the positive terminal of the battery is $+Q$ and the charge on the plate connected to the negative terminal is $-Q$. Charges are then induced on the other plates so that the sum of the charges on all plates, and the sum of charges on any pair of capacitor plates, is zero. However, the potential drop $V_1 = Q/C_1$ on one capacitor may be different from the potential drop $V_2 = Q/C_2$ on another capacitor, because, generally, the capacitors may have different capacitances. The series combination of two or three capacitors resembles a single capacitor with a smaller capacitance. Generally, any number of capacitors connected in series is equivalent to one capacitor whose capacitance (called the *equivalent capacitance*) is smaller than the smallest of the capacitances in the series combination. Charge on this equivalent capacitor is the same as the charge on any capacitor in a series combination: That is, *all*

capacitors of a series combination have the same charge. This occurs due to the conservation of charge in the circuit. When a charge Q in a series circuit is removed from a plate of the first capacitor (which we denote as $-Q$), it must be placed on a plate of the second capacitor (which we denote as $+Q$), and so on.



(a) Three capacitors are connected in series. The magnitude of the charge on each plate is Q . (b) The network of capacitors in (a) is equivalent to one capacitor that has a smaller capacitance than any of the individual capacitances in (a), and the charge on its plates is Q .

We can find an expression for the total (equivalent) capacitance by considering the voltages across the individual capacitors. The potentials across capacitors 1, 2, and 3 are, respectively, $V_1 = Q/C_1$, $V_2 = Q/C_2$, and $V_3 = Q/C_3$. These potentials must sum up to the voltage of the battery, giving the following potential balance:

Equation:

$$V = V_1 + V_2 + V_3.$$

Potential V is measured across an equivalent capacitor that holds charge Q and has an equivalent capacitance C_S . Entering the expressions for V_1 , V_2 , and V_3 , we get

Equation:

$$\frac{Q}{C_S} = \frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3}.$$

Canceling the charge Q , we obtain an expression containing the equivalent capacitance, C_S , of three capacitors connected in series:

Equation:

$$\frac{1}{C_S} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}.$$

This expression can be generalized to any number of capacitors in a series network.

Note:

Series Combination

For capacitors connected in a **series combination**, the reciprocal of the equivalent capacitance is the sum of reciprocals of individual capacitances:

Equation:

$$\frac{1}{C_S} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \dots.$$

Example:

Equivalent Capacitance of a Series Network

Find the total capacitance for three capacitors connected in series, given their individual capacitances are $1.000\ \mu\text{F}$, $5.000\ \mu\text{F}$, and $8.000\ \mu\text{F}$.

Strategy

Because there are only three capacitors in this network, we can find the equivalent capacitance by using [\[link\]](#) with three terms.

Solution

We enter the given capacitances into [\[link\]](#):

Equation:

$$\begin{aligned}\frac{1}{C_S} &= \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \\ &= \frac{1}{1.000\ \mu\text{F}} + \frac{1}{5.000\ \mu\text{F}} + \frac{1}{8.000\ \mu\text{F}} \\ \frac{1}{C_S} &= \frac{1.325}{\mu\text{F}}.\end{aligned}$$

Now we invert this result and obtain $C_S = \frac{\mu\text{F}}{1.325} = 0.755\ \mu\text{F}$.

Significance

Note that in a series network of capacitors, the equivalent capacitance is always less than the smallest individual capacitance in the network.

The Parallel Combination of Capacitors

A parallel combination of three capacitors, with one plate of each capacitor connected to one side of the circuit and the other plate connected to the other side, is illustrated in [\[link\]](#)(a). Since the capacitors are connected in parallel, *they all have the same voltage V across their plates*. However, each capacitor in the parallel network may store a different charge. To find the equivalent capacitance C_P of the parallel network, we note that the total charge Q stored by the network is the sum of all the individual charges:

Equation:

$$Q = Q_1 + Q_2 + Q_3.$$

On the left-hand side of this equation, we use the relation $Q = C_P V$, which holds for the entire network. On the right-hand side of the equation, we use the relations $Q_1 = C_1 V$, $Q_2 = C_2 V$, and $Q_3 = C_3 V$ for the three capacitors in the network. In this way we obtain

Equation:

$$C_P V = C_1 V + C_2 V + C_3 V.$$

This equation, when simplified, is the expression for the equivalent capacitance of the parallel network of three capacitors:

Equation:

$$C_P = C_1 + C_2 + C_3.$$

This expression is easily generalized to any number of capacitors connected in parallel in the network.

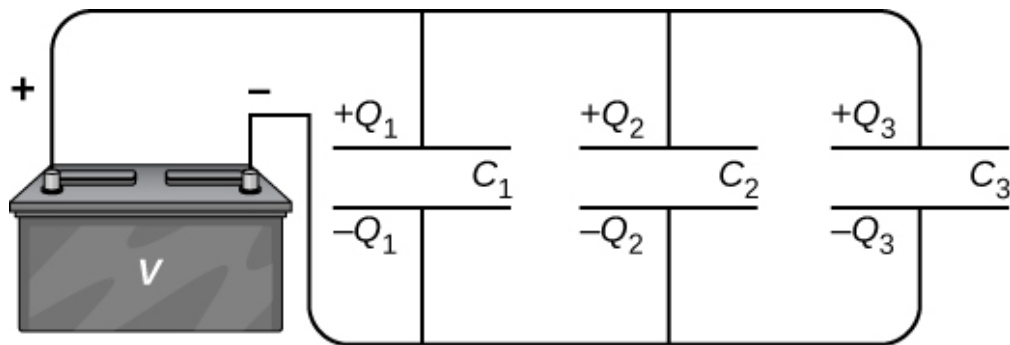
Note:

Parallel Combination

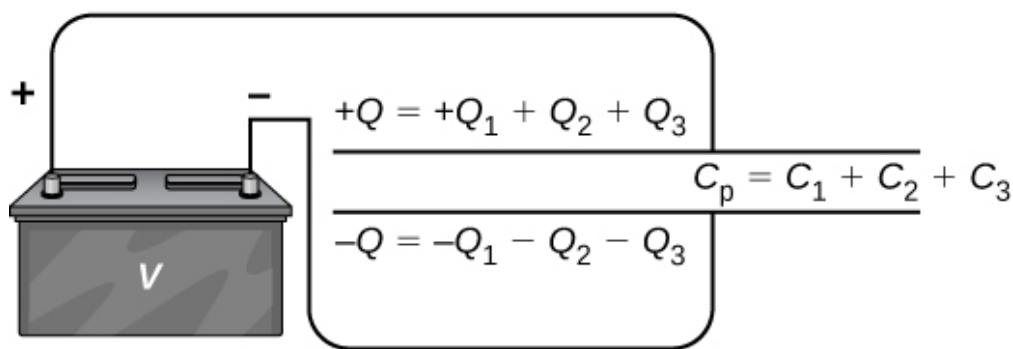
For capacitors connected in a **parallel combination**, the equivalent (net) capacitance is the sum of all individual capacitances in the network,

Equation:

$$C_P = C_1 + C_2 + C_3 + \cdots.$$



(a)



(b)

(a) Three capacitors are connected in parallel. Each capacitor is connected directly to the battery. (b) The charge on the equivalent capacitor is the sum of the charges on the individual capacitors.

Example:

Equivalent Capacitance of a Parallel Network

Find the net capacitance for three capacitors connected in parallel, given their individual capacitances are $1.0 \mu\text{F}$, $5.0 \mu\text{F}$, and $8.0 \mu\text{F}$.

Strategy

Because there are only three capacitors in this network, we can find the equivalent capacitance by using [\[link\]](#) with three terms.

Solution

Entering the given capacitances into [\[link\]](#) yields

Equation:

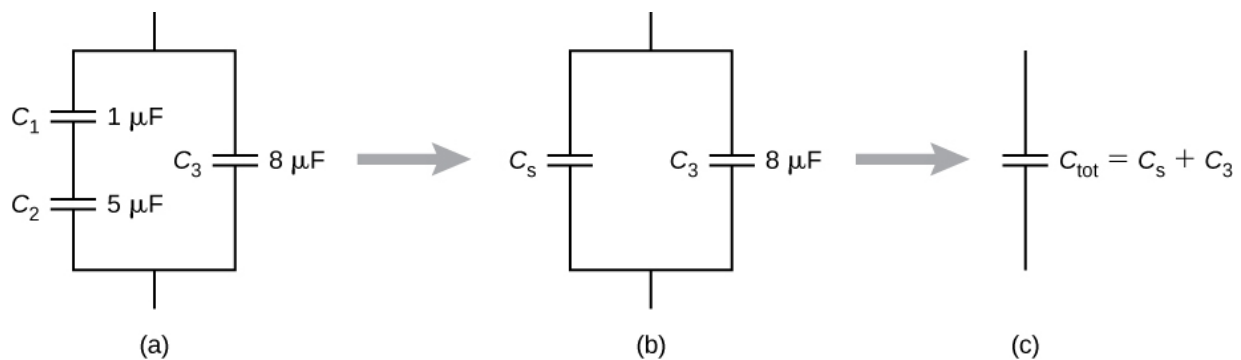
$$C_P = C_1 + C_2 + C_3 = 1.0 \mu\text{F} + 5.0 \mu\text{F} + 8.0 \mu\text{F}$$

$$C_P = 14.0 \mu\text{F}.$$

Significance

Note that in a parallel network of capacitors, the equivalent capacitance is always larger than any of the individual capacitances in the network.

Capacitor networks are usually some combination of series and parallel connections, as shown in [\[link\]](#). To find the net capacitance of such combinations, we identify parts that contain only series or only parallel connections, and find their equivalent capacitances. We repeat this process until we can determine the equivalent capacitance of the entire network. The following example illustrates this process.



(a) This circuit contains both series and parallel connections of capacitors. (b) C_1 and C_2 are in series; their equivalent capacitance is C_s . (c) The equivalent capacitance C_s is connected in parallel with C_3 . Thus, the equivalent capacitance of the entire network is the sum of C_s and C_3 .

Example:**Equivalent Capacitance of a Network**

Find the total capacitance of the combination of capacitors shown in [\[link\]](#).

Assume the capacitances are known to three decimal places

($C_1 = 1.000 \mu\text{F}$, $C_2 = 5.000 \mu\text{F}$, $C_3 = 8.000 \mu\text{F}$). Round your answer to three decimal places.

Strategy

We first identify which capacitors are in series and which are in parallel.

Capacitors C_1 and C_2 are in series. Their combination, labeled C_S , is in parallel with C_3 .

Solution

Since C_1 and C_2 are in series, their equivalent capacitance C_S is obtained with [\[link\]](#):

Equation:

$$\frac{1}{C_S} = \frac{1}{C_1} + \frac{1}{C_2} = \frac{1}{1.000 \mu\text{F}} + \frac{1}{5.000 \mu\text{F}} = \frac{1.200}{\mu\text{F}} \Rightarrow C_S = 0.833 \mu\text{F}.$$

Capacitance C_S is connected in parallel with the third capacitance C_3 , so we use [\[link\]](#) to find the equivalent capacitance C of the entire network:

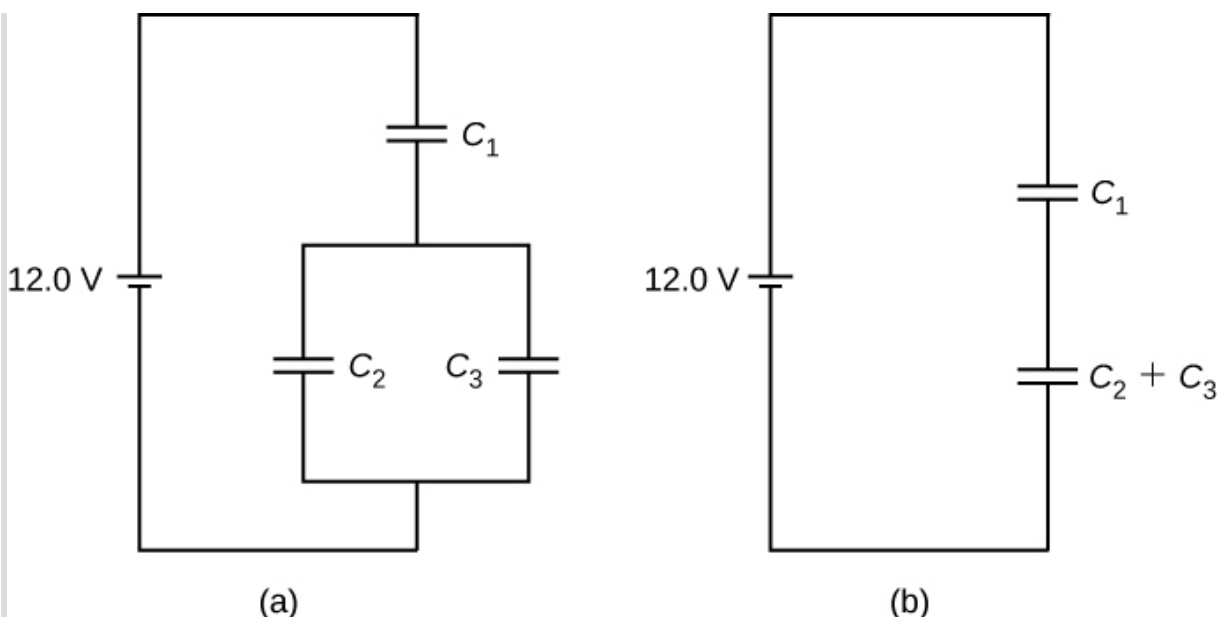
Equation:

$$C = C_S + C_3 = 0.833 \mu\text{F} + 8.000 \mu\text{F} = 8.833 \mu\text{F}.$$

Example:**Network of Capacitors**

Determine the net capacitance C of the capacitor combination shown in [\[link\]](#) when the capacitances are $C_1 = 12.0 \mu\text{F}$, $C_2 = 2.0 \mu\text{F}$, and $C_3 = 4.0 \mu\text{F}$.

When a 12.0-V potential difference is maintained across the combination, find the charge and the voltage across each capacitor.



(a) A capacitor combination. (b) An equivalent two-capacitor combination.

Strategy

We first compute the net capacitance C_{23} of the parallel connection C_2 and C_3 . Then C is the net capacitance of the series connection C_1 and C_{23} . We use the relation $C = Q/V$ to find the charges Q_1, Q_2 , and Q_3 , and the voltages V_1 , V_2 , and V_3 , across capacitors 1, 2, and 3, respectively.

Solution

The equivalent capacitance for C_2 and C_3 is

Equation:

$$C_{23} = C_2 + C_3 = 2.0 \mu\text{F} + 4.0 \mu\text{F} = 6.0 \mu\text{F}.$$

The entire three-capacitor combination is equivalent to two capacitors in series,

Equation:

$$\frac{1}{C} = \frac{1}{12.0 \mu\text{F}} + \frac{1}{6.0 \mu\text{F}} = \frac{1}{4.0 \mu\text{F}} \Rightarrow C = 4.0 \mu\text{F}.$$

Consider the equivalent two-capacitor combination in [\[link\]](#)(b). Since the capacitors are in series, they have the same charge, $Q_1 = Q_{23}$. Also, the

capacitors share the 12.0-V potential difference, so

Equation:

$$12.0 \text{ V} = V_1 + V_{23} = \frac{Q_1}{C_1} + \frac{Q_{23}}{C_{23}} = \frac{Q_1}{12.0 \mu\text{F}} + \frac{Q_1}{6.0 \mu\text{F}} \Rightarrow Q_1 = 48.0 \mu\text{C}.$$

Now the potential difference across capacitor 1 is

Equation:

$$V_1 = \frac{Q_1}{C_1} = \frac{48.0 \mu\text{C}}{12.0 \mu\text{F}} = 4.0 \text{ V}.$$

Because capacitors 2 and 3 are connected in parallel, they are at the same potential difference:

Equation:

$$V_2 = V_3 = 12.0 \text{ V} - 4.0 \text{ V} = 8.0 \text{ V}.$$

Hence, the charges on these two capacitors are, respectively,

Equation:

$$\begin{aligned} Q_2 &= C_2 V_2 = (2.0 \mu\text{F})(8.0 \text{ V}) = 16.0 \mu\text{C}, \\ Q_3 &= C_3 V_3 = (4.0 \mu\text{F})(8.0 \text{ V}) = 32.0 \mu\text{C}. \end{aligned}$$

Significance

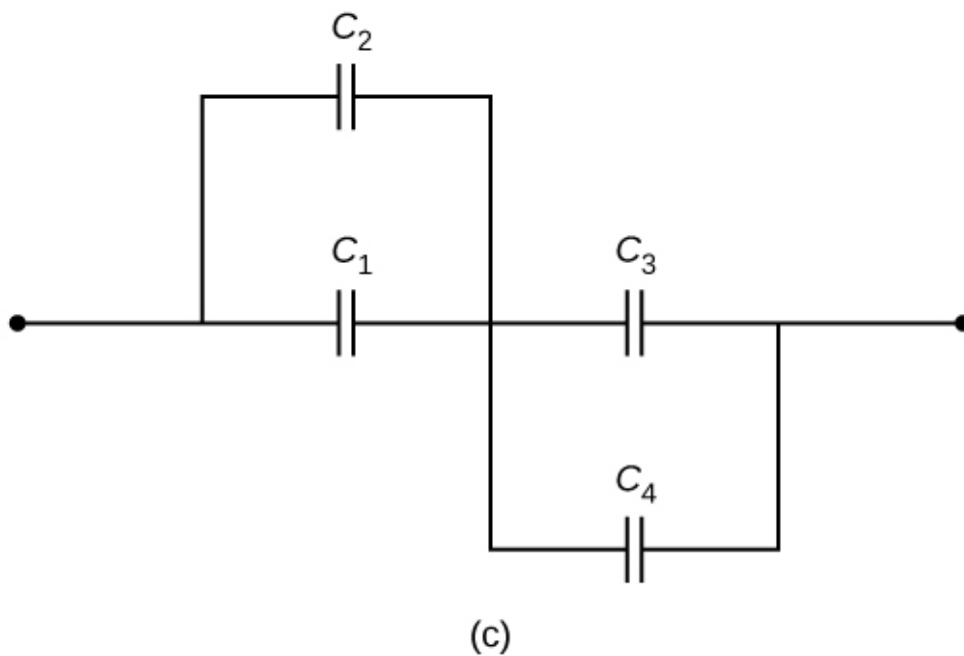
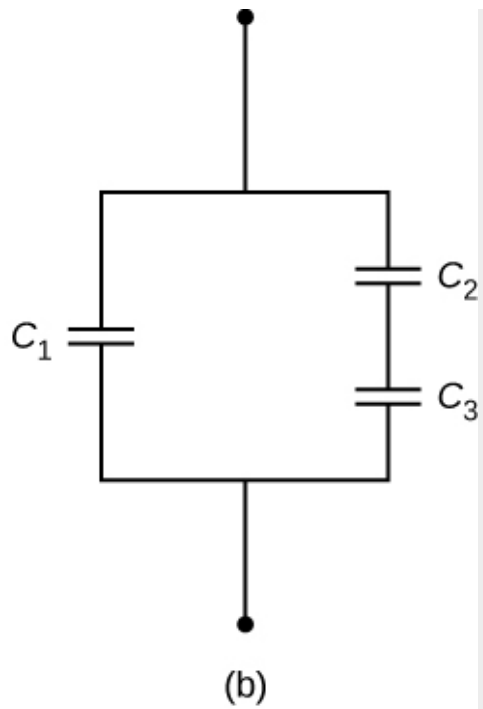
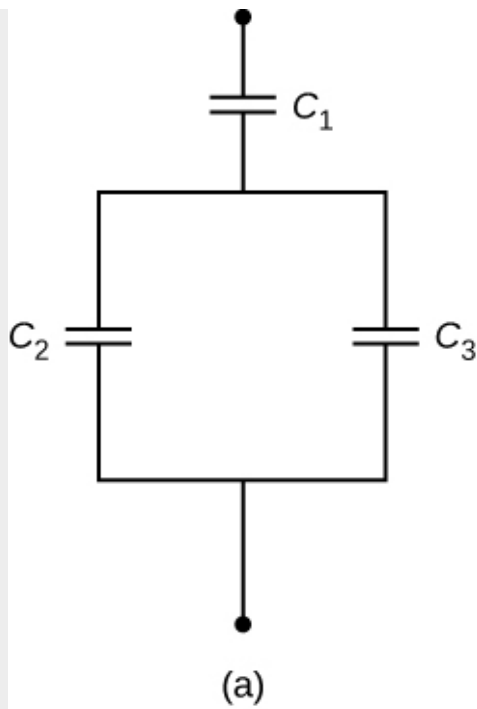
As expected, the net charge on the parallel combination of C_2 and C_3 is $Q_{23} = Q_2 + Q_3 = 48.0 \mu\text{C}$.

Note:

Exercise:

Problem:

Check Your Understanding Determine the net capacitance C of each network of capacitors shown below. Assume that $C_1 = 1.0 \text{ pF}$, $C_2 = 2.0 \text{ pF}$, $C_3 = 4.0 \text{ pF}$, and $C_4 = 5.0 \text{ pF}$. Find the charge on each capacitor, assuming there is a potential difference of 12.0 V across each network.



Solution:

a. $C = 0.86 \text{ pF}$, $Q_1 = 10 \text{ pC}$, $Q_2 = 3.4 \text{ pC}$, $Q_3 = 6.8 \text{ pC}$;

b. $C = 2.3 \text{ pF}$, $Q_1 = 12 \text{ pC}$, $Q_2 = Q_3 = 16 \text{ pC}$;

$$\text{c. } C = 2.3 \text{ pF}, Q_1 = 9.0 \text{ pC}, Q_2 = 18 \text{ pC}, Q_3 = 12 \text{ pC}, Q_4 = 15 \text{ pC}$$

Summary

- When several capacitors are connected in a series combination, the reciprocal of the equivalent capacitance is the sum of the reciprocals of the individual capacitances.
- When several capacitors are connected in a parallel combination, the equivalent capacitance is the sum of the individual capacitances.
- When a network of capacitors contains a combination of series and parallel connections, we identify the series and parallel networks, and compute their equivalent capacitances step by step until the entire network becomes reduced to one equivalent capacitance.

Conceptual Questions

Exercise:

Problem:

If you wish to store a large amount of charge in a capacitor bank, would you connect capacitors in series or in parallel? Explain.

Exercise:

Problem:

What is the maximum capacitance you can get by connecting three $1.0\text{-}\mu\text{F}$ capacitors? What is the minimum capacitance?

Solution:

$3.0 \mu\text{F}, 0.33 \mu\text{F}$

Problems

Exercise:

Problem:

A 4.00-pF is connected in series with an 8.00-pF capacitor and a 400-V potential difference is applied across the pair. (a) What is the charge on each capacitor? (b) What is the voltage across each capacitor?

Solution:

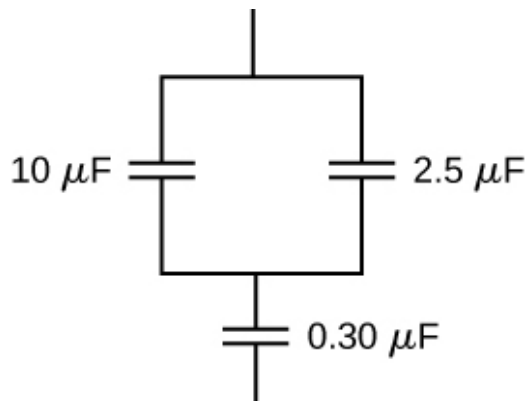
a. 1.07 nC; b. 267 V, 133 V

Exercise:**Problem:**

Three capacitors, with capacitances of $C_1 = 2.0 \mu\text{F}$, $C_2 = 3.0 \mu\text{F}$, and $C_3 = 6.0 \mu\text{F}$, respectively, are connected in parallel. A 500-V potential difference is applied across the combination. Determine the voltage across each capacitor and the charge on each capacitor.

Exercise:**Problem:**

Find the total capacitance of this combination of series and parallel capacitors shown below.

**Solution:**

$0.29 \mu\text{F}$

Exercise:

Problem:

Suppose you need a capacitor bank with a total capacitance of 0.750 F but you have only 1.50-mF capacitors at your disposal. What is the smallest number of capacitors you could connect together to achieve your goal, and how would you connect them?

Solution:

500 capacitors; connected in parallel

Exercise:**Problem:**

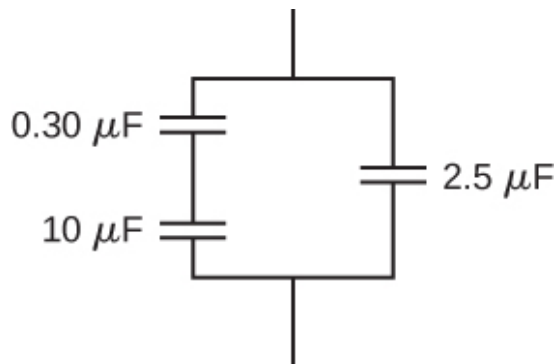
What total capacitances can you make by connecting a 5.00- μF and a 8.00- μF capacitor?

Solution:

3.08 μF (series) and 13.0 μF (parallel)

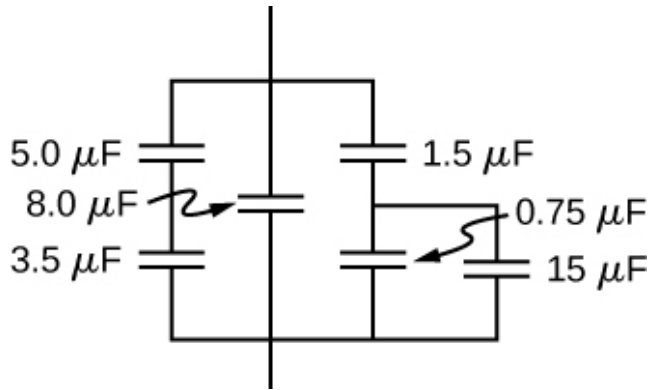
Exercise:**Problem:**

Find the equivalent capacitance of the combination of series and parallel capacitors shown below.

**Exercise:**

Problem:

Find the net capacitance of the combination of series and parallel capacitors shown below.



Solution:

$11.4 \mu\text{F}$

Exercise:**Problem:**

A 40-pF capacitor is charged to a potential difference of 500 V. Its terminals are then connected to those of an uncharged 10-pF capacitor. Calculate: (a) the original charge on the 40-pF capacitor; (b) the charge on each capacitor after the connection is made; and (c) the potential difference across the plates of each capacitor after the connection.

Exercise:**Problem:**

A $2.0\text{-}\mu\text{F}$ capacitor and a $4.0\text{-}\mu\text{F}$ capacitor are connected in series across a 1.0-kV potential. The charged capacitors are then disconnected from the source and connected to each other with terminals of like sign together. Find the charge on each capacitor and the voltage across each capacitor.

Solution:

0.89 mC; 1.78 mC; 444 V

Glossary

parallel combination

components in a circuit arranged with one side of each component connected to one side of the circuit and the other sides of the components connected to the other side of the circuit

series combination

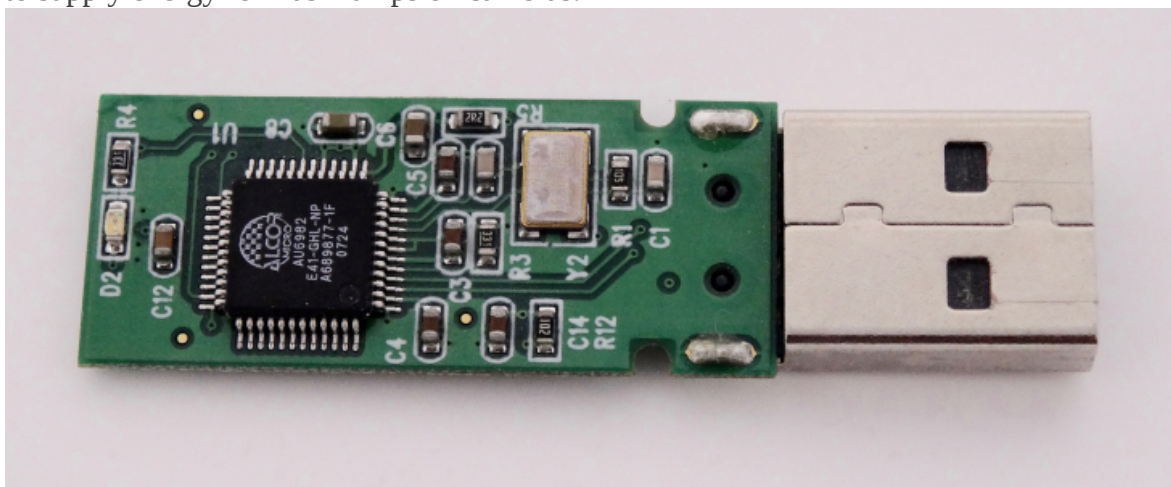
components in a circuit arranged in a row one after the other in a circuit

Energy Stored in a Capacitor

By the end of this section, you will be able to:

- Explain how energy is stored in a capacitor
- Use energy relations to determine the energy stored in a capacitor network

Most of us have seen dramatizations of medical personnel using a defibrillator to pass an electrical current through a patient's heart to get it to beat normally. Often realistic in detail, the person applying the shock directs another person to “make it 400 joules this time.” The energy delivered by the defibrillator is stored in a capacitor and can be adjusted to fit the situation. SI units of joules are often employed. Less dramatic is the use of capacitors in microelectronics to supply energy when batteries are charged ([link](#)). Capacitors are also used to supply energy for flash lamps on cameras.



The capacitors on the circuit board for an electronic device follow a labeling convention that identifies each one with a code that begins with the letter “C.” (credit: Windell Oskay)

The energy U_C stored in a capacitor is electrostatic potential energy and is thus related to the charge Q and voltage V between the capacitor plates. A charged capacitor stores energy in the electrical field between its plates. As the capacitor is being charged, the electrical field builds up. When a charged capacitor is disconnected from a battery, its energy remains in the field in the space between its plates.

To gain insight into how this energy may be expressed (in terms of Q and V), consider a charged, empty, parallel-plate capacitor; that is, a capacitor without a dielectric but with a vacuum between its plates. The space between its plates has a volume Ad , and it is filled with a uniform electrostatic field E . The total energy U_C of the capacitor is contained within this space. The **energy density** u_E in this space is simply U_C divided by the volume Ad . If we know the energy density, the energy can be found as $U_C = u_E(Ad)$. We will learn in

[Electromagnetic Waves](#) (after completing the study of Maxwell's equations) that the energy density u_E in a region of free space occupied by an electrical field E depends only on the magnitude of the field and is

Note:

Equation:

$$u_E = \frac{1}{2} \varepsilon_0 E^2.$$

If we multiply the energy density by the volume between the plates, we obtain the amount of energy stored between the plates of a parallel-plate capacitor:

$$U_C = u_E(Ad) = \frac{1}{2} \varepsilon_0 E^2 Ad = \frac{1}{2} \varepsilon_0 \frac{V^2}{d^2} Ad = \frac{1}{2} V^2 \varepsilon_0 \frac{A}{d} = \frac{1}{2} V^2 C.$$

In this derivation, we used the fact that the electrical field between the plates is uniform so that $E = V/d$ and $C = \varepsilon_0 A/d$. Because $C = Q/V$, we can express this result in other equivalent forms:

Note:

Equation:

$$U_C = \frac{1}{2} V^2 C = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} QV.$$

The expression in [\[link\]](#) for the energy stored in a parallel-plate capacitor is generally valid for all types of capacitors. To see this, consider any uncharged capacitor (not necessarily a parallel-plate type). At some instant, we connect it across a battery, giving it a potential difference $V = q/C$ between its plates. Initially, the charge on the plates is $Q = 0$. As the capacitor is being charged, the charge gradually builds up on its plates, and after some time, it reaches the value Q . To move an infinitesimal charge dq from the negative plate to the positive plate (from a lower to a higher potential), the amount of work dW that must be done on dq is $dW = Vdq = \frac{q}{C} dq$.

This work becomes the energy stored in the electrical field of the capacitor. In order to charge the capacitor to a charge Q , the total work required is

Equation:

$$W = \int_0^{W(Q)} dW = \int_0^Q \frac{q}{C} dq = \frac{1}{2} \frac{Q^2}{C}.$$

Since the geometry of the capacitor has not been specified, this equation holds for any type of capacitor. The total work W needed to charge a capacitor is the electrical potential energy U_C stored in it, or $U_C = W$. When the charge is expressed in coulombs, potential is expressed in volts, and the capacitance is expressed in farads, this relation gives the energy in joules.

Knowing that the energy stored in a capacitor is $U_C = Q^2/(2C)$, we can now find the energy density u_E stored in a vacuum between the plates of a charged parallel-plate capacitor. We just have to divide U_C by the volume Ad of space between its plates and take into account that for a parallel-plate capacitor, we have $E = \sigma/\epsilon_0$ and $C = \epsilon_0 A/d$. Therefore, we obtain

Equation:

$$u_E = \frac{U_C}{Ad} = \frac{1}{2} \frac{Q^2}{C} \frac{1}{Ad} = \frac{1}{2} \frac{Q^2}{\epsilon_0 A/d} \frac{1}{Ad} = \frac{1}{2} \frac{1}{\epsilon_0} \left(\frac{Q}{A} \right)^2 = \frac{\sigma^2}{2\epsilon_0} = \frac{(E\epsilon_0)^2}{2\epsilon_0} = \frac{\epsilon_0}{2} E^2.$$

We see that this expression for the density of energy stored in a parallel-plate capacitor is in accordance with the general relation expressed in [\[link\]](#). We could repeat this calculation for either a spherical capacitor or a cylindrical capacitor—or other capacitors—and in all cases, we would end up with the general relation given by [\[link\]](#).

Example:

Energy Stored in a Capacitor

Calculate the energy stored in the capacitor network in [\[link\]](#)(a) when the capacitors are fully charged and when the capacitances are $C_1 = 12.0 \mu\text{F}$, $C_2 = 2.0 \mu\text{F}$, and $C_3 = 4.0 \mu\text{F}$, respectively.

Strategy

We use [\[link\]](#) to find the energy U_1 , U_2 , and U_3 stored in capacitors 1, 2, and 3, respectively. The total energy is the sum of all these energies.

Solution

We identify $C_1 = 12.0 \mu\text{F}$ and $V_1 = 4.0 \text{ V}$, $C_2 = 2.0 \mu\text{F}$ and $V_2 = 8.0 \text{ V}$, $C_3 = 4.0 \mu\text{F}$ and $V_3 = 8.0 \text{ V}$. The energies stored in these capacitors are

Equation:

$$\begin{aligned} U_1 &= \frac{1}{2} C_1 V_1^2 = \frac{1}{2} (12.0 \mu\text{F}) (4.0 \text{ V})^2 = 96 \mu\text{J}, \\ U_2 &= \frac{1}{2} C_2 V_2^2 = \frac{1}{2} (2.0 \mu\text{F}) (8.0 \text{ V})^2 = 64 \mu\text{J}, \\ U_3 &= \frac{1}{2} C_3 V_3^2 = \frac{1}{2} (4.0 \mu\text{F}) (8.0 \text{ V})^2 = 130 \mu\text{J}. \end{aligned}$$

The total energy stored in this network is

Equation:

$$U_C = U_1 + U_2 + U_3 = 96 \mu\text{J} + 64 \mu\text{J} + 130 \mu\text{J} = 0.29 \text{ mJ}.$$

Significance

We can verify this result by calculating the energy stored in the single 4.0- μF capacitor, which is found to be equivalent to the entire network. The voltage across the network is 12.0 V. The total energy obtained in this way agrees with our previously obtained result, $U_C = \frac{1}{2} CV^2 = \frac{1}{2} (4.0 \mu\text{F})(12.0 \text{ V})^2 = 0.29 \text{ mJ}$.

Note:

Exercise:

Problem:

Check Your Understanding The potential difference across a 5.0-pF capacitor is 0.40 V. (a) What is the energy stored in this capacitor? (b) The potential difference is now increased to 1.20 V. By what factor is the stored energy increased?

Solution:

a. $4.0 \times 10^{-13} \text{ J}$; b. 9 times

In a cardiac emergency, a portable electronic device known as an automated external defibrillator (AED) can be a lifesaver. A **defibrillator** ([\[link\]](#)) delivers a large charge in a short burst, or a shock, to a person's heart to correct abnormal heart rhythm (an arrhythmia). A heart attack can arise from the onset of fast, irregular beating of the heart—called cardiac or ventricular fibrillation. Applying a large shock of electrical energy can terminate the arrhythmia and allow the body's natural pacemaker to resume its normal rhythm. Today, it is common for ambulances to carry AEDs. AEDs are also found in many public places. These are designed to be used by lay persons. The device automatically diagnoses the patient's heart rhythm and then applies the shock with appropriate energy and waveform. CPR (cardiopulmonary resuscitation) is recommended in many cases before using a defibrillator.



Automated external defibrillators are found in many public places. These portable units provide verbal instructions for use in the important first few minutes for a person suffering a cardiac attack. (credit: Owain Davies)

Example:**Capacitance of a Heart Defibrillator**

A heart defibrillator delivers $4.00 \times 10^2 \text{ J}$ of energy by discharging a capacitor initially at $1.00 \times 10^4 \text{ V}$. What is its capacitance?

Strategy

We are given U_C and V , and we are asked to find the capacitance C . We solve [\[link\]](#) for C and substitute.

Solution

Solving this expression for C and entering the given values yields

$$C = 2 \frac{U_C}{V^2} = 2 \frac{4.00 \times 10^2 \text{ J}}{(1.00 \times 10^4 \text{ V})^2} = 8.00 \mu\text{F}.$$

Summary

- Capacitors are used to supply energy to a variety of devices, including defibrillators, microelectronics such as calculators, and flash lamps.
- The energy stored in a capacitor is the work required to charge the capacitor, beginning with no charge on its plates. The energy is stored in the electrical field in the space between the capacitor plates. It depends on the amount of electrical charge on the plates and on the potential difference between the plates.
- The energy stored in a capacitor network is the sum of the energies stored on individual capacitors in the network. It can be computed as the energy stored in the equivalent capacitor of the network.

Conceptual Questions

Exercise:

Problem:

If you wish to store a large amount of energy in a capacitor bank, would you connect capacitors in series or parallel? Explain.

Problems

Exercise:

Problem:

How much energy is stored in an $8.00\text{-}\mu\text{F}$ capacitor whose plates are at a potential difference of 6.00 V ?

Exercise:

Problem:

A capacitor has a charge of $2.5\text{ }\mu\text{C}$ when connected to a 6.0-V battery. How much energy is stored in this capacitor?

Solution:

$7.5\text{ }\mu\text{J}$

Exercise:

Problem:

How much energy is stored in the electrical field of a metal sphere of radius 2.0 m that is kept at a 10.0-V potential?

Exercise:

Problem:

(a) What is the energy stored in the $10.0\text{-}\mu\text{F}$ capacitor of a heart defibrillator charged to $9.00 \times 10^3 \text{ V}$? (b) Find the amount of the stored charge.

Solution:

a. 405 J; b. 90.0 mC

Exercise:**Problem:**

In open-heart surgery, a much smaller amount of energy will defibrillate the heart. (a) What voltage is applied to the $8.00\text{-}\mu\text{F}$ capacitor of a heart defibrillator that stores 40.0 J of energy? (b) Find the amount of the stored charge.

Exercise:**Problem:**

A $165\text{-}\mu\text{F}$ capacitor is used in conjunction with a dc motor. How much energy is stored in it when 119 V is applied?

Solution:

1.17 J

Exercise:**Problem:**

Suppose you have a 9.00-V battery, a $2.00\text{-}\mu\text{F}$ capacitor, and a $7.40\text{-}\mu\text{F}$ capacitor. (a) Find the charge and energy stored if the capacitors are connected to the battery in series. (b) Do the same for a parallel connection.

Exercise:**Problem:**

An anxious physicist worries that the two metal shelves of a wood frame bookcase might obtain a high voltage if charged by static electricity, perhaps produced by friction. (a) What is the capacitance of the empty shelves if they have area $1.00 \times 10^2 \text{ m}^2$ and are 0.200 m apart? (b) What is the voltage between them if opposite charges of magnitude 2.00 nC are placed on them? (c) To show that this voltage poses a small hazard, calculate the energy stored. (d) The actual shelves have an area 100 times smaller than these hypothetical shelves with a connection to the same voltage. Are his fears justified?

Solution:

a. $4.43 \times 10^{-9} \text{ F}$; b. 0.453 V; c. $4.53 \times 10^{-10} \text{ J}$; d. no

Exercise:**Problem:**

A parallel-plate capacitor is made of two square plates 25 cm on a side and 1.0 mm apart. The capacitor is connected to a 50.0-V battery. With the battery still connected, the plates are pulled apart to a separation of 2.00 mm. What are the energies stored in the capacitor before and after the plates are pulled farther apart? Why does the energy decrease even though work is done in separating the plates?

Exercise:**Problem:**

Suppose that the capacitance of a variable capacitor can be manually changed from 100 pF to 800 pF by turning a dial, connected to one set of plates by a shaft, from 0° to 180° . With the dial set at 180° (corresponding to $C = 800$ pF), the capacitor is connected to a 500-V source. After charging, the capacitor is disconnected from the source, and the dial is turned to 0° . If friction is negligible, how much work is required to turn the dial from 180° to 0° ?

Solution:

0.7 mJ

Glossary

energy density

energy stored in a capacitor divided by the volume between the plates

Capacitor with a Dielectric

By the end of this section, you will be able to:

- Describe the effects a dielectric in a capacitor has on capacitance and other properties
- Calculate the capacitance of a capacitor containing a dielectric

As we discussed earlier, an insulating material placed between the plates of a capacitor is called a dielectric. Inserting a dielectric between the plates of a capacitor affects its capacitance. To see why, let's consider an experiment described in [\[link\]](#). Initially, a capacitor with capacitance C_0 when there is air between its plates is charged by a battery to voltage V_0 . When the capacitor is fully charged, the battery is disconnected. A charge Q_0 then resides on the plates, and the potential difference between the plates is measured to be V_0 . Now, suppose we insert a dielectric that *totally* fills the gap between the plates. If we monitor the voltage, we find that the voltmeter reading has dropped to a *smaller* value V . We write this new voltage value as a fraction of the original voltage V_0 , with a positive number κ , $\kappa > 1$:

Equation:

$$V = \frac{1}{\kappa} V_0.$$

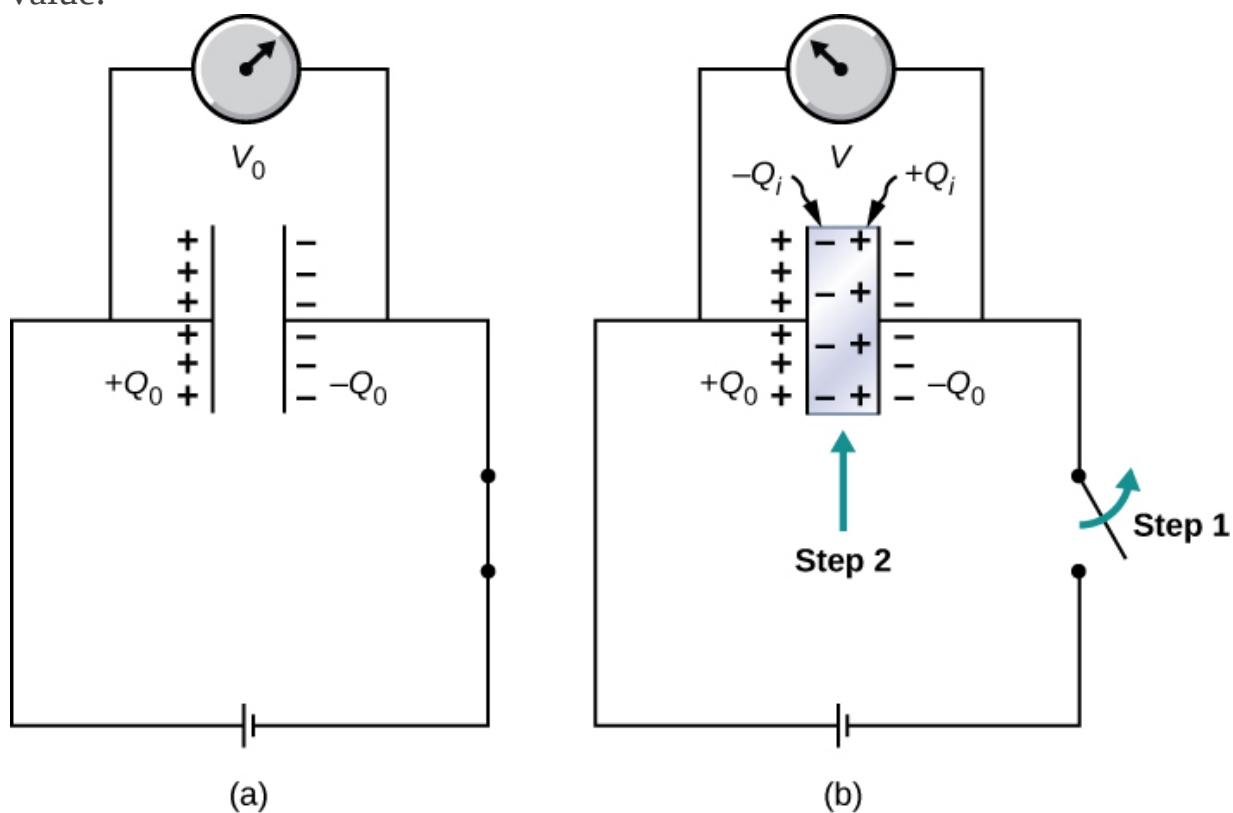
The constant κ in this equation is called the **dielectric constant** of the material between the plates, and its value is characteristic for the material. A detailed explanation for why the dielectric reduces the voltage is given in the next section. Different materials have different dielectric constants (a table of values for typical materials is provided in the next section). Once the battery becomes disconnected, there is no path for a charge to flow to the battery from the capacitor plates. Hence, the insertion of the dielectric has no effect on the charge on the plate, which remains at a value of Q_0 . Therefore, we find that the capacitance of the capacitor with a dielectric is

Note:

Equation:

$$C = \frac{Q_0}{V} = \frac{Q_0}{V_0/\kappa} = \kappa \frac{Q_0}{V_0} = \kappa C_0.$$

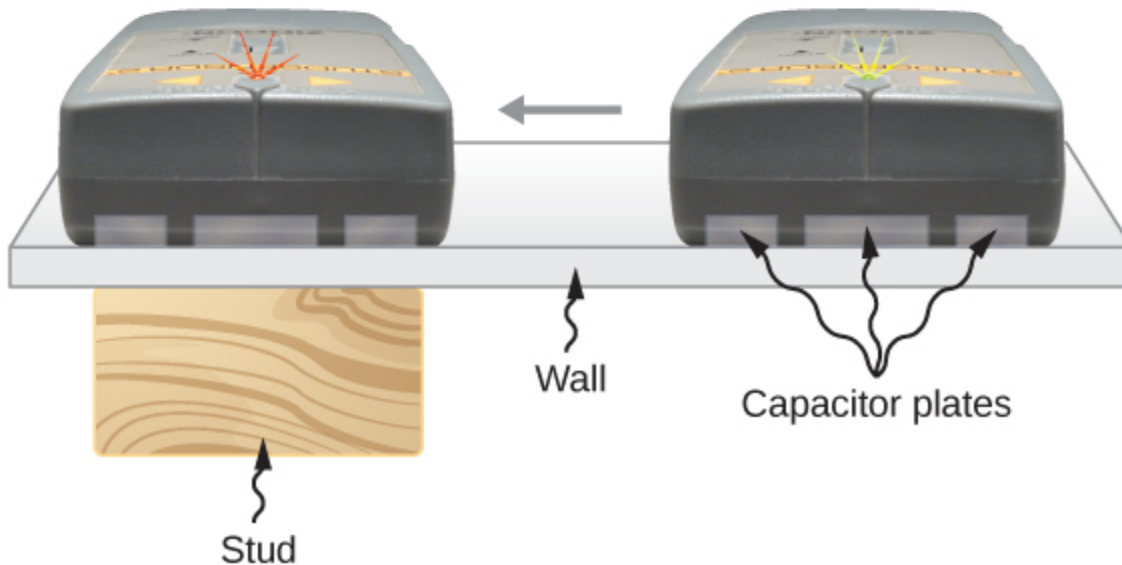
This equation tells us that the *capacitance* C_0 of an empty (vacuum) capacitor can be increased by a factor of κ when we insert a dielectric material to completely fill the space between its plates. Note that [\[link\]](#) can also be used for an empty capacitor by setting $\kappa = 1$. In other words, we can say that the dielectric constant of the vacuum is 1, which is a reference value.



(a) When fully charged, a vacuum capacitor has a voltage V_0 and charge Q_0 (the charges remain on plate's inner surfaces; the schematic indicates the sign of charge on each plate). (b) In step 1, the battery is disconnected. Then, in step 2, a dielectric (that is electrically neutral)

is inserted into the charged capacitor. When the voltage across the capacitor is now measured, it is found that the voltage value has decreased to $V = V_0/\kappa$. The schematic indicates the sign of the induced charge that is now present on the surfaces of the dielectric material between the plates.

The principle expressed by [\[link\]](#) is widely used in the construction industry ([\[link\]](#)). Metal plates in an electronic stud finder act effectively as a capacitor. You place a stud finder with its flat side on the wall and move it continually in the horizontal direction. When the finder moves over a wooden stud, the capacitance of its plates changes, because wood has a different dielectric constant than a gypsum wall. This change triggers a signal in a circuit, and thus the stud is detected.



An electronic stud finder is used to detect wooden studs behind drywall. (credit top: modification of work by Jane Whitney)

The electrical energy stored by a capacitor is also affected by the presence of a dielectric. When the energy stored in an empty capacitor is U_0 , the energy U stored in a capacitor with a dielectric is smaller by a factor of κ ,

Note:

Equation:

$$U = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} \frac{Q_0^2}{\kappa C_0} = \frac{1}{\kappa} U_0.$$

As a dielectric material sample is brought near an empty charged capacitor, the sample reacts to the electrical field of the charges on the capacitor plates. Just as we learned in [Electric Charges and Fields](#) on electrostatics, there will be the induced charges on the surface of the sample; however, they are not free charges like in a conductor, because a perfect insulator does not have freely moving charges. These induced charges on the dielectric surface are of an opposite sign to the free charges on the plates of the capacitor, and so they are attracted by the free charges on the plates. Consequently, the dielectric is “pulled” into the gap, and the work to polarize the dielectric material between the plates is done at the expense of the stored electrical energy, which is reduced, in accordance with [\[link\]](#).

Example:

Inserting a Dielectric into an Isolated Capacitor

An empty 20.0-pF capacitor is charged to a potential difference of 40.0 V. The charging battery is then disconnected, and a piece of Teflon™ with a dielectric constant of 2.1 is inserted to completely fill the space between the capacitor plates (see [\[link\]](#)). What are the values of (a) the capacitance, (b) the charge of the plate, (c) the potential difference between the plates, and (d) the energy stored in the capacitor with and without dielectric?

Strategy

We identify the original capacitance $C_0 = 20.0$ pF and the original potential difference $V_0 = 40.0$ V between the plates. We combine [\[link\]](#) with other relations involving capacitance and substitute.

Solution

- a. The capacitance increases to

Equation:

$$C = \kappa C_0 = 2.1(20.0 \text{ pF}) = 42.0 \text{ pF}.$$

- b. Without dielectric, the charge on the plates is

Equation:

$$Q_0 = C_0 V_0 = (20.0 \text{ pF})(40.0 \text{ V}) = 0.8 \text{ nC}.$$

Since the battery is disconnected before the dielectric is inserted, the plate charge is unaffected by the dielectric and remains at 0.8 nC.

- c. With the dielectric, the potential difference becomes

Equation:

$$V = \frac{1}{\kappa} V_0 = \frac{1}{2.1} 40.0 \text{ V} = 19.0 \text{ V}.$$

- d. The stored energy without the dielectric is

Equation:

$$U_0 = \frac{1}{2} C_0 V_0^2 = \frac{1}{2} (20.0 \text{ pF})(40.0 \text{ V})^2 = 16.0 \text{ nJ}.$$

With the dielectric inserted, we use [\[link\]](#) to find that the stored energy decreases to

Equation:

$$U = \frac{1}{\kappa} U_0 = \frac{1}{2.1} 16.0 \text{ nJ} = 7.6 \text{ nJ}.$$

Significance

Notice that the effect of a dielectric on the capacitance of a capacitor is a drastic increase of its capacitance. This effect is far more profound than a mere change in the geometry of a capacitor.

Note:

Exercise:**Problem:**

Check Your Understanding When a dielectric is inserted into an isolated and charged capacitor, the stored energy decreases to 33% of its original value. (a) What is the dielectric constant? (b) How does the capacitance change?

Solution:

a. 3.0; b. $C = 3.0 C_0$

Summary

- The capacitance of an empty capacitor is increased by a factor of κ when the space between its plates is completely filled by a dielectric with dielectric constant κ .
- Each dielectric material has its specific dielectric constant.
- The energy stored in an empty isolated capacitor is decreased by a factor of κ when the space between its plates is completely filled with a dielectric with dielectric constant κ while disconnecting the battery and keeping the charge on the capacitor constant.

Conceptual Questions**Exercise:****Problem:**

Discuss what would happen if a conducting slab rather than a dielectric were inserted into the gap between the capacitor plates.

Solution:

answers may vary

Exercise:**Problem:**

Discuss how the energy stored in an empty but charged capacitor changes when a dielectric is inserted if (a) the capacitor is isolated so that its charge does not change; (b) the capacitor remains connected to a battery so that the potential difference between its plates does not change.

Problems**Exercise:****Problem:**

Show that for a given dielectric material, the maximum energy a parallel-plate capacitor can store is directly proportional to the volume of dielectric.

Exercise:**Problem:**

An air-filled capacitor is made from two flat parallel plates 1.0 mm apart. The inside area of each plate is 8.0 cm^2 . (a) What is the capacitance of this set of plates? (b) If the region between the plates is filled with a material whose dielectric constant is 6.0, what is the new capacitance?

Solution:

a. 7.1 pF; b. 42 pF

Exercise:

Problem:

A capacitor is made from two concentric spheres, one with radius 5.00 cm, the other with radius 8.00 cm. (a) What is the capacitance of this set of conductors? (b) If the region between the conductors is filled with a material whose dielectric constant is 6.00, what is the capacitance of the system?

Exercise:**Problem:**

A parallel-plate capacitor has charge of magnitude $9.00 \mu\text{C}$ on each plate and capacitance $3.00 \mu\text{F}$ when there is air between the plates. The plates are separated by 2.00 mm. With the charge on the plates kept constant, a dielectric with $\kappa = 5$ is inserted between the plates, completely filling the volume between the plates. (a) What is the potential difference between the plates of the capacitor, before and after the dielectric has been inserted? (b) What is the electrical field at the point midway between the plates before and after the dielectric is inserted?

Solution:

a. before 3.00 V; after 0.600 V; b. before 1500 V/m; after 300 V/m

Exercise:

Problem:

Some cell walls in the human body have a layer of negative charge on the inside surface. Suppose that the surface charge densities are $\pm 0.50 \times 10^{-3} \text{C/m}^2$, the cell wall is $5.0 \times 10^{-9} \text{m}$ thick, and the cell wall material has a dielectric constant of $\kappa = 5.4$. (a) Find the magnitude of the electric field in the wall between two charge layers. (b) Find the potential difference between the inside and the outside of the cell. Which is at higher potential? (c) A typical cell in the human body has volume 10^{-16}m^3 . Estimate the total electrical field energy stored in the wall of a cell of this size when assuming that the cell is spherical. (*Hint: Calculate the volume of the cell wall.*)

Exercise:**Problem:**

A parallel-plate capacitor with only air between its plates is charged by connecting the capacitor to a battery. The capacitor is then disconnected from the battery, without any of the charge leaving the plates. (a) A voltmeter reads 45.0 V when placed across the capacitor. When a dielectric is inserted between the plates, completely filling the space, the voltmeter reads 11.5 V. What is the dielectric constant of the material? (b) What will the voltmeter read if the dielectric is now pulled away out so it fills only one-third of the space between the plates?

Solution:

a. 3.91; b. 22.8 V

Glossary

dielectric constant

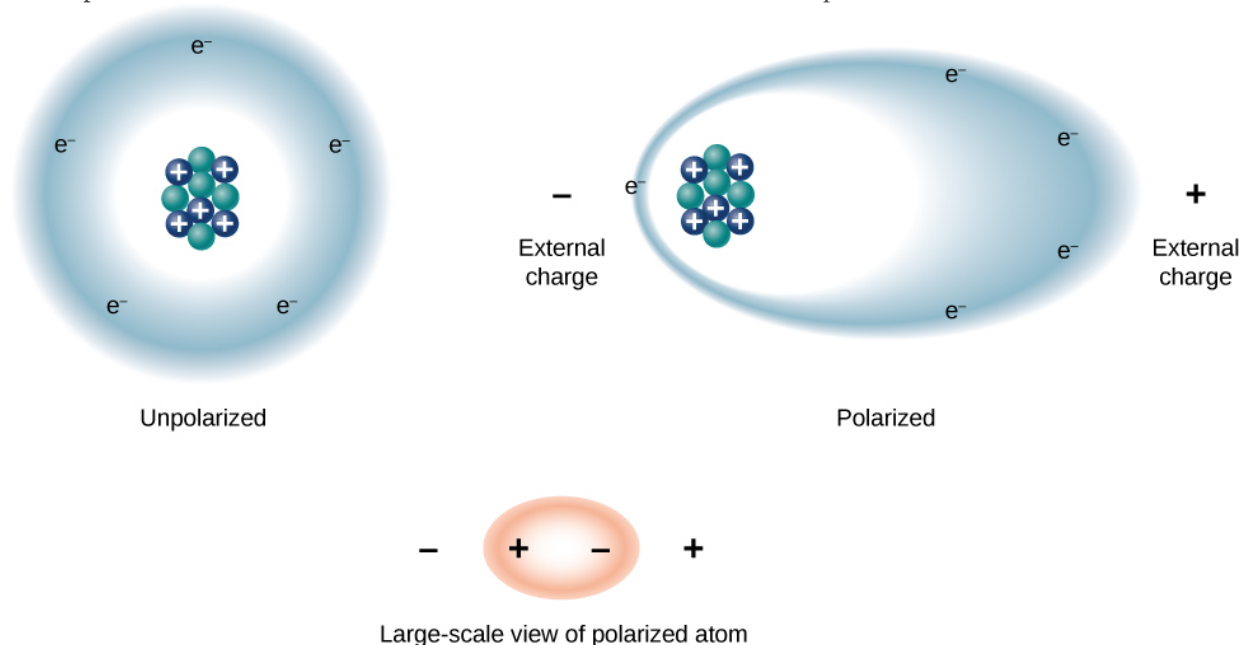
factor by which capacitance increases when a dielectric is inserted between the plates of a capacitor

Molecular Model of a Dielectric

By the end of this section, you will be able to:

- Explain the polarization of a dielectric in a uniform electrical field
- Describe the effect of a polarized dielectric on the electrical field between capacitor plates
- Explain dielectric breakdown

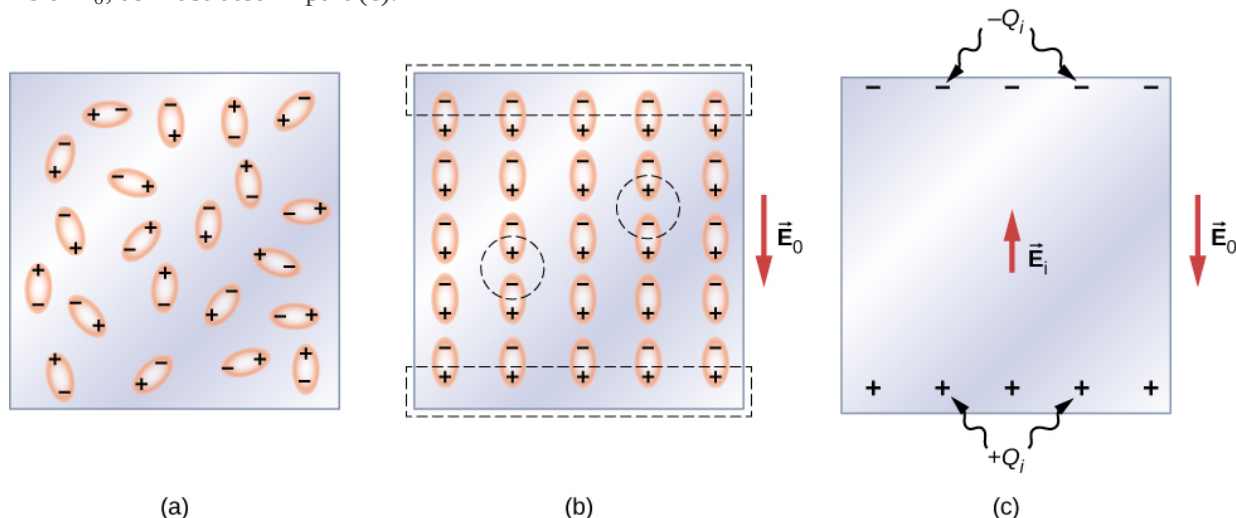
We can understand the effect of a dielectric on capacitance by looking at its behavior at the molecular level. As we have seen in earlier chapters, in general, all molecules can be classified as either *polar* or *nonpolar*. There is a net separation of positive and negative charges in an isolated polar molecule, whereas there is no charge separation in an isolated nonpolar molecule ([\[link\]](#)). In other words, polar molecules have permanent *electric-dipole moments* and nonpolar molecules do not. For example, a molecule of water is polar, and a molecule of oxygen is nonpolar. Nonpolar molecules can become polar in the presence of an external electrical field, which is called *induced polarization*.



The concept of polarization: In an unpolarized atom or molecule, a negatively charged electron cloud is evenly distributed around positively charged centers, whereas a polarized atom or molecule has an excess of negative charge at one side so that the other side has an excess of positive charge. However, the entire system remains electrically neutral. The charge polarization may be caused by an external electrical field. Some molecules and atoms are permanently polarized (electric dipoles) even in the absence of an external electrical field (polar molecules and atoms).

Let's first consider a dielectric composed of polar molecules. In the absence of any external electrical field, the electric dipoles are oriented randomly, as illustrated in [\[link\]\(a\)](#). However, if the dielectric is placed in an external electrical field \vec{E}_0 , the polar molecules align with the external field, as shown in part (b) of the figure. Opposite charges on adjacent dipoles within the volume of dielectric neutralize each other, so there is no net charge within the dielectric (see the dashed circles in part (b)). However, this is not the case very close to the upper and lower surfaces that border the dielectric (the region enclosed by the dashed rectangles in part (b)), where the alignment does produce a net charge. Since the

external electrical field merely aligns the dipoles, the dielectric as a whole is neutral, and the surface charges induced on its opposite faces are equal and opposite. These **induced surface charges** $+Q_i$ and $-Q_i$ produce an additional electrical field \vec{E}_i (an **induced electrical field**), which *opposes* the external field \vec{E}_0 , as illustrated in part (c).



A dielectric with polar molecules: (a) In the absence of an external electrical field; (b) in the presence of an external electrical field \vec{E}_0 . The dashed lines indicate the regions immediately adjacent to the capacitor plates. (c) The induced electrical field \vec{E}_i inside the dielectric produced by the induced surface charge Q_i of the dielectric. Note that, in reality, the individual molecules are not perfectly aligned with an external field because of thermal fluctuations; however, the *average* alignment is along the field lines as shown.

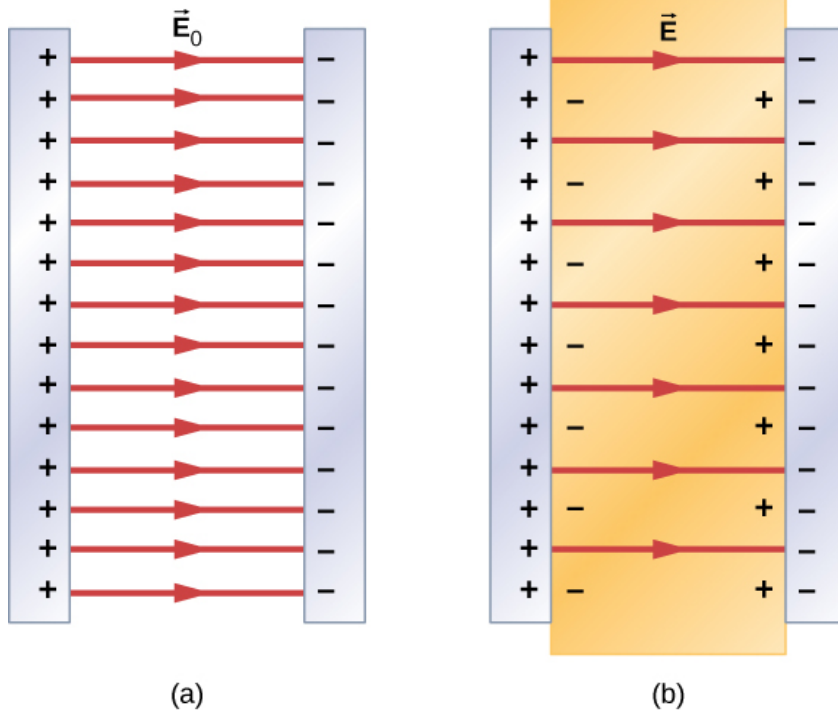
The same effect is produced when the molecules of a dielectric are nonpolar. In this case, a nonpolar molecule acquires an **induced electric-dipole moment** because the external field \vec{E}_0 causes a separation between its positive and negative charges. The induced dipoles of the nonpolar molecules align with \vec{E}_0 in the same way as the permanent dipoles of the polar molecules are aligned (shown in part (b)). Hence, the electrical field within the dielectric is weakened regardless of whether its molecules are polar or nonpolar.

Therefore, when the region between the parallel plates of a charged capacitor, such as that shown in [\[link\]](#)(a), is filled with a dielectric, within the dielectric there is an electrical field \vec{E}_0 due to the *free* charge Q_0 on the capacitor plates and an electrical field \vec{E}_i due to the induced charge Q_i on the surfaces of the dielectric. Their vector sum gives the net electrical field \vec{E} within the dielectric between the capacitor plates (shown in part (b) of the figure):

Equation:

$$\vec{E} = \vec{E}_0 + \vec{E}_i.$$

This net field can be considered to be the field produced by an *effective charge* $Q_0 - Q_i$ on the capacitor.



Electrical field: (a) In an empty capacitor, electrical field \vec{E}_0 . (b) In a dielectric-filled capacitor, electrical field \vec{E} .

In most dielectrics, the net electrical field \vec{E} is proportional to the field \vec{E}_0 produced by the free charge. In terms of these two electrical fields, the dielectric constant κ of the material is defined as

Note:

Equation:

$$\kappa = \frac{E_0}{E}.$$

Since \vec{E}_0 and \vec{E}_i point in opposite directions, the magnitude E is smaller than the magnitude E_0 and therefore $\kappa > 1$. Combining [\[link\]](#) with [\[link\]](#), and rearranging the terms, yields the following expression for the induced electrical field in a dielectric:

Note:

Equation:

$$\vec{E}_i = \left(\frac{1}{\kappa} - 1 \right) \vec{E}_0.$$

When the magnitude of an external electrical field becomes too large, the molecules of dielectric material start to become ionized. A molecule or an atom is ionized when one or more electrons are removed from it and become free electrons, no longer bound to the molecular or atomic structure. When this happens, the material can conduct, thereby allowing charge to move through the dielectric from one capacitor plate to the other. This phenomenon is called **dielectric breakdown**. ([link](#) shows typical random-path patterns of electrical discharge during dielectric breakdown.) The critical value, E_c , of the electrical field at which the molecules of an insulator become ionized is called the **dielectric strength** of the material. The dielectric strength imposes a limit on the voltage that can be applied for a given plate separation in a capacitor. For example, the dielectric strength of air is $E_c = 3.0 \text{ MV/m}$, so for an air-filled capacitor with a plate separation of $d = 1.00 \text{ mm}$, the limit on the potential difference that can be safely applied across its plates without causing dielectric breakdown is $V = E_c d = (3.0 \times 10^6 \text{ V/m})(1.00 \times 10^{-3} \text{ m}) = 3.0 \text{ kV}$.

However, this limit becomes 60.0 kV when the same capacitor is filled with Teflon™, whose dielectric strength is about 60.0 MV/m. Because of this limit imposed by the dielectric strength, the amount of charge that an air-filled capacitor can store is only $Q_0 = \kappa_{\text{air}} C_0 (3.0 \text{ kV})$ and the charge stored on the same Teflon™-filled capacitor can be as much as

Equation:

$$Q = \kappa_{\text{teflon}} C_0 (60.0 \text{ kV}) = \kappa_{\text{teflon}} \frac{Q_0}{\kappa_{\text{air}} (3.0 \text{ kV})} (60.0 \text{ kV}) = 20 \frac{\kappa_{\text{teflon}}}{\kappa_{\text{air}}} Q_0 = 20 \frac{2.1}{1.00059} Q_0 \cong 42 Q_0,$$

which is about 42 times greater than a charge stored on an air-filled capacitor. Typical values of dielectric constants and dielectric strengths for various materials are given in [link](#). Notice that the dielectric constant κ is exactly 1.0 for a vacuum (the empty space serves as a reference condition) and very close to 1.0 for air under normal conditions (normal pressure at room temperature). These two values are so close that, in fact, the properties of an air-filled capacitor are essentially the same as those of an empty capacitor.

Material	Dielectric constant κ	Dielectric strength $E_c [\times 10^6 \text{ V/m}]$
Vacuum	1	∞
Dry air (1 atm)	1.00059	3.0
Teflon™	2.1	60 to 173
Paraffin	2.3	11

Material	Dielectric constant κ	Dielectric strength $E_c [\times 10^6 \text{V/m}]$
Silicon oil	2.5	10 to 15
Polystyrene	2.56	19.7
Nylon	3.4	14
Paper	3.7	16
Fused quartz	3.78	8
Glass	4 to 6	9.8 to 13.8
Concrete	4.5	—
Bakelite	4.9	24
Diamond	5.5	2,000
Pyrex glass	5.6	14
Mica	6.0	118
Neoprene rubber	6.7	15.7 to 26.7
Water	80	—
Sulfuric acid	84 to 100	—
Titanium dioxide	86 to 173	—
Strontium titanate	310	8
Barium titanate	1,200 to 10,000	—
Calcium copper titanate	> 250,000	—

Representative Values of Dielectric Constants and Dielectric Strengths of Various Materials at Room Temperature

Not all substances listed in the table are good insulators, despite their high dielectric constants. Water, for example, consists of polar molecules and has a large dielectric constant of about 80. In a water molecule, electrons are more likely found around the oxygen nucleus than around the hydrogen nuclei. This makes the oxygen end of the molecule slightly negative and leaves the hydrogens end slightly positive, which makes the molecule easy to align along an external electrical field, and thus water has a large dielectric constant. However, the polar nature of water molecules also makes water a good solvent for many substances, which produces undesirable effects, because any concentration of free ions in water conducts electricity.

Example:**Electrical Field and Induced Surface Charge**

Suppose that the distance between the plates of the capacitor in [\[link\]](#) is 2.0 mm and the area of each plate is $4.5 \times 10^{-3} \text{ m}^2$. Determine: (a) the electrical field between the plates before and after the Teflon™ is inserted, and (b) the surface charge induced on the Teflon™ surfaces.

Strategy

In part (a), we know that the voltage across the empty capacitor is $V_0 = 40 \text{ V}$, so to find the electrical fields we use the relation $V = Ed$ and [\[link\]](#). In part (b), knowing the magnitude of the electrical field, we use the expression for the magnitude of electrical field near a charged plate $E = \sigma/\epsilon_0$, where σ is a uniform surface charge density caused by the surface charge. We use the value of free charge $Q_0 = 8.0 \times 10^{-10} \text{ C}$ obtained in [\[link\]](#).

Solution

- a. The electrical field E_0 between the plates of an empty capacitor is

Equation:

$$E_0 = \frac{V_0}{d} = \frac{40 \text{ V}}{2.0 \times 10^{-3} \text{ m}} = 2.0 \times 10^4 \text{ V/m}.$$

The electrical field E with the Teflon™ in place is

Equation:

$$E = \frac{1}{\kappa} E_0 = \frac{1}{2.1} 2.0 \times 10^4 \text{ V/m} = 9.5 \times 10^3 \text{ V/m}.$$

- b. The effective charge on the capacitor is the difference between the free charge Q_0 and the induced charge Q_i . The electrical field in the Teflon™ is caused by this effective charge. Thus

Equation:

$$E = \frac{1}{\epsilon_0} \sigma = \frac{1}{\epsilon_0} \frac{Q_0 - Q_i}{A}.$$

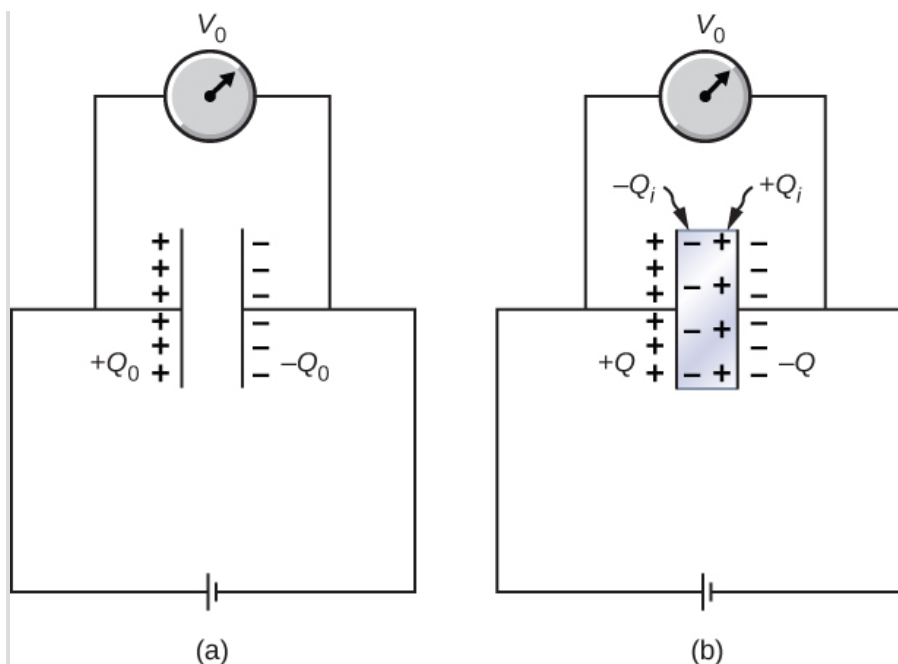
We invert this equation to obtain Q_i , which yields

Equation:

$$\begin{aligned} Q_i &= Q_0 - \epsilon_0 A E \\ &= 8.0 \times 10^{-10} \text{ C} - \left(8.85 \times 10^{-12} \frac{\text{C}^2}{\text{N} \cdot \text{m}^2} \right) (4.5 \times 10^{-3} \text{ m}^2) (9.5 \times 10^3 \frac{\text{V}}{\text{m}}) \\ &= 4.2 \times 10^{-10} \text{ C} = 0.42 \text{ nC}. \end{aligned}$$

Example:**Inserting a Dielectric into a Capacitor Connected to a Battery**

When a battery of voltage V_0 is connected across an empty capacitor of capacitance C_0 , the charge on its plates is Q_0 , and the electrical field between its plates is E_0 . A dielectric of dielectric constant κ is inserted between the plates *while the battery remains in place*, as shown in [\[link\]](#). (a) Find the capacitance C , the voltage V across the capacitor, and the electrical field E between the plates after the dielectric is inserted. (b) Obtain an expression for the free charge Q on the plates of the filled capacitor and the induced charge Q_i on the dielectric surface in terms of the original plate charge Q_0 .



A dielectric is inserted into the charged capacitor while the capacitor remains connected to the battery.

Strategy

We identify the known values: V_0 , C_0 , E_0 , κ , and Q_0 . Our task is to express the unknown values in terms of these known values.

Solution

(a) The capacitance of the filled capacitor is $C = \kappa C_0$. Since the battery is always connected to the capacitor plates, the potential difference between them does not change; hence, $V = V_0$. Because of that, the electrical field in the filled capacitor is the same as the field in the empty capacitor, so we can obtain directly that

Equation:

$$E = \frac{V}{d} = \frac{V_0}{d} = E_0.$$

(b) For the filled capacitor, the free charge on the plates is

Equation:

$$Q = CV = (\kappa C_0)V_0 = \kappa(C_0V_0) = \kappa Q_0.$$

The electrical field E in the filled capacitor is due to the effective charge $Q - Q_i$ ([link](#)(b)). Since $E = E_0$, we have

Equation:

$$\frac{Q - Q_i}{\epsilon_0 A} = \frac{Q_0}{\epsilon_0 A}.$$

Solving this equation for Q_i , we obtain for the induced charge

Equation:

$$Q_i = Q - Q_0 = \kappa Q_0 - Q_0 = (\kappa - 1)Q_0.$$

Significance

Notice that for materials with dielectric constants larger than 2 (see [\[link\]](#)), the induced charge on the surface of dielectric is larger than the charge on the plates of a vacuum capacitor. The opposite is true for gasses like air whose dielectric constant is smaller than 2.

Note:

Exercise:

Problem:

Check Your Understanding Continuing with [\[link\]](#), show that when the battery is connected across the plates the energy stored in dielectric-filled capacitor is $U = \kappa U_0$ (larger than the energy U_0 of an empty capacitor kept at the same voltage). Compare this result with the result $U = U_0/\kappa$ found previously for an isolated, charged capacitor.

Note:

Exercise:

Problem:

Check Your Understanding Repeat the calculations of [\[link\]](#) for the case in which the battery remains connected while the dielectric is placed in the capacitor.

Solution:

a. $C_0 = 20 \text{ pF}$, $C = 42 \text{ pF}$; b. $Q_0 = 0.8 \text{ nC}$, $Q = 1.7 \text{ nC}$; c. $V_0 = V = 40 \text{ V}$; d. $U_0 = 16 \text{ nJ}$, $U = 34 \text{ nJ}$

Summary

- When a dielectric is inserted between the plates of a capacitor, equal and opposite surface charge is induced on the two faces of the dielectric. The induced surface charge produces an induced electrical field that opposes the field of the free charge on the capacitor plates.
- The dielectric constant of a material is the ratio of the electrical field in vacuum to the net electrical field in the material. A capacitor filled with dielectric has a larger capacitance than an empty capacitor.
- The dielectric strength of an insulator represents a critical value of electrical field at which the molecules in an insulating material start to become ionized. When this happens, the material can conduct and dielectric breakdown is observed.

Key Equations

Capacitance	$C = \frac{Q}{V}$
Capacitance of a parallel-plate capacitor	$C = \epsilon_0 \frac{A}{d}$
Capacitance of a vacuum spherical capacitor	$C = 4\pi\epsilon_0 \frac{R_1 R_2}{R_2 - R_1}$
Capacitance of a vacuum cylindrical capacitor	$C = \frac{2\pi\epsilon_0 l}{\ln(R_2/R_1)}$
Capacitance of a series combination	$\frac{1}{C_s} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \dots$
Capacitance of a parallel combination	$C_P = C_1 + C_2 + C_3 + \dots$
Energy density	$u_E = \frac{1}{2}\epsilon_0 E^2$
Energy stored in a capacitor	$U_C = \frac{1}{2} V^2 C = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} QV$
Capacitance of a capacitor with dielectric	$C = \kappa C_0$
Energy stored in an isolated capacitor with dielectric	$U = \frac{1}{\kappa} U_0$
Dielectric constant	$\kappa = \frac{E_0}{E}$
Induced electrical field in a dielectric	$\vec{E}_i = \left(\frac{1}{\kappa} - 1\right) \vec{E}_0$

Conceptual Questions

Exercise:

Problem: Distinguish between dielectric strength and dielectric constant.

Solution:

Dielectric strength is a critical value of an electrical field above which an insulator starts to conduct; a dielectric constant is the ratio of the electrical field in vacuum to the net electrical field in a material.

Exercise:

Problem: Water is a good solvent because it has a high dielectric constant. Explain.

Exercise:

Problem:

Water has a high dielectric constant. Explain why it is then not used as a dielectric material in capacitors.

Solution:

Water is a good solvent.

Exercise:**Problem:**

Elaborate on why molecules in a dielectric material experience net forces on them in a non-uniform electrical field but not in a uniform field.

Exercise:**Problem:**

Explain why the dielectric constant of a substance containing permanent molecular electric dipoles decreases with increasing temperature.

Solution:

When energy of thermal motion is large (high temperature), an electrical field must be large too in order to keep electric dipoles aligned with it.

Exercise:**Problem:**

Give a reason why a dielectric material increases capacitance compared with what it would be with air between the plates of a capacitor. How does a dielectric material also allow a greater voltage to be applied to a capacitor? (The dielectric thus increases C and permits a greater V .)

Exercise:**Problem:**

Elaborate on the way in which the polar character of water molecules helps to explain water's relatively large dielectric constant.

Solution:

answers may vary

Exercise:**Problem:**

Sparks will occur between the plates of an air-filled capacitor at a lower voltage when the air is humid than when it is dry. Discuss why, considering the polar character of water molecules.

Problems**Exercise:**

Problem:

Two flat plates containing equal and opposite charges are separated by material 4.0 mm thick with a dielectric constant of 5.0. If the electrical field in the dielectric is 1.5 MV/m, what are (a) the charge density on the capacitor plates, and (b) the induced charge density on the surfaces of the dielectric?

Exercise:**Problem:**

For a TeflonTM-filled, parallel-plate capacitor, the area of the plate is 50.0 cm² and the spacing between the plates is 0.50 mm. If the capacitor is connected to a 200-V battery, find (a) the free charge on the capacitor plates, (b) the electrical field in the dielectric, and (c) the induced charge on the dielectric surfaces.

Solution:

a. 37 nC; b. 0.4 MV/m; c. 19 nC

Exercise:**Problem:**

Find the capacitance of a parallel-plate capacitor having plates with a surface area of 5.00 m² and separated by 0.100 mm of TeflonTM.

Exercise:**Problem:**

(a) What is the capacitance of a parallel-plate capacitor with plates of area 1.50 m² that are separated by 0.0200 mm of neoprene rubber? (b) What charge does it hold when 9.00 V is applied to it?

Solution:

a. 4.4 μF; b. 4.0×10^{-5} C

Exercise:**Problem:**

Two parallel plates have equal and opposite charges. When the space between the plates is evacuated, the electrical field is $E = 3.20 \times 10^5$ V/m. When the space is filled with dielectric, the electrical field is $E = 2.50 \times 10^5$ V/m. (a) What is the surface charge density on each surface of the dielectric? (b) What is the dielectric constant?

Exercise:**Problem:**

The dielectric to be used in a parallel-plate capacitor has a dielectric constant of 3.60 and a dielectric strength of 1.60×10^7 V/m. The capacitor has to have a capacitance of 1.25 nF and must be able to withstand a maximum potential difference 5.5 kV. What is the minimum area the plates of the capacitor may have?

Solution:

$$0.0135 \text{ m}^2$$

Exercise:**Problem:**

When a 360-nF air capacitor is connected to a power supply, the energy stored in the capacitor is $18.5 \mu\text{J}$. While the capacitor is connected to the power supply, a slab of dielectric is inserted that completely fills the space between the plates. This increases the stored energy by $23.2 \mu\text{J}$. (a) What is the potential difference between the capacitor plates? (b) What is the dielectric constant of the slab?

Exercise:**Problem:**

A parallel-plate capacitor has square plates that are 8.00 cm on each side and 3.80 mm apart. The space between the plates is completely filled with two square slabs of dielectric, each 8.00 cm on a side and 1.90 mm thick. One slab is Pyrex glass and the other slab is polystyrene. If the potential difference between the plates is 86.0 V, find how much electrical energy can be stored in this capacitor.

Solution:

$$0.185 \mu\text{J}$$

Additional Problems**Exercise:****Problem:**

A capacitor is made from two flat parallel plates placed 0.40 mm apart. When a charge of $0.020 \mu\text{C}$ is placed on the plates the potential difference between them is 250 V. (a) What is the capacitance of the plates? (b) What is the area of each plate? (c) What is the charge on the plates when the potential difference between them is 500 V? (d) What maximum potential difference can be applied between the plates so that the magnitude of electrical fields between the plates does not exceed 3.0 MV/m?

Exercise:**Problem:**

An air-filled (empty) parallel-plate capacitor is made from two square plates that are 25 cm on each side and 1.0 mm apart. The capacitor is connected to a 50-V battery and fully charged. It is then disconnected from the battery and its plates are pulled apart to a separation of 2.00 mm. (a) What is the capacitance of this new capacitor? (b) What is the charge on each plate? (c) What is the electrical field between the plates?

Solution:

$$\text{a. } 0.277 \text{ nF; b. } 27.7 \text{ nC; c. } 50 \text{ kV/m}$$

Exercise:

Problem:

Suppose that the capacitance of a variable capacitor can be manually changed from 100 to 800 pF by turning a dial connected to one set of plates by a shaft, from 0° to 180° . With the dial set at 180° (corresponding to $C = 800$ pF), the capacitor is connected to a 500-V source. After charging, the capacitor is disconnected from the source, and the dial is turned to 0° . (a) What is the charge on the capacitor? (b) What is the voltage across the capacitor when the dial is set to 0° ?

Exercise:**Problem:**

Earth can be considered as a spherical capacitor with two plates, where the negative plate is the surface of Earth and the positive plate is the bottom of the ionosphere, which is located at an altitude of approximately 70 km. The potential difference between Earth's surface and the ionosphere is about 350,000 V. (a) Calculate the capacitance of this system. (b) Find the total charge on this capacitor. (c) Find the energy stored in this system.

Solution:

a. 0.065 F; b. 23,000 C; c. 4.0 GJ

Exercise:**Problem:**

A $4.00\text{-}\mu\text{F}$ capacitor and a $6.00\text{-}\mu\text{F}$ capacitor are connected in parallel across a 600-V supply line. (a) Find the charge on each capacitor and voltage across each. (b) The charged capacitors are disconnected from the line and from each other. They are then reconnected to each other with terminals of unlike sign together. Find the final charge on each capacitor and the voltage across each.

Exercise:**Problem:**

Three capacitors having capacitances of 8.40, 8.40, and $4.20\text{ }\mu\text{F}$, respectively, are connected in series across a 36.0-V potential difference. (a) What is the charge on the $4.20\text{-}\mu\text{F}$ capacitor? (b) The capacitors are disconnected from the potential difference without allowing them to discharge. They are then reconnected in parallel with each other with the positively charged plates connected together. What is the voltage across each capacitor in the parallel combination?

Solution:

a. $75.6\text{ }\mu\text{C}$; b. 10.8 V

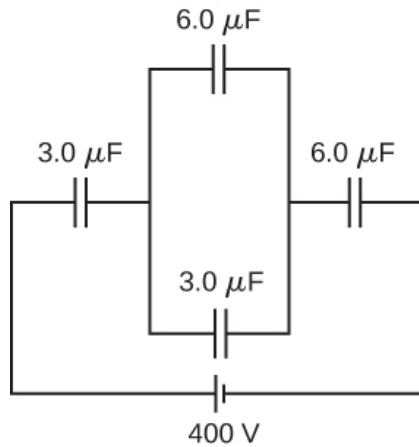
Exercise:**Problem:**

A parallel-plate capacitor with capacitance $5.0\text{ }\mu\text{F}$ is charged with a 12.0-V battery, after which the battery is disconnected. Determine the minimum work required to increase the separation between the plates by a factor of 3.

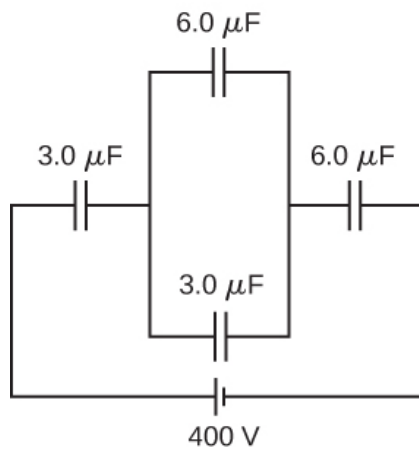
Exercise:

Problem:

(a) How much energy is stored in the electrical fields in the capacitors (in total) shown below? (b) Is this energy equal to the work done by the 400-V source in charging the capacitors?

**Solution:**

a. $0.13\ \text{J}$; b. no, because of resistive heating in connecting wires that is always present, but the circuit schematic does not indicate resistors

**Exercise:****Problem:**

Three capacitors having capacitances 8.4 , 8.4 , and $4.2\ \mu\text{F}$ are connected in series across a 36.0-V potential difference. (a) What is the total energy stored in all three capacitors? (b) The capacitors are disconnected from the potential difference without allowing them to discharge. They are then reconnected in parallel with each other with the positively charged plates connected together. What is the total energy now stored in the capacitors?

Exercise:

Problem:

(a) An $8.00\text{-}\mu\text{F}$ capacitor is connected in parallel to another capacitor, producing a total capacitance of $5.00\text{ }\mu\text{F}$. What is the capacitance of the second capacitor? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. $-3.00\text{ }\mu\text{F}$; b. You cannot have a negative C_2 capacitance. c. The assumption that they were hooked up in parallel, rather than in series, is incorrect. A parallel connection always produces a greater capacitance, while here a smaller capacitance was assumed. This could only happen if the capacitors are connected in series.

Exercise:**Problem:**

(a) On a particular day, it takes $9.60 \times 10^3\text{ J}$ of electrical energy to start a truck's engine. Calculate the capacitance of a capacitor that could store that amount of energy at 12.0 V . (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:**Problem:**

(a) A certain parallel-plate capacitor has plates of area 4.00 m^2 , separated by 0.0100 mm of nylon, and stores 0.170 C of charge. What is the applied voltage? (b) What is unreasonable about this result? (c) Which assumptions are responsible or inconsistent?

Solution:

a. 14.2 kV ; b. The voltage is unreasonably large, more than 100 times the breakdown voltage of nylon. c. The assumed charge is unreasonably large and cannot be stored in a capacitor of these dimensions.

Exercise:**Problem:**

A prankster applies 450 V to an $80.0\text{-}\mu\text{F}$ capacitor and then tosses it to an unsuspecting victim. The victim's finger is burned by the discharge of the capacitor through 0.200 g of flesh. Estimate, what is the temperature increase of the flesh? Is it reasonable to assume that no thermodynamic phase change happened?

Challenge Problems**Exercise:**

Problem:

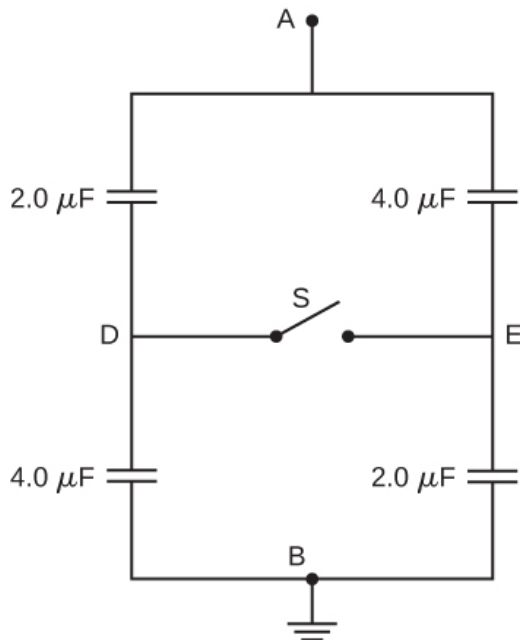
A spherical capacitor is formed from two concentric spherical conducting spheres separated by vacuum. The inner sphere has radius 12.5 cm and the outer sphere has radius 14.8 cm. A potential difference of 120 V is applied to the capacitor. (a) What is the capacitance of the capacitor? (b) What is the magnitude of the electrical field at $r = 12.6$ cm, just outside the inner sphere? (c) What is the magnitude of the electrical field at $r = 14.7$ cm, just inside the outer sphere? (d) For a parallel-plate capacitor the electrical field is uniform in the region between the plates, except near the edges of the plates. Is this also true for a spherical capacitor?

Solution:

a. 89.6 pF; b. 6.09 kV/m; c. 4.47 kV/m; d. no

Exercise:**Problem:**

The network of capacitors shown below are all uncharged when a 300-V potential is applied between points A and B with the switch S open. (a) What is the potential difference $V_E - V_D$? (b) What is the potential at point E after the switch is closed? (c) How much charge flows through the switch after it is closed?

**Exercise:****Problem:**

Electronic flash units for cameras contain a capacitor for storing the energy used to produce the flash. In one such unit the flash lasts for $1/675$ fraction of a second with an average light power output of 270 kW. (a) If the conversion of electrical energy to light is 95% efficient (because the rest of the energy goes to thermal energy), how much energy must be stored in the capacitor for one flash? (b) The capacitor has a potential difference between its plates of 125 V when the stored energy equals the value stored in part (a). What is the capacitance?

Solution:

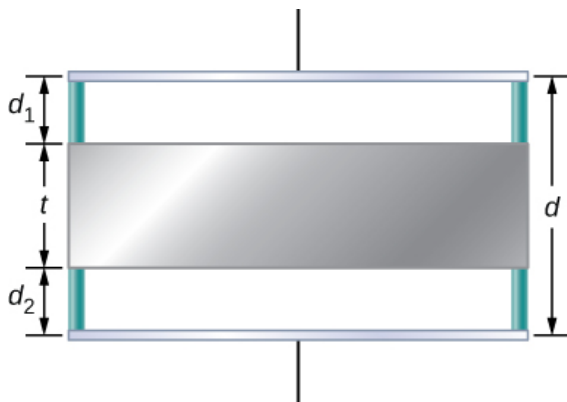
a. 421 J; b. 53.9 mF

Exercise:**Problem:**

A spherical capacitor is formed from two concentric spherical conducting shells separated by a vacuum. The inner sphere has radius 12.5 cm and the outer sphere has radius 14.8 cm. A potential difference of 120 V is applied to the capacitor. (a) What is the energy density at $r = 12.6$ cm, just outside the inner sphere? (b) What is the energy density at $r = 14.7$ cm, just inside the outer sphere? (c) For the parallel-plate capacitor the energy density is uniform in the region between the plates, except near the edges of the plates. Is this also true for the spherical capacitor?

Exercise:**Problem:**

A metal plate of thickness t is held in place between two capacitor plates by plastic pegs, as shown below. The effect of the pegs on the capacitance is negligible. The area of each capacitor plate and the area of the top and bottom surfaces of the inserted plate are all A . What is the capacitance of this system?



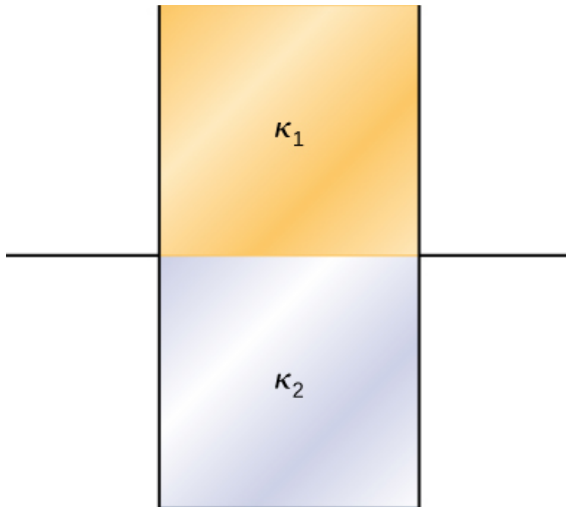
Solution:

$$C = \epsilon_0 A / (d_1 + d_2)$$

Exercise:**Problem:**

A parallel-plate capacitor is filled with two dielectrics, as shown below. When the plate area is A and separation between plates is d , show that the capacitance is given by

$$C = \epsilon_0 \frac{A}{d} \frac{\kappa_1 + \kappa_2}{2}.$$

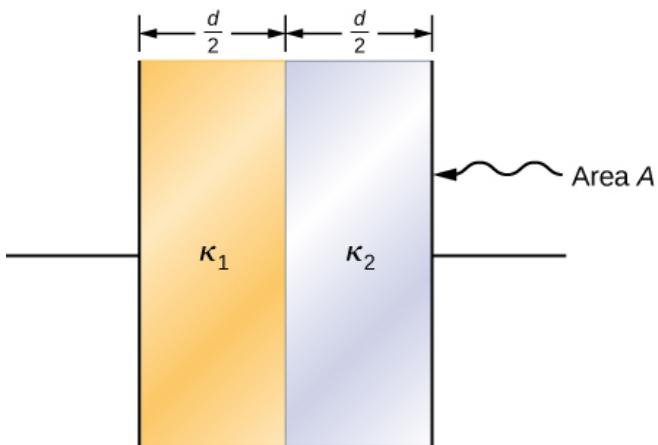


Exercise:

Problem:

A parallel-plate capacitor is filled with two dielectrics, as shown below. Show that the capacitance is given by

$$C = 2\epsilon_0 \frac{A}{d} \frac{\kappa_1 \kappa_2}{\kappa_1 + \kappa_2}.$$



Solution:

proof

Exercise:

Problem:

A capacitor has parallel plates of area 12 cm^2 separated by 2.0 mm . The space between the plates is filled with polystyrene. (a) Find the maximum permissible voltage across the capacitor to avoid dielectric breakdown. (b) When the voltage equals the value found in part (a), find the surface charge density on the surface of the dielectric.

Glossary

dielectric breakdown

phenomenon that occurs when an insulator becomes a conductor in a strong electrical field

dielectric strength

critical electrical field strength above which molecules in insulator begin to break down and the insulator starts to conduct

induced electric-dipole moment

dipole moment that a nonpolar molecule may acquire when it is placed in an electrical field

induced electrical field

electrical field in the dielectric due to the presence of induced charges

induced surface charges

charges that occur on a dielectric surface due to its polarization

Introduction

class="introduction"

Magnetic resonance imaging (MRI) uses superconducting magnets and produces high-resolution images without the danger of radiation. The image on the left shows the spacing of vertebrae along a human spinal column, with the circle indicating where the vertebrae are too close due to a ruptured disc. On the right is a picture of the MRI instrument, which surrounds the patient on all sides. A large amount of electrical current is

required to
operate the
electromagnets
(credit right:
modification of
work by “digital
cat”/Flickr).



In this chapter, we study the electrical current through a material, where the electrical current is the rate of flow of charge. We also examine a characteristic of materials known as the resistance. Resistance is a measure of how much a material impedes the flow of charge, and it will be shown that the resistance depends on temperature. In general, a good conductor, such as copper, gold, or silver, has very low resistance. Some materials, called superconductors, have zero resistance at very low temperatures.

High currents are required for the operation of electromagnets. Superconductors can be used to make electromagnets that are 10 times stronger than the strongest conventional electromagnets. These superconducting magnets are used in the construction of magnetic resonance imaging (MRI) devices that can be used to make high-resolution images of the human body. The chapter-opening picture shows an MRI image of the vertebrae of a human subject and the MRI device itself.

Superconducting magnets have many other uses. For example, superconducting magnets are used in the Large Hadron Collider (LHC) to curve the path of protons in the ring.

Electrical Current

By the end of this section, you will be able to:

- Describe an electrical current
- Define the unit of electrical current
- Explain the direction of current flow

Up to now, we have considered primarily static charges. When charges did move, they were accelerated in response to an electrical field created by a voltage difference. The charges lost potential energy and gained kinetic energy as they traveled through a potential difference where the electrical field did work on the charge.

Although charges do not require a material to flow through, the majority of this chapter deals with understanding the movement of charges through a material. The rate at which the charges flow past a location—that is, the amount of charge per unit time—is known as the *electrical current*. When charges flow through a medium, the current depends on the voltage applied, the material through which the charges flow, and the state of the material. Of particular interest is the motion of charges in a conducting wire. In previous chapters, charges were accelerated due to the force provided by an electrical field, losing potential energy and gaining kinetic energy. In this chapter, we discuss the situation of the force provided by an electrical field in a conductor, where charges lose kinetic energy to the material reaching a constant velocity, known as the “*drift velocity*.” This is analogous to an object falling through the atmosphere and losing kinetic energy to the air, reaching a constant terminal velocity.

If you have ever taken a course in first aid or safety, you may have heard that in the event of electric shock, it is the current, not the voltage, which is the important factor on the severity of the shock and the amount of damage to the human body. Current is measured in units called amperes; you may have noticed that circuit breakers in your home and fuses in your car are rated in amps (or amperes). But what is the ampere and what does it measure?

Defining Current and the Ampere

Electrical current is defined to be the rate at which charge flows. When there is a large current present, such as that used to run a refrigerator, a large amount of charge moves through the wire in a small amount of time. If the current is small, such as that used to operate a handheld calculator, a small amount of charge moves through the circuit over a long period of time.

Note:

Electrical Current

The average electrical current I is the rate at which charge flows,

Equation:

$$I_{\text{ave}} = \frac{\Delta Q}{\Delta t},$$

where ΔQ is the amount of net charge passing through a given cross-sectional area in time Δt ([\[link\]](#)). The SI unit for current is the **ampere** (A), named for the French physicist André-Marie Ampère (1775–1836). Since $I = \frac{\Delta Q}{\Delta t}$, we see that an ampere is defined as one coulomb of charge passing through a given area per second:

Equation:

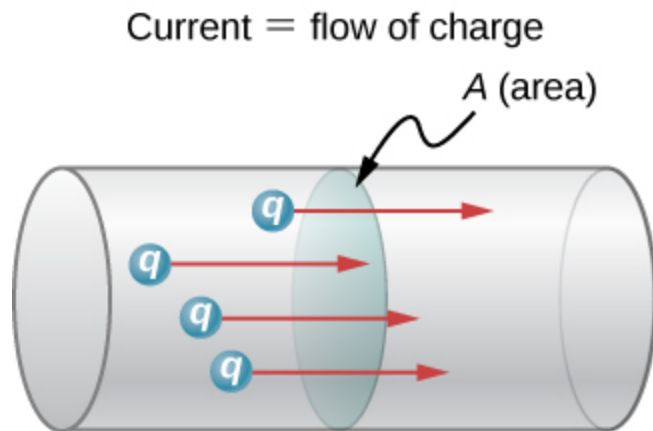
$$1\text{A} \equiv 1 \frac{\text{C}}{\text{s}}.$$

The instantaneous electrical current, or simply the **electrical current**, is the time derivative of the charge that flows and is found by taking the limit of the average electrical current as $\Delta t \rightarrow 0$:

Equation:

$$I = \lim_{\Delta t \rightarrow 0} \frac{\Delta Q}{\Delta t} = \frac{dQ}{dt}.$$

Most electrical appliances are rated in amperes (or amps) required for proper operation, as are fuses and circuit breakers.



The rate of flow of charge is current. An ampere is the flow of one coulomb of charge through an area in one second. A current of one amp would result from 6.25×10^{18} electrons flowing through the area A each second.

Example:

Calculating the Average Current

The main purpose of a battery in a car or truck is to run the electric starter motor, which starts the engine. The operation of starting the vehicle requires a large current to be supplied by the battery. Once the engine starts, a device called an alternator takes over supplying the electric power required for running the vehicle and for charging the battery.

(a) What is the average current involved when a truck battery sets in motion 720 C of charge in 4.00 s while starting an engine? (b) How long does it take 1.00 C of charge to flow from the battery?

Strategy

We can use the definition of the average current in the equation $I = \frac{\Delta Q}{\Delta t}$ to find the average current in part (a), since charge and time are given. For part (b), once we know the average current, we can use its definition $I = \frac{\Delta Q}{\Delta t}$ to find the time required for 1.00 C of charge to flow from the battery.

Solution

a. Entering the given values for charge and time into the definition of current gives

Equation:

$$I = \frac{\Delta Q}{\Delta t} = \frac{720 \text{ C}}{4.00 \text{ s}} = 180 \text{ C/s} = 180 \text{ A}.$$

b. Solving the relationship $I = \frac{\Delta Q}{\Delta t}$ for time Δt and entering the known values for charge and current gives

Equation:

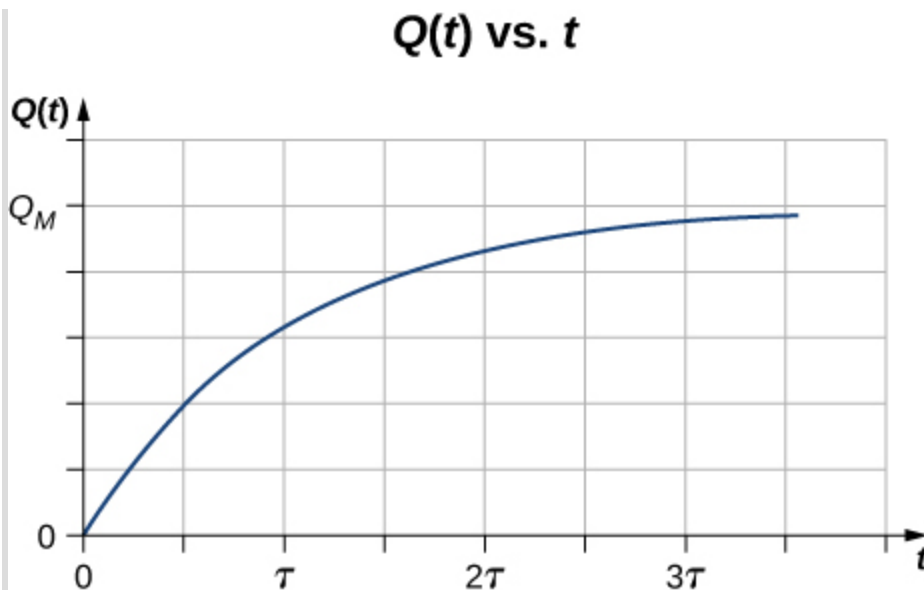
$$\Delta t = \frac{\Delta Q}{I} = \frac{1.00 \text{ C}}{180 \text{ C/s}} = 5.56 \times 10^{-3} \text{ s} = 5.56 \text{ ms}.$$

Significance

a. This large value for current illustrates the fact that a large charge is moved in a small amount of time. The currents in these “starter motors” are fairly large to overcome the inertia of the engine. b. A high current requires a short time to supply a large amount of charge. This large current is needed to supply the large amount of energy needed to start the engine.

Example:**Calculating Instantaneous Currents**

Consider a charge moving through a cross-section of a wire where the charge is modeled as $Q(t) = Q_M (1 - e^{-t/\tau})$. Here, Q_M is the charge after a long period of time, as time approaches infinity, with units of coulombs, and τ is a time constant with units of seconds (see [\[link\]](#)). What is the current through the wire?



A graph of the charge moving through a cross-section of a wire over time.

Strategy

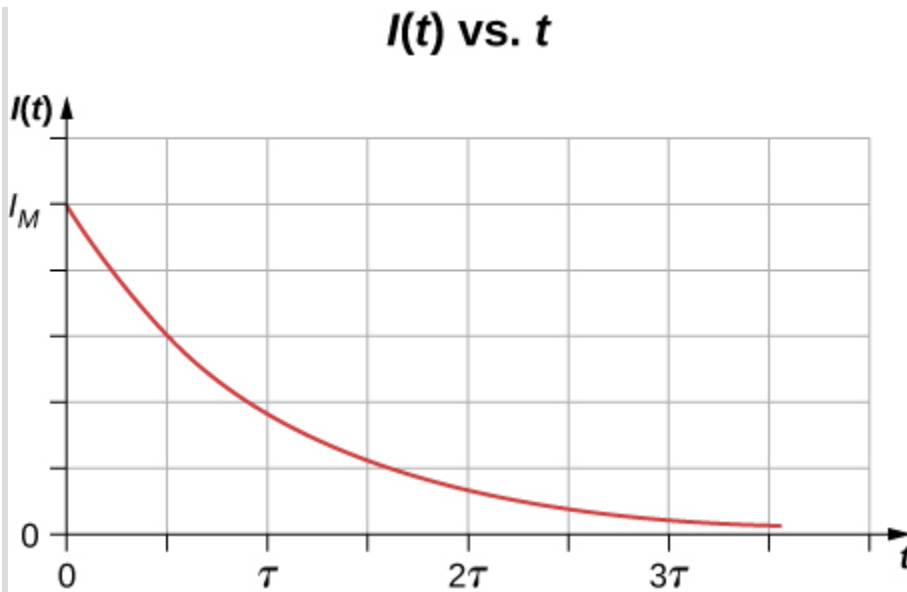
The current through the cross-section can be found from $I = \frac{dQ}{dt}$. Notice from the figure that the charge increases to Q_M and the derivative decreases, approaching zero, as time increases ([\[link\]](#)).

Solution

The derivative can be found using $\frac{d}{dx} e^u = e^u \frac{du}{dx}$.

Equation:

$$I = \frac{dQ}{dt} = \frac{d}{dt} \left[Q_M \left(1 - e^{-t/\tau} \right) \right] = \frac{Q_M}{\tau} e^{-t/\tau}.$$



A graph of the current flowing through the wire over time.

Significance

The current through the wire in question decreases exponentially, as shown in [\[link\]](#). In later chapters, it will be shown that a time-dependent current appears when a capacitor charges or discharges through a resistor. Recall that a capacitor is a device that stores charge. You will learn about the resistor in [Model of Conduction in Metals](#).

Note:

Exercise:

Problem:

Check Your Understanding Handheld calculators often use small solar cells to supply the energy required to complete the calculations needed to complete your next physics exam. The current needed to run your calculator can be as small as 0.30 mA. How long would it take for 1.00 C of charge to flow from the solar cells? Can solar cells be used, instead of batteries, to start traditional internal combustion engines presently used in most cars and trucks?

Solution:

The time for 1.00 C of charge to flow would be

$$\Delta t = \frac{\Delta Q}{I} = \frac{1.00 \text{ C}}{0.300 \times 10^{-3} \text{ C/s}} = 3.33 \times 10^3 \text{ s, slightly less than an hour.}$$
 This is quite different from the 5.55 ms for the truck battery. The calculator takes a very small amount of energy to operate, unlike the truck's starter motor. There are several reasons that vehicles use batteries and not solar cells. Aside from the obvious fact that a light source to run the solar cells for a car or truck is not always available, the large amount of current needed to start the engine cannot easily be supplied by present-day solar cells. Solar cells can possibly be used to charge the batteries. Charging the battery requires a small amount of energy when compared to the energy required to run the engine and the other accessories such as the heater and air conditioner. Present day solar-powered cars are powered by solar panels, which may power an electric motor, instead of an internal combustion engine.

Note:**Exercise:**

Problem:

Check Your Understanding Circuit breakers in a home are rated in amperes, normally in a range from 10 amps to 30 amps, and are used to protect the residents from harm and their appliances from damage due to large currents. A single 15-amp circuit breaker may be used to protect several outlets in the living room, whereas a single 20-amp circuit breaker may be used to protect the refrigerator in the kitchen. What can you deduce from this about current used by the various appliances?

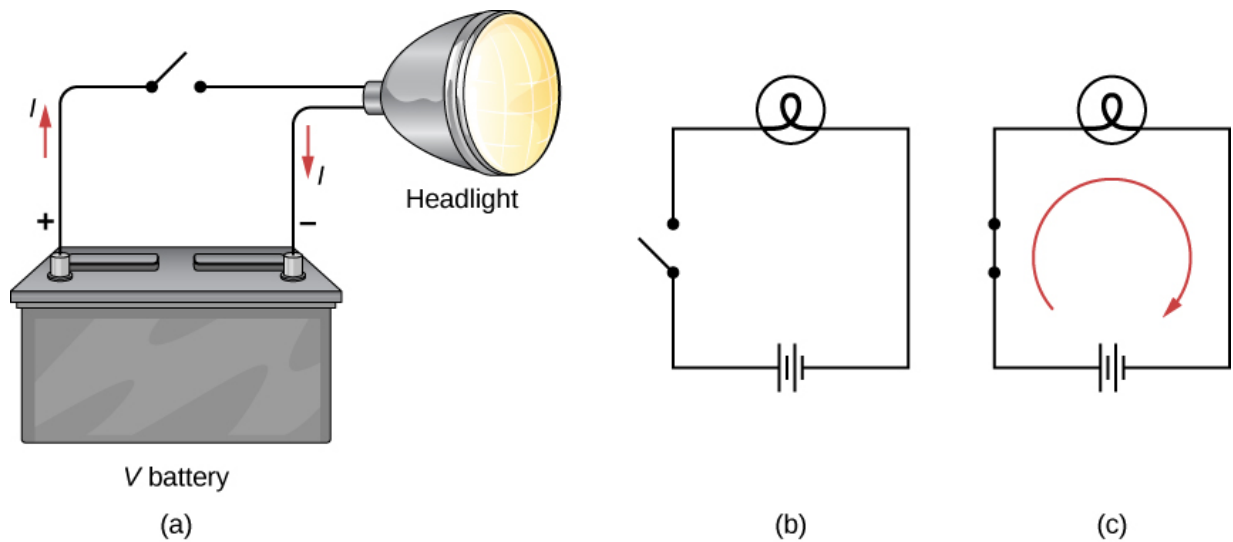
Solution:

The total current needed by all the appliances in the living room (a few lamps, a television, and your laptop) draw less current and require less power than the refrigerator.

Current in a Circuit

In the previous paragraphs, we defined the current as the charge that flows through a cross-sectional area per unit time. In order for charge to flow through an appliance, such as the headlight shown in [\[link\]](#), there must be a complete path (or **circuit**) from the positive terminal to the negative terminal. Consider a simple circuit of a car battery, a switch, a headlight lamp, and wires that provide a current path between the components. In order for the lamp to light, there must be a complete path for current flow. In other words, a charge must be able to leave the positive terminal of the battery, travel through the component, and back to the negative terminal of the battery. The switch is there to control the circuit. Part (a) of the figure shows the simple circuit of a car battery, a switch, a conducting path, and a headlight lamp. Also shown is the **schematic** of the circuit [part (b)]. A schematic is a graphical representation of a circuit and is very useful in visualizing the main features of a circuit. Schematics use standardized symbols to represent the components in a circuits and solid lines to represent the wires connecting the components. The battery is shown as a

series of long and short lines, representing the historic voltaic pile. The lamp is shown as a circle with a loop inside, representing the filament of an incandescent bulb. The switch is shown as two points with a conducting bar to connect the two points and the wires connecting the components are shown as solid lines. The schematic in part (c) shows the direction of current flow when the switch is closed.

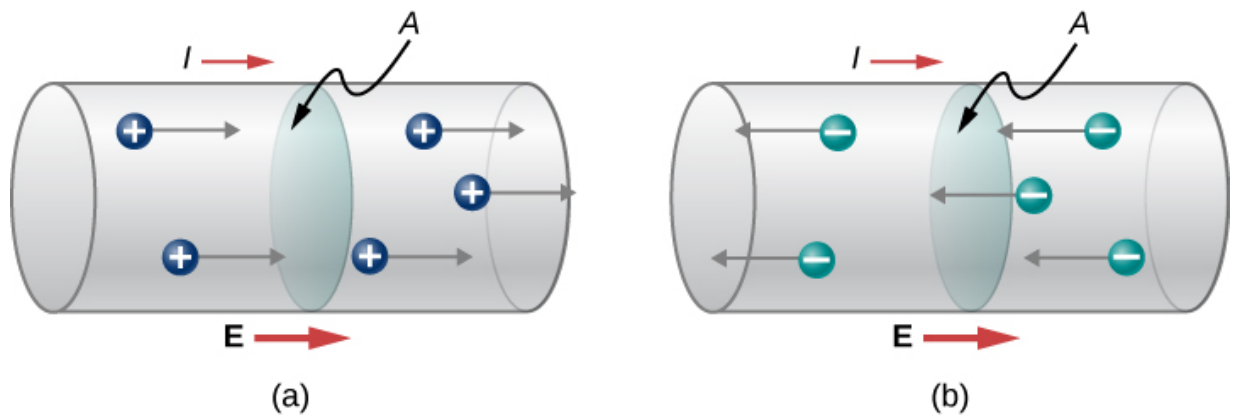


(a) A simple electric circuit of a headlight (lamp), a battery, and a switch. When the switch is closed, an uninterrupted path for current to flow through is supplied by conducting wires connecting a load to the terminals of a battery. (b) In this schematic, the battery is represented by parallel lines, which resemble plates in the original design of a battery. The longer lines indicate the positive terminal. The conducting wires are shown as solid lines. The switch is shown, in the open position, as two terminals with a line representing a conducting bar that can make contact between the two terminals. The lamp is represented by a circle encompassing a filament, as would be seen in an incandescent light bulb. (c) When the switch is closed, the circuit is complete and current flows from the positive terminal to the negative terminal of the battery.

When the switch is closed in [\[link\]](#)(c), there is a complete path for charges to flow, from the positive terminal of the battery, through the switch, then through the headlight and back to the negative terminal of the battery. Note that the direction of current flow is from positive to negative. The direction of **conventional current** is always represented in the direction that positive charge would flow, from the positive terminal to the negative terminal.

The conventional current flows from the positive terminal to the negative terminal, but depending on the actual situation, positive charges, negative charges, or both may move. In metal wires, for example, current is carried by electrons—that is, negative charges move. In ionic solutions, such as salt water, both positive and negative charges move. This is also true in nerve cells. A Van de Graaff generator, used for nuclear research, can produce a current of pure positive charges, such as protons. In the Tevatron Accelerator at Fermilab, before it was shut down in 2011, beams of protons and antiprotons traveling in opposite directions were collided. The protons are positive and therefore their current is in the same direction as they travel. The antiprotons are negatively charged and thus their current is in the opposite direction that the actual particles travel.

A closer look at the current flowing through a wire is shown in [\[link\]](#). The figure illustrates the movement of charged particles that compose a current. The fact that conventional current is taken to be in the direction that positive charge would flow can be traced back to American scientist and statesman Benjamin Franklin in the 1700s. Having no knowledge of the particles that make up the atom (namely the proton, electron, and neutron), Franklin believed that electrical current flowed from a material that had more of an “electrical fluid” and to a material that had less of this “electrical fluid.” He coined the term *positive* for the material that had more of this electrical fluid and *negative* for the material that lacked the electrical fluid. He surmised that current would flow from the material with more electrical fluid—the positive material—to the negative material, which has less electrical fluid. Franklin called this direction of current a positive current flow. This was pretty advanced thinking for a man who knew nothing about the atom.



Current I is the rate at which charge moves through an area A , such as the cross-section of a wire. Conventional current is defined to move in the direction of the electrical field. (a) Positive charges move in the direction of the electrical field, which is the same direction as conventional current. (b) Negative charges move in the direction opposite to the electrical field. Conventional current is in the direction opposite to the movement of negative charge. The flow of electrons is sometimes referred to as electronic flow.

We now know that a material is positive if it has a greater number of protons than electrons, and it is negative if it has a greater number of electrons than protons. In a conducting metal, the current flow is due primarily to electrons flowing from the negative material to the positive material, but for historical reasons, we consider the positive current flow and the current is shown to flow from the positive terminal of the battery to the negative terminal.

It is important to realize that an electrical field is present in conductors and is responsible for producing the current ([link](#)). In previous chapters, we considered the static electrical case, where charges in a conductor quickly redistribute themselves on the surface of the conductor in order to cancel out the external electrical field and restore equilibrium. In the case of an electrical circuit, the charges are prevented from ever reaching equilibrium by an external source of electric potential, such as a battery. The energy

needed to move the charge is supplied by the electric potential from the battery.

Although the electrical field is responsible for the motion of the charges in the conductor, the work done on the charges by the electrical field does not increase the kinetic energy of the charges. We will show that the electrical field is responsible for keeping the electric charges moving at a “drift velocity.”

Summary

- The average electrical current I_{ave} is the rate at which charge flows, given by $I_{\text{ave}} = \frac{\Delta Q}{\Delta t}$, where ΔQ is the amount of charge passing through an area in time Δt .
- The instantaneous electrical current, or simply the current I , is the rate at which charge flows. Taking the limit as the change in time approaches zero, we have $I = \frac{dQ}{dt}$, where $\frac{dQ}{dt}$ is the time derivative of the charge.
- The direction of conventional current is taken as the direction in which positive charge moves. In a simple direct-current (DC) circuit, this will be from the positive terminal of the battery to the negative terminal.
- The SI unit for current is the ampere, or simply the amp (A), where $1 \text{ A} = 1 \text{ C/s}$.
- Current consists of the flow of free charges, such as electrons, protons, and ions.

Conceptual Questions

Exercise:

Problem:

Can a wire carry a current and still be neutral—that is, have a total charge of zero? Explain.

Solution:

If a wire is carrying a current, charges enter the wire from the voltage source's positive terminal and leave at the negative terminal, so the total charge remains zero while the current flows through it.

Exercise:

Problem:

Car batteries are rated in ampere-hours ($A \cdot h$). To what physical quantity do ampere-hours correspond (voltage, current, charge, energy, power,...)?

Exercise:

Problem:

When working with high-power electric circuits, it is advised that whenever possible, you work “one-handed” or “keep one hand in your pocket.” Why is this a sensible suggestion?

Solution:

Using one hand will reduce the possibility of “completing the circuit” and having current run through your body, especially current running through your heart.

Problems

Exercise:

Problem:

A Van de Graaff generator is one of the original particle accelerators and can be used to accelerate charged particles like protons or electrons. You may have seen it used to make human hair stand on end or produce large sparks. One application of the Van de Graaff generator is to create X-rays by bombarding a hard metal target with the beam. Consider a beam of protons at 1.00 keV and a current of 5.00 mA produced by the generator. (a) What is the speed of the protons? (b) How many protons are produced each second?

Solution:

a. $v = 4.38 \times 10^5 \frac{\text{m}}{\text{s}};$

b. $\Delta q = 5.00 \times 10^{-3} \text{C},$ no. of protons $= 3.13 \times 10^{16}$

Exercise:**Problem:**

A cathode ray tube (CRT) is a device that produces a focused beam of electrons in a vacuum. The electrons strike a phosphor-coated glass screen at the end of the tube, which produces a bright spot of light. The position of the bright spot of light on the screen can be adjusted by deflecting the electrons with electrical fields, magnetic fields, or both. Although the CRT tube was once commonly found in televisions, computer displays, and oscilloscopes, newer appliances use a liquid crystal display (LCD) or plasma screen. You still may come across a CRT in your study of science. Consider a CRT with an electron beam average current of $25.00 \mu\text{A}$. How many electrons strike the screen every minute?

Exercise:**Problem:**

How many electrons flow through a point in a wire in 3.00 s if there is a constant current of $I = 4.00 \text{ A}$?

Solution:

$$I = \frac{\Delta Q}{\Delta t}, \quad \Delta Q = 12.00 \text{ C}$$

$$\text{no. of electrons} = 7.5 \times 10^{19}$$

Exercise:

Problem:

A conductor carries a current that is decreasing exponentially with time. The current is modeled as $I = I_0 e^{-t/\tau}$, where $I_0 = 3.00$ A is the current at time $t = 0.00$ s and $\tau = 0.50$ s is the time constant. How much charge flows through the conductor between $t = 0.00$ s and $t = 3\tau$?

Exercise:**Problem:**

The quantity of charge through a conductor is modeled as $Q = 4.00 \frac{\text{C}}{\text{s}^4} t^4 - 1.00 \frac{\text{C}}{\text{s}} t + 6.00$ mC.

What is the current at time $t = 3.00$ s?

Solution:

$$I(t) = 0.016 \frac{\text{C}}{\text{s}^4} t^3 - 0.001 \frac{\text{C}}{\text{s}}$$
$$I(3.00 \text{ s}) = 0.431 \text{ A}$$

Exercise:**Problem:**

The current through a conductor is modeled as $I(t) = I_m \sin(2\pi [60 \text{ Hz}]t)$. Write an equation for the charge as a function of time.

Exercise:**Problem:**

The charge on a capacitor in a circuit is modeled as $Q(t) = Q_{\text{max}} \cos(\omega t + \phi)$. What is the current through the circuit as a function of time?

Solution:

$$I(t) = -I_{\max} \sin(\omega t + \phi)$$

Glossary

ampere (amp)

SI unit for current; $1 \text{ A} = 1 \text{ C/s}$

circuit

complete path that an electrical current travels along

conventional current

current that flows through a circuit from the positive terminal of a battery through the circuit to the negative terminal of the battery

electrical current

rate at which charge flows, $I = \frac{dQ}{dt}$

schematic

graphical representation of a circuit using standardized symbols for components and solid lines for the wire connecting the components

Model of Conduction in Metals

By the end of this section, you will be able to:

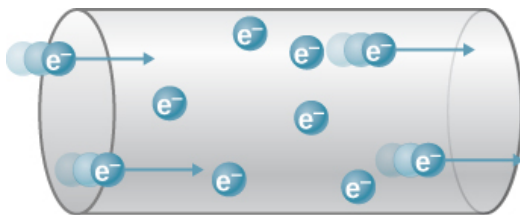
- Define the drift velocity of charges moving through a metal
- Define the vector current density
- Describe the operation of an incandescent lamp

When electrons move through a conducting wire, they do not move at a constant velocity, that is, the electrons do not move in a straight line at a constant speed. Rather, they interact with and collide with atoms and other free electrons in the conductor. Thus, the electrons move in a zig-zag fashion and drift through the wire. We should also note that even though it is convenient to discuss the direction of current, current is a scalar quantity. When discussing the velocity of charges in a current, it is more appropriate to discuss the current density. We will come back to this idea at the end of this section.

Drift Velocity

Electrical signals move very rapidly. Telephone conversations carried by currents in wires cover large distances without noticeable delays. Lights come on as soon as a light switch is moved to the 'on' position. Most electrical signals carried by currents travel at speeds on the order of 10^8 m/s, a significant fraction of the speed of light. Interestingly, the individual charges that make up the current move much slower on average, typically drifting at speeds on the order of 10^{-4} m/s. How do we reconcile these two speeds, and what does it tell us about standard conductors?

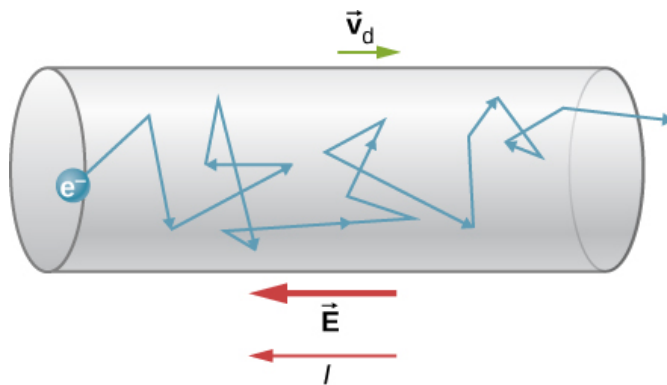
The high speed of electrical signals results from the fact that the force between charges acts rapidly at a distance. Thus, when a free charge is forced into a wire, as in [\[link\]](#), the incoming charge pushes other charges ahead of it due to the repulsive force between like charges. These moving charges push on charges farther down the line. The density of charge in a system cannot easily be increased, so the signal is passed on rapidly. The resulting electrical shock wave moves through the system at nearly the speed of light. To be precise, this fast-moving signal, or shock wave, is a rapidly propagating change in the electrical field.



When charged particles are forced into this volume of a conductor, an equal number are quickly forced to leave. The repulsion between like charges makes it difficult to increase the number of charges in a volume. Thus, as one charge enters, another

leaves almost immediately, carrying the signal rapidly forward.

Good conductors have large numbers of free charges. In metals, the free charges are free electrons. (In fact, good electrical conductors are often good heat conductors too, because large numbers of free electrons can transport thermal energy as well as carry electrical current.) [\[link\]](#) shows how free electrons move through an ordinary conductor. The distance that an individual electron can move between collisions with atoms or other electrons is quite small. The electron paths thus appear nearly random, like the motion of atoms in a gas. But there is an electrical field in the conductor that causes the electrons to drift in the direction shown (opposite to the field, since they are negative). The **drift velocity** \vec{v}_d is the average velocity of the free charges. Drift velocity is quite small, since there are so many free charges. If we have an estimate of the density of free electrons in a conductor, we can calculate the drift velocity for a given current. The larger the density, the lower the velocity required for a given current.



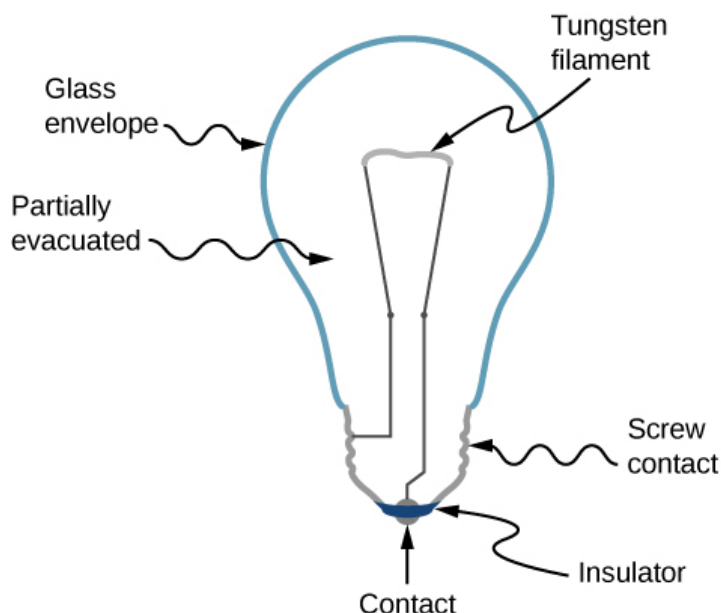
Free electrons moving in a conductor make many collisions with other electrons and other particles.

A typical path of one electron is shown. The average velocity of the free charges is called the drift velocity \vec{v}_d and for electrons, it is in the direction opposite to the electrical field. The

collisions normally transfer energy to the conductor, requiring a constant supply of energy to maintain a steady current.

Free-electron collisions transfer energy to the atoms of the conductor. The electrical field does work in moving the electrons through a distance, but that work does not increase the kinetic energy (nor speed) of the electrons. The work is transferred to the conductor's atoms, often increasing temperature. Thus, a continuous power input is required to keep a current flowing. (An exception is superconductors, for reasons we shall explore in a later chapter. Superconductors can have a steady current without a continual supply of energy—a great energy savings.) For a conductor that is not a

superconductor, the supply of energy can be useful, as in an incandescent light bulb filament ([link](#)). The supply of energy is necessary to increase the temperature of the tungsten filament, so that the filament glows.



The incandescent lamp is a simple design. A tungsten filament is placed in a partially evacuated glass envelope. One end of the filament is attached to the screw base, which is made out of a conducting material. The second end of the filament is attached to a second contact in the base of the bulb. The two contacts are separated by an insulating material. Current flows through the filament, and the temperature of the filament becomes large enough to cause the filament to glow and produce light. However, these bulbs are not very energy efficient, as evident from the heat coming from the bulb. In the year 2012, the United States, along with many other countries, began to phase out incandescent lamps in favor of more energy-efficient lamps, such as light-emitting diode (LED) lamps and compact fluorescent lamps (CFL) (credit right: modification of work by Serge Saint).

We can obtain an expression for the relationship between current and drift velocity by considering the number of free charges in a segment of wire, as illustrated in [link](#). The number of free charges per unit volume, or the number density of free charges, is given the symbol n where

$n = \frac{\text{number of charges}}{\text{volume}}$. The value of n depends on the material. The shaded segment has a volume $Av_d dt$, so that the number of free charges in the volume is $nAv_d dt$. The charge dQ in this segment is thus $qnAv_d dt$, where q is the amount of charge on each carrier. (The magnitude of the charge of electrons is $q = 1.60 \times 10^{-19}$ C.) Current is charge moved per unit time; thus, if all the original charges move out of this segment in time dt , the current is

Equation:

$$I = \frac{dQ}{dt} = qnAv_d.$$

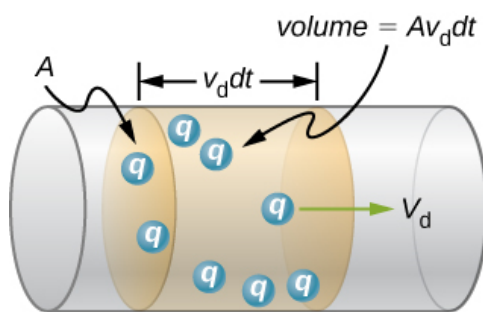
Rearranging terms gives

Note:

Equation:

$$v_d = \frac{I}{nqA}$$

where v_d is the drift velocity, n is the free charge density, A is the cross-sectional area of the wire, and I is the current through the wire. The carriers of the current each have charge q and move with a drift velocity of magnitude v_d .

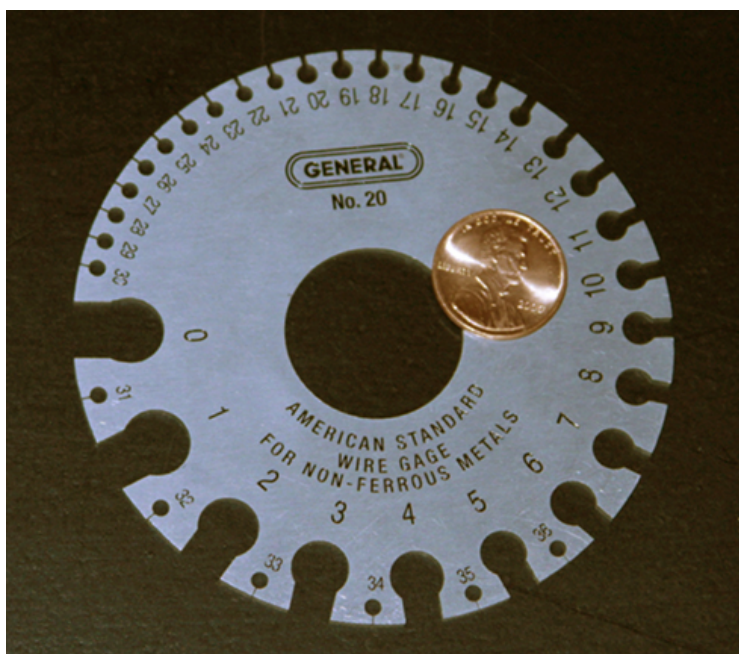


All the charges in the shaded volume of this wire move out in a time dt , having a drift velocity of magnitude v_d .

Note that simple drift velocity is not the entire story. The speed of an electron is sometimes much greater than its drift velocity. In addition, not all of the electrons in a conductor can move freely, and those that do move might move somewhat faster or slower than the drift velocity. So what do we mean by free electrons?

Atoms in a metallic conductor are packed in the form of a lattice structure. Some electrons are far enough away from the atomic nuclei that they do not experience the attraction of the nuclei as strongly as the inner electrons do. These are the free electrons. They are not bound to a single atom but can instead move freely among the atoms in a “sea” of electrons. When an electrical field is applied, these free electrons respond by accelerating. As they move, they collide with the atoms in the lattice and with other electrons, generating thermal energy, and the conductor gets warmer. In an insulator, the organization of the atoms and the structure do not allow for such free electrons.

As you know, electric power is usually supplied to equipment and appliances through round wires made of a conducting material (copper, aluminum, silver, or gold) that are stranded or solid. The diameter of the wire determines the current-carrying capacity—the larger the diameter, the greater the current-carrying capacity. Even though the current-carrying capacity is determined by the diameter, wire is not normally characterized by the diameter directly. Instead, wire is commonly sold in a unit known as “gauge.” Wires are manufactured by passing the material through circular forms called “drawing dies.” In order to make thinner wires, manufacturers draw the wires through multiple dies of successively thinner diameter. Historically, the gauge of the wire was related to the number of drawing processes required to manufacture the wire. For this reason, the larger the gauge, the smaller the diameter. In the United States, the American Wire Gauge (AWG) was developed to standardize the system. Household wiring commonly consists of 10-gauge (2.588-mm diameter) to 14-gauge (1.628-mm diameter) wire. A device used to measure the gauge of wire is shown in [\[link\]](#).



A device for measuring the gauge of electrical wire. As you can see, higher gauge numbers indicate thinner wires. (credit: Joseph J. Trout)

Example:

Calculating Drift Velocity in a Common Wire

Calculate the drift velocity of electrons in a copper wire with a diameter of 2.053 mm (12-gauge) carrying a 20.0-A current, given that there is one free electron per copper atom. (Household wiring often contains 12-gauge copper wire, and the maximum current allowed in such wire is usually 20.0 A.) The density of copper is $8.80 \times 10^3 \text{ kg/m}^3$ and the atomic mass of copper is 63.54 g/mol.

Strategy

We can calculate the drift velocity using the equation $I = nqAv_d$. The current is $I = 20.00 \text{ A}$ and $q = 1.60 \times 10^{-19} \text{ C}$ is the charge of an electron. We can calculate the area of a cross-section of the wire using the formula $A = \pi r^2$, where r is one-half the diameter. The given diameter is 2.053 mm, so r is 1.0265 mm. We are given the density of copper, $8.80 \times 10^3 \text{ kg/m}^3$, and the atomic mass of copper is 63.54 g/mol. We can use these two quantities along with Avogadro's number, $6.02 \times 10^{23} \text{ atoms/mol}$, to determine n , the number of free electrons per cubic meter.

Solution

First, we calculate the density of free electrons in copper. There is one free electron per copper atom. Therefore, the number of free electrons is the same as the number of copper atoms per m^3 . We can now find n as follows:

Equation:

$$\begin{aligned} n &= \frac{1 e^-}{\text{atom}} \times \frac{6.02 \times 10^{23} \text{ atoms}}{\text{mol}} \times \frac{1 \text{ mol}}{63.54 \text{ g}} \times \frac{1000 \text{ g}}{\text{kg}} \times \frac{8.80 \times 10^3 \text{ kg}}{1 \text{ m}^3} \\ &= 8.34 \times 10^{28} e^-/\text{m}^3. \end{aligned}$$

The cross-sectional area of the wire is

Equation:

$$A = \pi r^2 = \pi \left(\frac{2.05 \times 10^{-3} \text{ m}}{2} \right)^2 = 3.30 \times 10^{-6} \text{ m}^2.$$

Rearranging $I = nqAv_d$ to isolate drift velocity gives

Equation:

$$v_d = \frac{I}{nqA} = \frac{20.00 \text{ A}}{(8.34 \times 10^{28} / \text{m}^3)(-1.60 \times 10^{-19} \text{ C})(3.30 \times 10^{-6} \text{ m}^2)} = -4.54 \times 10^{-4} \text{ m/s}.$$

Significance

The minus sign indicates that the negative charges are moving in the direction opposite to conventional current. The small value for drift velocity (on the order of 10^{-4} m/s) confirms that the signal moves on the order of 10^{12} times faster (about 10^8 m/s) than the charges that carry it.

Note:

Exercise:

Problem:

Check Your Understanding In [\[link\]](#), the drift velocity was calculated for a 2.053-mm diameter (12-gauge) copper wire carrying a 20-amp current. Would the drift velocity change for a 1.628-mm diameter (14-gauge) wire carrying the same 20-amp current?

Solution:

The diameter of the 14-gauge wire is smaller than the diameter of the 12-gauge wire. Since the drift velocity is inversely proportional to the cross-sectional area, the drift velocity in the 14-gauge wire is larger than the drift velocity in the 12-gauge wire carrying the same current. The number of electrons per cubic meter will remain constant.

Current Density

Although it is often convenient to attach a negative or positive sign to indicate the overall direction of motion of the charges, current is a scalar quantity, $I = \frac{dQ}{dt}$. It is often necessary to discuss the details of the motion of the charge, instead of discussing the overall motion of the charges. In such cases, it is necessary to discuss the current density, \vec{J} , a vector quantity. The **current density** is the flow of charge through an infinitesimal area, divided by the area. The current density must take into account the local magnitude and direction of the charge flow, which varies from point to point. The unit of current density is ampere per meter squared, and the direction is defined as the direction of net flow of positive charges through the area.

The relationship between the current and the current density can be seen in [\[link\]](#). The differential current flow through the area $d\vec{A}$ is found as

Equation:

$$dI = \vec{J} \cdot d\vec{A} = JdA \cos \theta,$$

where θ is the angle between the area and the current density. The total current passing through area $d\vec{A}$ can be found by integrating over the area,

Note:

Equation:

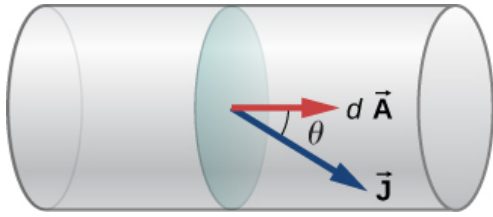
$$I = \iint_{\text{area}} \vec{J} \cdot d\vec{A}.$$

Consider the magnitude of the current density, which is the current divided by the area:

Equation:

$$J = \frac{I}{A} = \frac{n|q|Av_d}{A} = n|q|v_d.$$

Thus, the current density is $\vec{J} = nq\vec{v}_d$. If q is positive, \vec{v}_d is in the same direction as the electrical field \vec{E} . If q is negative, \vec{v}_d is in the opposite direction of \vec{E} . Either way, the direction of the current density \vec{J} is in the direction of the electrical field \vec{E} .



The current density \vec{J} is defined as the current passing through an infinitesimal cross-sectional area divided by the area. The direction of the current density is the direction of the net flow of positive charges and the magnitude is equal to the current divided by the infinitesimal area.

Example:

Calculating the Current Density in a Wire

The current supplied to a lamp with a 100-W light bulb is 0.87 amps. The lamp is wired using a copper wire with diameter 2.588 mm (10-gauge). Find the magnitude of the current density.

Strategy

The current density is the current moving through an infinitesimal cross-sectional area divided by the area. We can calculate the magnitude of the current density using $J = \frac{I}{A}$. The current is given as 0.87 A. The cross-sectional area can be calculated to be $A = 5.26 \text{ mm}^2$.

Solution

Calculate the current density using the given current $I = 0.87 \text{ A}$ and the area, found to be $A = 5.26 \text{ mm}^2$.

Equation:

$$J = \frac{I}{A} = \frac{0.87 \text{ A}}{5.26 \times 10^{-6} \text{ m}^2} = 1.65 \times 10^5 \frac{\text{A}}{\text{m}^2}.$$

Significance

The current density in a conducting wire depends on the current through the conducting wire and the cross-sectional area of the wire. For a given current, as the diameter of the wire increases, the charge density decreases.

Note:

Exercise:

Problem:

Check Your Understanding The current density is proportional to the current and inversely proportional to the area. If the current density in a conducting wire increases, what would happen to the drift velocity of the charges in the wire?

Solution:

The current density in a conducting wire increases due to an increase in current. The drift velocity is inversely proportional to the current ($v_d = \frac{I}{nqA}$), so the drift velocity would decrease.

What is the significance of the current density? The current density is proportional to the current, and the current is the number of charges that pass through a cross-sectional area per second. The charges move through the conductor, accelerated by the electric force provided by the electrical field. The electrical field is created when a voltage is applied across the conductor. In [Ohm's Law](#), we will use this relationship between the current density and the electrical field to examine the relationship between the current through a conductor and the voltage applied.

Summary

- The current through a conductor depends mainly on the motion of free electrons.
- When an electrical field is applied to a conductor, the free electrons in a conductor do not move through a conductor at a constant speed and direction; instead, the motion is almost random due to collisions with atoms and other free electrons.
- Even though the electrons move in a nearly random fashion, when an electrical field is applied to the conductor, the overall velocity of the electrons can be defined in terms of a drift velocity.
- The current density is a vector quantity defined as the current through an infinitesimal area divided by the area.
- The current can be found from the current density, $I = \iint_{\text{area}} \vec{J} \cdot d\vec{A}$.
- An incandescent light bulb is a filament of wire enclosed in a glass bulb that is partially evacuated. Current runs through the filament, where the electrical energy is converted to light and heat.

Conceptual Questions**Exercise:****Problem:**

Incandescent light bulbs are being replaced with more efficient LED and CFL light bulbs. Is there any obvious evidence that incandescent light bulbs might not be that energy efficient? Is energy converted into anything but visible light?

Exercise:

Problem:

It was stated that the motion of an electron appears nearly random when an electrical field is applied to the conductor. What makes the motion nearly random and differentiates it from the random motion of molecules in a gas?

Solution:

Even though the electrons collide with atoms and other electrons in the wire, they travel from the negative terminal to the positive terminal, so they drift in one direction. Gas molecules travel in completely random directions.

Exercise:**Problem:**

Electric circuits are sometimes explained using a conceptual model of water flowing through a pipe. In this conceptual model, the voltage source is represented as a pump that pumps water through pipes and the pipes connect components in the circuit. Is a conceptual model of water flowing through a pipe an adequate representation of the circuit? How are electrons and wires similar to water molecules and pipes? How are they different?

Exercise:

Problem: An incandescent light bulb is partially evacuated. Why do you suppose that is?

Solution:

In the early years of light bulbs, the bulbs are partially evacuated to reduce the amount of heat conducted through the air to the glass envelope. Dissipating the heat would cool the filament, increasing the amount of energy needed to produce light from the filament. It also protects the glass from the heat produced from the hot filament. If the glass heats, it expands, and as it cools, it contracts. This expansion and contraction could cause the glass to become brittle and crack, reducing the life of the bulbs. Many bulbs are now partially filled with an inert gas. It is also useful to remove the oxygen to reduce the possibility of the filament actually burning. When the original filaments were replaced with more efficient tungsten filaments, atoms from the tungsten would evaporate off the filament at such high temperatures. The atoms collide with the atoms of the inert gas and land back on the filament.

Problems**Exercise:**

Problem:

An aluminum wire 1.628 mm in diameter (14-gauge) carries a current of 3.00 amps. (a) What is the absolute value of the charge density in the wire? (b) What is the drift velocity of the electrons? (c) What would be the drift velocity if the same gauge copper were used instead of aluminum? The density of copper is 8.96 g/cm^3 and the density of aluminum is 2.70 g/cm^3 . The molar mass of aluminum is 26.98 g/mol and the molar mass of copper is 63.5 g/mol . Assume each atom of metal contributes one free electron.

Exercise:**Problem:**

The current of an electron beam has a measured current of $I = 50.00 \mu\text{A}$ with a radius of 1.00 mm. What is the magnitude of the current density of the beam?

Solution:

$$|J| = 15.92 \text{ A/m}^2$$

Exercise:**Problem:**

A high-energy proton accelerator produces a proton beam with a radius of $r = 0.90 \text{ mm}$. The beam current is $I = 9.00 \mu\text{A}$ and is constant. The charge density of the beam is $n = 6.00 \times 10^{11}$ protons per cubic meter. (a) What is the current density of the beam? (b) What is the drift velocity of the beam? (c) How much time does it take for 1.00×10^{10} protons to be emitted by the accelerator?

Exercise:**Problem:**

Consider a wire of a circular cross-section with a radius of $R = 3.00 \text{ mm}$. The magnitude of the current density is modeled as $J = cr^2 = 5.00 \times 10^6 \frac{\text{A}}{\text{m}^4} r^2$. What is the current through the inner section of the wire from the center to $r = 0.5R$?

Solution:

$$I = 3.98 \times 10^{-5} \text{ A}$$

Exercise:**Problem:**

A cylindrical wire has a current density from the center of the wire's cross section as $J = Cr^2$ where r is in meters, J is in amps per square meter, and $C = 10^3 \text{ A/m}^4$. This current density continues to the end of the wire at a radius of 1.0 mm. Calculate the current just outside of this wire.

Exercise:

Problem:

The current supplied to an air conditioner unit is 4.00 amps. The air conditioner is wired using a 10-gauge (diameter 2.588 mm) wire. The charge density is $n = 8.48 \times 10^{28} \frac{\text{electrons}}{\text{m}^3}$. Find the magnitude of (a) current density and (b) the drift velocity.

Solution:

a. $|J| = 7.60 \times 10^5 \frac{\text{A}}{\text{m}^2}$; b. $v_d = 5.60 \times 10^{-5} \frac{\text{m}}{\text{s}}$

Glossary

current density

flow of charge through a cross-sectional area divided by the area

drift velocity

velocity of a charge as it moves nearly randomly through a conductor, experiencing multiple collisions, averaged over a length of a conductor, whose magnitude is the length of conductor traveled divided by the time it takes for the charges to travel the length

Resistivity and Resistance

By the end of this section, you will be able to:

- Differentiate between resistance and resistivity
- Define the term conductivity
- Describe the electrical component known as a resistor
- State the relationship between resistance of a resistor and its length, cross-sectional area, and resistivity
- State the relationship between resistivity and temperature

What drives current? We can think of various devices—such as batteries, generators, wall outlets, and so on—that are necessary to maintain a current. All such devices create a potential difference and are referred to as voltage sources. When a voltage source is connected to a conductor, it applies a potential difference V that creates an electrical field. The electrical field, in turn, exerts force on free charges, causing current. The amount of current depends not only on the magnitude of the voltage, but also on the characteristics of the material that the current is flowing through. The material can resist the flow of the charges, and the measure of how much a material resists the flow of charges is known as the *resistivity*. This resistivity is crudely analogous to the friction between two materials that resists motion.

Resistivity

When a voltage is applied to a conductor, an electrical field \vec{E} is created, and charges in the conductor feel a force due to the electrical field. The current density \vec{J} that results depends on the electrical field and the properties of the material. This dependence can be very complex. In some materials, including metals at a given temperature, the current density is approximately proportional to the electrical field. In these cases, the current density can be modeled as

Equation:

$$\vec{J} = \sigma \vec{E},$$

where σ is the **electrical conductivity**. The electrical conductivity is analogous to thermal conductivity and is a measure of a material's ability to conduct or transmit electricity. Conductors have a higher electrical conductivity than insulators. Since the electrical conductivity is $\sigma = J/E$, the units are

Equation:

$$\sigma = \frac{[J]}{[E]} = \frac{\text{A/m}^2}{\text{V/m}} = \frac{\text{A}}{\text{V} \cdot \text{m}}.$$

Here, we define a unit named the **ohm** with the Greek symbol uppercase omega, Ω . The unit is named after Georg Simon Ohm, whom we will discuss later in this chapter. The Ω is used to avoid confusion with the number 0. One ohm equals one volt per amp: $1 \Omega = 1 \text{ V/A}$. The units of electrical conductivity are therefore $(\Omega \cdot \text{m})^{-1}$.

Conductivity is an intrinsic property of a material. Another intrinsic property of a material is the **resistivity**, or electrical resistivity. The resistivity of a material is a measure of how strongly a material opposes the flow of electrical current. The symbol for resistivity is the lowercase Greek letter rho, ρ , and resistivity is the reciprocal of electrical conductivity:

Equation:

$$\rho = \frac{1}{\sigma}.$$

The unit of resistivity in SI units is the ohm-meter ($\Omega \cdot \text{m}$). We can define the resistivity in terms of the electrical field and the current density,

Note:

Equation:

$$\rho = \frac{E}{J}.$$

The greater the resistivity, the larger the field needed to produce a given current density. The lower the resistivity, the larger the current density produced by a given electrical field. Good conductors have a high conductivity and low resistivity. Good insulators have a low conductivity and a high resistivity. [\[link\]](#) lists resistivity and conductivity values for various materials.

Material	Conductivity, σ ($\Omega \cdot \text{m}$)⁻¹	Resistivity, ρ ($\Omega \cdot \text{m}$)	Temperature Coefficient, α ($^{\circ}\text{C}$)⁻¹
<i>Conductors</i>			
Silver	6.29×10^7	1.59×10^{-8}	0.0038
Copper	5.95×10^7	1.68×10^{-8}	0.0039
Gold	4.10×10^7	2.44×10^{-8}	0.0034
Aluminum	3.77×10^7	2.65×10^{-8}	0.0039

Material	Conductivity, σ ($\Omega \cdot \text{m}$)⁻¹	Resistivity, ρ ($\Omega \cdot \text{m}$)	Temperature Coefficient, α ($^{\circ}\text{C}$)⁻¹
Tungsten	1.79×10^7	5.60×10^{-8}	0.0045
Iron	1.03×10^7	9.71×10^{-8}	0.0065
Platinum	0.94×10^7	10.60×10^{-8}	0.0039
Steel	0.50×10^7	20.00×10^{-8}	
Lead	0.45×10^7	22.00×10^{-8}	
Manganin (Cu, Mn, Ni alloy)	0.21×10^7	48.20×10^{-8}	0.000002
Constantan (Cu, Ni alloy)	0.20×10^7	49.00×10^{-8}	0.00003
Mercury	0.10×10^7	98.00×10^{-8}	0.0009
Nichrome (Ni, Fe, Cr alloy)	0.10×10^7	100.00×10^{-8}	0.0004
<i>Semiconductors</i> [1]			
Carbon (pure)	2.86×10^4	3.50×10^{-5}	-0.0005
Carbon	$(2.86 - 1.67) \times 10^{-6}$	$(3.5 - 60) \times 10^{-5}$	-0.0005
Germanium (pure)		600×10^{-3}	-0.048
Germanium		$(1 - 600) \times 10^{-3}$	-0.050
Silicon (pure)		2300	-0.075
Silicon		0.1 - 2300	-0.07
<i>Insulators</i>			
Amber	2.00×10^{-15}	5×10^{14}	
Glass	$10^{-9} - 10^{-14}$	$10^9 - 10^{14}$	

Material	Conductivity, σ $(\Omega \cdot \text{m})^{-1}$	Resistivity, ρ $(\Omega \cdot \text{m})$	Temperature Coefficient, α $(^\circ \text{C})^{-1}$
Lucite	$<10^{-13}$	$>10^{13}$	
Mica	$10^{-11} - 10^{-15}$	$10^{11} - 10^{15}$	
Quartz (fused)	1.33×10^{-18}	75×10^{16}	
Rubber (hard)	$10^{-13} - 10^{-16}$	$10^{13} - 10^{16}$	
Sulfur	10^{-15}	10^{15}	
Teflon TM	$<10^{-13}$	$>10^{13}$	
Wood	$10^{-8} - 10^{-11}$	$10^8 - 10^{11}$	

Resistivities and Conductivities of Various Materials at 20 °C[1] Values depend strongly on amounts and types of impurities.

The materials listed in the table are separated into categories of conductors, semiconductors, and insulators, based on broad groupings of resistivity. Conductors have the smallest resistivity, and insulators have the largest; semiconductors have intermediate resistivity. Conductors have varying but large, free charge densities, whereas most charges in insulators are bound to atoms and are not free to move. Semiconductors are intermediate, having far fewer free charges than conductors, but having properties that make the number of free charges depend strongly on the type and amount of impurities in the semiconductor. These unique properties of semiconductors are put to use in modern electronics, as we will explore in later chapters.

Note:
Exercise:

Problem:

Check Your Understanding Copper wires are routinely used for extension cords and house wiring for several reasons. Copper has the highest electrical conductivity rating, and therefore the lowest resistivity rating, of all nonprecious metals. Also important is the tensile strength, where the tensile strength is a measure of the force required to pull an object to the point where it breaks. The tensile strength of a material is the maximum amount of tensile stress it can take before breaking. Copper has a high tensile strength, $2 \times 10^8 \frac{\text{N}}{\text{m}^2}$. A third important characteristic is ductility. Ductility is a measure of a material's ability to be drawn into wires and a measure of the flexibility of the material, and copper has a high ductility. Summarizing, for a conductor to be a suitable candidate for making wire, there are at least three important characteristics: low resistivity, high tensile strength, and high ductility. What other materials are used for wiring and what are the advantages and disadvantages?

Solution:

Silver, gold, and aluminum are all used for making wires. All four materials have a high conductivity, silver having the highest. All four can easily be drawn into wires and have a high tensile strength, though not as high as copper. The obvious disadvantage of gold and silver is the cost, but silver and gold wires are used for special applications, such as speaker wires. Gold does not oxidize, making better connections between components. Aluminum wires do have their drawbacks. Aluminum has a higher resistivity than copper, so a larger diameter is needed to match the resistance per length of copper wires, but aluminum is cheaper than copper, so this is not a major drawback. Aluminum wires do not have as high of a ductility and tensile strength as copper, but the ductility and tensile strength is within acceptable levels. There are a few concerns that must be addressed in using aluminum and care must be used when making connections. Aluminum has a higher rate of thermal expansion than copper, which can lead to loose connections and a possible fire hazard. The oxidation of aluminum does not conduct and can cause problems. Special techniques must be used when using aluminum wires and components, such as electrical outlets, must be designed to accept aluminum wires.

Temperature Dependence of Resistivity

Looking back at [\[link\]](#), you will see a column labeled “Temperature Coefficient.” The resistivity of some materials has a strong temperature dependence. In some materials, such as copper, the resistivity increases with increasing temperature. In fact, in most conducting metals, the resistivity increases with increasing temperature. The increasing temperature causes increased vibrations of the atoms in the lattice structure of the metals, which impede the motion of the electrons. In other materials, such as carbon, the resistivity decreases with increasing temperature. In many materials, the dependence is approximately linear and can be modeled using a linear equation:

Note:

Equation:

$$\rho \approx \rho_0 [1 + \alpha (T - T_0)],$$

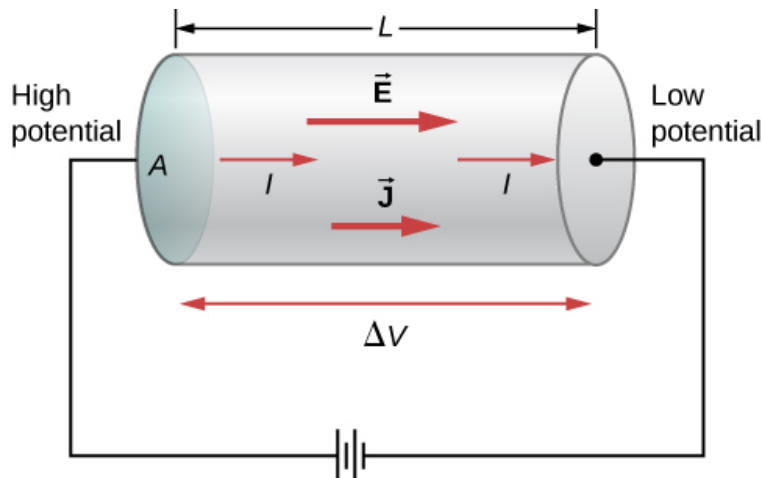
where ρ is the resistivity of the material at temperature T , α is the temperature coefficient of the material, and ρ_0 is the resistivity at T_0 , usually taken as $T_0 = 20.00^\circ\text{C}$.

Note also that the temperature coefficient α is negative for the semiconductors listed in [\[link\]](#), meaning that their resistivity decreases with increasing temperature. They become better conductors at higher temperature, because increased thermal agitation increases the number of free charges available to carry current. This property of decreasing ρ with temperature is also related to the type and amount of impurities present in the semiconductors.

Resistance

We now consider the resistance of a wire or component. The resistance is a measure of how difficult it is to pass current through a wire or component. Resistance depends on the resistivity. The resistivity is a characteristic of the material used to fabricate a wire or other electrical component, whereas the resistance is a characteristic of the wire or component.

To calculate the resistance, consider a section of conducting wire with cross-sectional area A , length L , and resistivity ρ . A battery is connected across the conductor, providing a potential difference ΔV across it ([\[link\]](#)). The potential difference produces an electrical field that is proportional to the current density, according to $\vec{E} = \rho \vec{J}$.



A potential provided by a battery is applied to a segment of a conductor with a cross-sectional area A and a length L .

The magnitude of the electrical field across the segment of the conductor is equal to the voltage divided by the length, $E = V/L$, and the magnitude of the current density is equal to the current divided by the cross-sectional area, $J = I/A$. Using this information and recalling that the electrical field is proportional to the resistivity and the current density, we can see that the voltage is proportional to the current:

Equation:

$$\begin{aligned} E &= \rho J \\ \frac{V}{L} &= \rho \frac{I}{A} \\ V &= \left(\rho \frac{L}{A}\right) I. \end{aligned}$$

Note:

Resistance

The ratio of the voltage to the current is defined as the **resistance** R :

Equation:

$$R \equiv \frac{V}{I}.$$

The resistance of a cylindrical segment of a conductor is equal to the resistivity of the material times the length divided by the area:

Equation:

$$R \equiv \frac{V}{I} = \rho \frac{L}{A}.$$

The unit of resistance is the ohm, Ω . For a given voltage, the higher the resistance, the lower the current.

Resistors

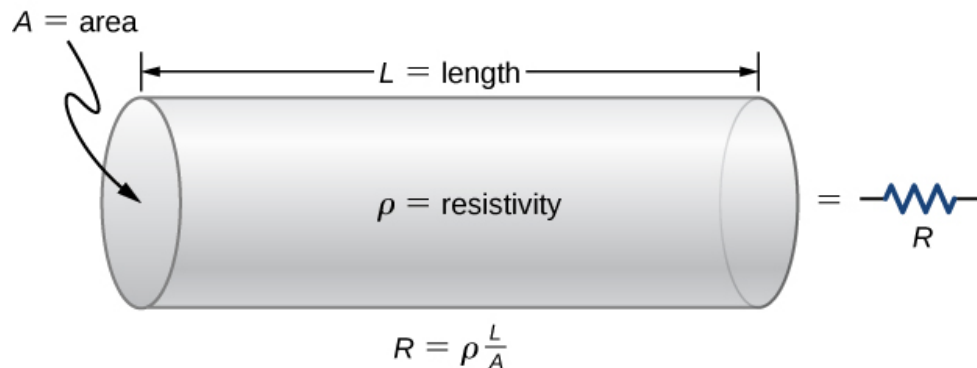
A common component in electronic circuits is the resistor. The resistor can be used to reduce current flow or provide a voltage drop. [\[link\]](#) shows the symbols used for a resistor in schematic diagrams of a circuit. Two commonly used standards for circuit diagrams are provided by the American National Standard Institute (ANSI, pronounced “AN-see”) and the International Electrotechnical Commission (IEC). Both systems are commonly used. We use the ANSI standard in this text for its visual recognition, but we note that for larger, more complex circuits, the IEC standard may have a cleaner presentation, making it easier to read.



Symbols for a resistor used in circuit diagrams. (a) The ANSI symbol; (b) the IEC symbol.

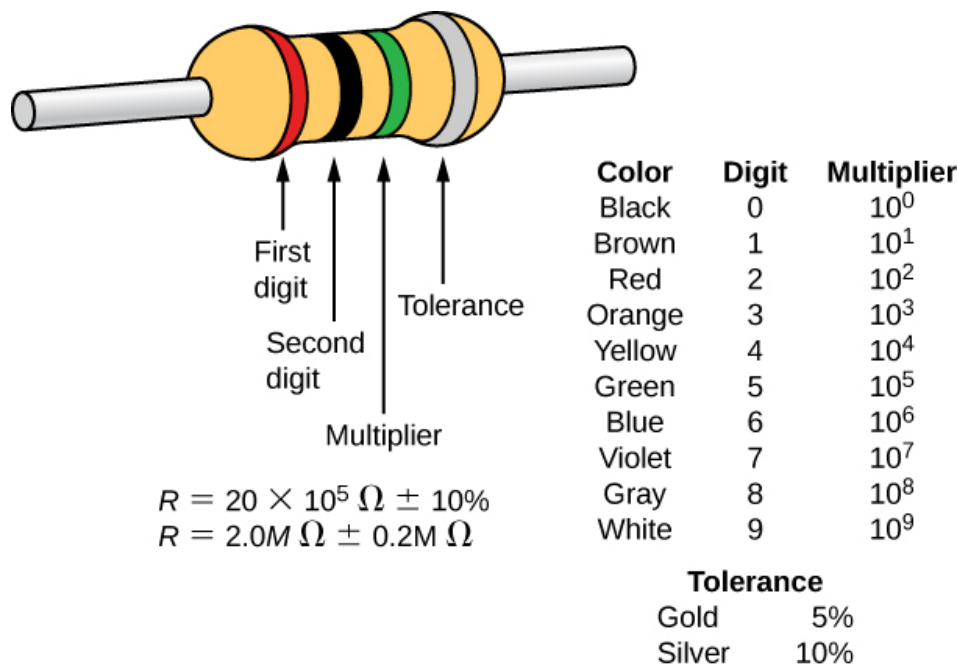
Material and shape dependence of resistance

A resistor can be modeled as a cylinder with a cross-sectional area A and a length L , made of a material with a resistivity ρ ([link](#)). The resistance of the resistor is $R = \rho \frac{L}{A}$.



A model of a resistor as a uniform cylinder of length L and cross-sectional area A . Its resistance to the flow of current is analogous to the resistance posed by a pipe to fluid flow. The longer the cylinder, the greater its resistance. The larger its cross-sectional area A , the smaller its resistance.

The most common material used to make a resistor is carbon. A carbon track is wrapped around a ceramic core, and two copper leads are attached. A second type of resistor is the metal film resistor, which also has a ceramic core. The track is made from a metal oxide material, which has semiconductive properties similar to carbon. Again, copper leads are inserted into the ends of the resistor. The resistor is then painted and marked for identification. A resistor has four colored bands, as shown in [link](#).



Many resistors resemble the figure shown above. The four bands are used to identify the resistor. The first two colored bands represent the first two digits of the resistance of the resistor. The third color is the multiplier. The fourth color represents the tolerance of the resistor.

The resistor shown has a resistance of $20 \times 10^5 \Omega \pm 10\%$.

Resistances range over many orders of magnitude. Some ceramic insulators, such as those used to support power lines, have resistances of $10^{12} \Omega$ or more. A dry person may have a hand-to-foot resistance of $10^5 \Omega$, whereas the resistance of the human heart is about $10^3 \Omega$. A meter-long piece of large-diameter copper wire may have a resistance of $10^{-5} \Omega$, and superconductors have no resistance at all at low temperatures. As we have seen, resistance is related to the shape of an object and the material of which it is composed.

Example:

Current Density, Resistance, and Electrical field for a Current-Carrying Wire

Calculate the current density, resistance, and electrical field of a 5-m length of copper wire with a diameter of 2.053 mm (12-gauge) carrying a current of $I = 10 \text{ mA}$.

Strategy

We can calculate the current density by first finding the cross-sectional area of the wire, which is $A = 3.31 \text{ mm}^2$, and the definition of current density $J = \frac{I}{A}$. The resistance can be found using the length of the wire $L = 5.00 \text{ m}$, the area, and the resistivity of copper $\rho = 1.68 \times 10^{-8} \Omega \cdot \text{m}$, where $R = \rho \frac{L}{A}$. The resistivity and current density can be used to find the electrical field.

Solution

First, we calculate the current density:

Equation:

$$J = \frac{I}{A} = \frac{10 \times 10^{-3} \text{ A}}{3.31 \times 10^{-6} \text{ m}^2} = 3.02 \times 10^3 \frac{\text{A}}{\text{m}^2}.$$

The resistance of the wire is

Equation:

$$R = \rho \frac{L}{A} = (1.68 \times 10^{-8} \Omega \cdot \text{m}) \frac{5.00 \text{ m}}{3.31 \times 10^{-6} \text{ m}^2} = 0.025 \Omega.$$

Finally, we can find the electrical field:

Equation:

$$E = \rho J = 1.68 \times 10^{-8} \Omega \cdot \text{m} \left(3.02 \times 10^3 \frac{\text{A}}{\text{m}^2} \right) = 5.07 \times 10^{-5} \frac{\text{V}}{\text{m}}.$$

Significance

From these results, it is not surprising that copper is used for wires for carrying current because the resistance is quite small. Note that the current density and electrical field are independent of the length of the wire, but the voltage depends on the length.

Note:

View this [interactive simulation](#) to see what the effects of the cross-sectional area, the length, and the resistivity of a wire are on the resistance of a conductor. Adjust the variables using slide bars and see if the resistance becomes smaller or larger.

The resistance of an object also depends on temperature, since R_0 is directly proportional to ρ . For a cylinder, we know $R = \rho \frac{L}{A}$, so if L and A do not change greatly with temperature, R has the same temperature dependence as ρ . (Examination of the coefficients of linear expansion shows them to be about two orders of magnitude less than typical temperature coefficients of resistivity, so the effect of temperature on L and A is about two orders of magnitude less than on ρ .) Thus,

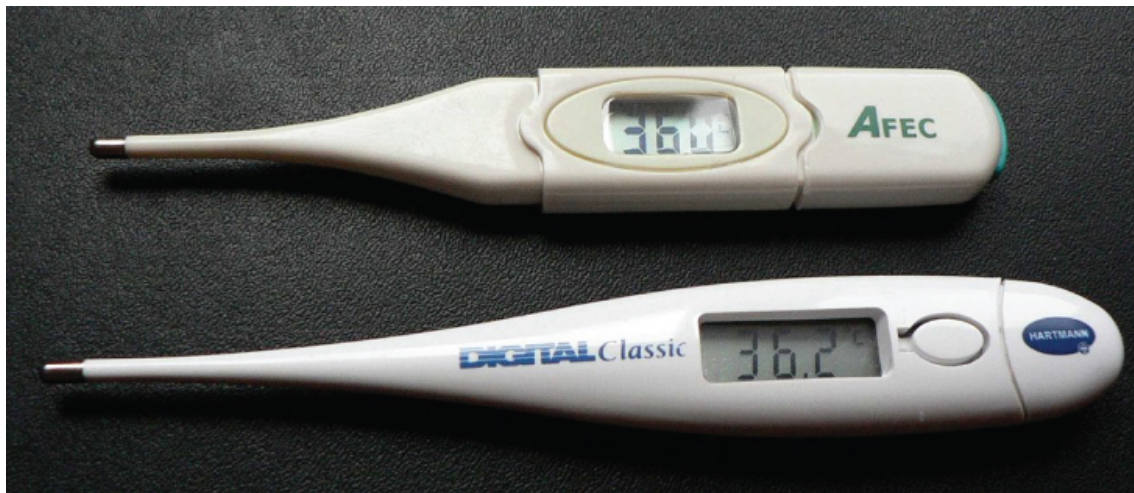
Note:

Equation:

$$R = R_0(1 + \alpha \Delta T)$$

is the temperature dependence of the resistance of an object, where R_0 is the original resistance (usually taken to be $20.00\text{ }^{\circ}\text{C}$) and R is the resistance after a temperature change ΔT . The color code gives the resistance of the resistor at a temperature of $T = 20.00\text{ }^{\circ}\text{C}$.

Numerous thermometers are based on the effect of temperature on resistance ([link](#)). One of the most common thermometers is based on the thermistor, a semiconductor crystal with a strong temperature dependence, the resistance of which is measured to obtain its temperature. The device is small, so that it quickly comes into thermal equilibrium with the part of a person it touches.



These familiar thermometers are based on the automated measurement of a thermistor's temperature-dependent resistance.

Example:

Calculating Resistance

Although caution must be used in applying $\rho = \rho_0(1 + \alpha\Delta T)$ and $R = R_0(1 + \alpha\Delta T)$ for temperature changes greater than $100\text{ }^{\circ}\text{C}$, for tungsten, the equations work reasonably well for very large temperature changes. A tungsten filament at $20\text{ }^{\circ}\text{C}$ has a resistance of $0.350\text{ }\Omega$. What would the resistance be if the temperature is increased to $2850\text{ }^{\circ}\text{C}$?

Strategy

This is a straightforward application of $R = R_0(1 + \alpha\Delta T)$, since the original resistance of the filament is given as $R_0 = 0.350\text{ }\Omega$ and the temperature change is $\Delta T = 2830\text{ }^{\circ}\text{C}$.

Solution

The resistance of the hotter filament R is obtained by entering known values into the above equation:

Equation:

$$R = R_0 (1 + \alpha \Delta T) = (0.350 \, \Omega) \left[1 + \left(\frac{4.5 \times 10^{-3}}{^{\circ}\text{C}} \right) (2830 \, ^{\circ}\text{C}) \right] = 4.8 \, \Omega .$$

Significance

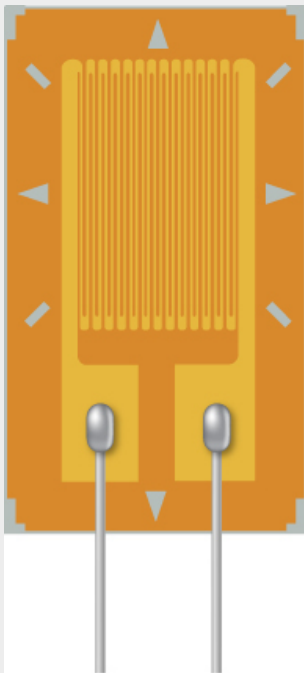
Notice that the resistance changes by more than a factor of 10 as the filament warms to the high temperature and the current through the filament depends on the resistance of the filament and the voltage applied. If the filament is used in an incandescent light bulb, the initial current through the filament when the bulb is first energized will be higher than the current after the filament reaches the operating temperature.

Note:

Exercise:

Problem:

Check Your Understanding A strain gauge is an electrical device to measure strain, as shown below. It consists of a flexible, insulating backing that supports a conduction foil pattern. The resistance of the foil changes as the backing is stretched. How does the strain gauge resistance change? Is the strain gauge affected by temperature changes?



Solution:

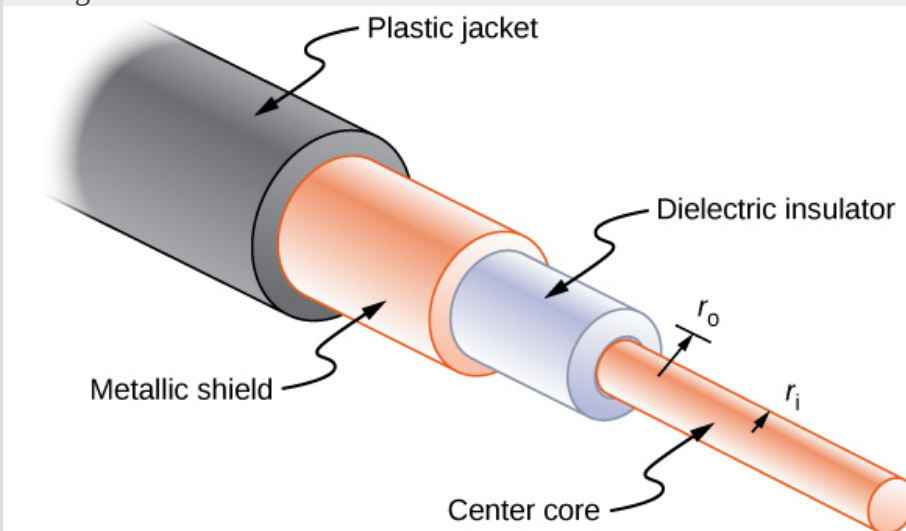
The foil pattern stretches as the backing stretches, and the foil tracks become longer and thinner. Since the resistance is calculated as $R = \rho \frac{L}{A}$, the resistance increases as the foil tracks are stretched. When the temperature changes, so does the resistivity of the foil tracks, changing the resistance. One way to combat this is to use two strain gauges, one used as a

reference and the other used to measure the strain. The two strain gauges are kept at a constant temperature

Example:

The Resistance of Coaxial Cable

Long cables can sometimes act like antennas, picking up electronic noise, which are signals from other equipment and appliances. Coaxial cables are used for many applications that require this noise to be eliminated. For example, they can be found in the home in cable TV connections or other audiovisual connections. Coaxial cables consist of an inner conductor of radius r_i surrounded by a second, outer concentric conductor with radius r_o ([link](#)). The space between the two is normally filled with an insulator such as polyethylene plastic. A small amount of radial leakage current occurs between the two conductors. Determine the resistance of a coaxial cable of length L .



Coaxial cables consist of two concentric conductors separated by insulation. They are often used in cable TV or other audiovisual connections.

Strategy

We cannot use the equation $R = \rho \frac{L}{A}$ directly. Instead, we look at concentric cylindrical shells, with thickness dr , and integrate.

Solution

We first find an expression for dR and then integrate from r_i to r_o ,

Equation:

$$dR = \frac{\rho}{A} dr = \frac{\rho}{2\pi r L} dr,$$

$$R = \int_{r_i}^{r_o} dR = \int_{r_i}^{r_o} \frac{\rho}{2\pi r L} dr = \frac{\rho}{2\pi L} \int_{r_i}^{r_o} \frac{1}{r} dr = \frac{\rho}{2\pi L} \ln \frac{r_o}{r_i}.$$

Significance

The resistance of a coaxial cable depends on its length, the inner and outer radii, and the resistivity of the material separating the two conductors. Since this resistance is not infinite, a small leakage current occurs between the two conductors. This leakage current leads to the attenuation (or weakening) of the signal being sent through the cable.

Note:

Exercise:

Problem:

Check Your Understanding The resistance between the two conductors of a coaxial cable depends on the resistivity of the material separating the two conductors, the length of the cable and the inner and outer radius of the two conductor. If you are designing a coaxial cable, how does the resistance between the two conductors depend on these variables?

Solution:

The longer the length, the smaller the resistance. The greater the resistivity, the higher the resistance. The larger the difference between the outer radius and the inner radius, that is, the greater the ratio between the two, the greater the resistance. If you are attempting to maximize the resistance, the choice of the values for these variables will depend on the application. For example, if the cable must be flexible, the choice of materials may be limited.

Note:

View this [simulation](#) to see how the voltage applied and the resistance of the material the current flows through affects the current through the material. You can visualize the collisions of the electrons and the atoms of the material effect the temperature of the material.

Summary

- Resistance has units of ohms (Ω), related to volts and amperes by $1 \Omega = 1 \text{ V/A}$.
- The resistance R of a cylinder of length L and cross-sectional area A is $R = \frac{\rho L}{A}$, where ρ is the resistivity of the material.

- Values of ρ in [\[link\]](#) show that materials fall into three groups—conductors, semiconductors, and insulators.
- Temperature affects resistivity; for relatively small temperature changes ΔT , resistivity is $\rho = \rho_0 (1 + \alpha \Delta T)$, where ρ_0 is the original resistivity and α is the temperature coefficient of resistivity.
- The resistance R of an object also varies with temperature: $R = R_0 (1 + \alpha \Delta T)$, where R_0 is the original resistance, and R is the resistance after the temperature change.

Conceptual Questions

Exercise:

Problem:

The IR drop across a resistor means that there is a change in potential or voltage across the resistor. Is there any change in current as it passes through a resistor? Explain.

Exercise:

Problem:

Do impurities in semiconducting materials listed in [\[link\]](#) supply free charges? (*Hint:* Examine the range of resistivity for each and determine whether the pure semiconductor has the higher or lower conductivity.)

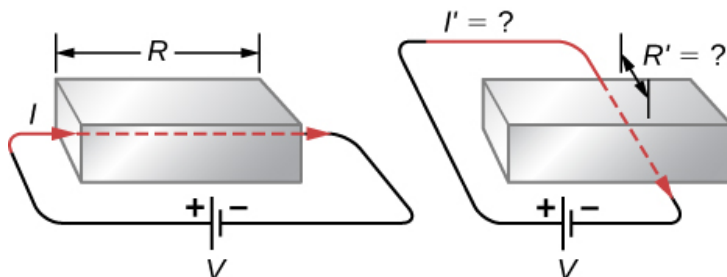
Solution:

In carbon, resistivity increases with the amount of impurities, meaning fewer free charges. In silicon and germanium, impurities decrease resistivity, meaning more free electrons.

Exercise:

Problem:

Does the resistance of an object depend on the path current takes through it? Consider, for example, a rectangular bar—is its resistance the same along its length as across its width?



Exercise:

Problem:

If aluminum and copper wires of the same length have the same resistance, which has the larger diameter? Why?

Solution:

Copper has a lower resistivity than aluminum, so if length is the same, copper must have the smaller diameter.

Problems**Exercise:****Problem:**

What current flows through the bulb of a 3.00-V flashlight when its hot resistance is $3.60\ \Omega$?

Exercise:**Problem:**

Calculate the effective resistance of a pocket calculator that has a 1.35-V battery and through which 0.200 mA flows.

Solution:

$$R = 6.750\ \text{k}\Omega$$

Exercise:**Problem:**

How many volts are supplied to operate an indicator light on a DVD player that has a resistance of $140\ \Omega$, given that 25.0 mA passes through it?

Exercise:**Problem:**

What is the resistance of a 20.0-m-long piece of 12-gauge copper wire having a 2.053-mm diameter?

Solution:

$$R = 0.10\ \Omega$$

Exercise:

Problem:

The diameter of 0-gauge copper wire is 8.252 mm. Find the resistance of a 1.00-km length of such wire used for power transmission.

Exercise:**Problem:**

If the 0.100-mm-diameter tungsten filament in a light bulb is to have a resistance of $0.200\ \Omega$ at $20.0\ ^\circ\text{C}$, how long should it be?

Solution:

$$R = \rho \frac{L}{A}$$

$$L = 3\ \text{cm}$$

Exercise:**Problem:**

A lead rod has a length of 30.00 cm and a resistance of $5.00\ \mu\Omega$. What is the radius of the rod?

Exercise:**Problem:**

Find the ratio of the diameter of aluminum to copper wire, if they have the same resistance per unit length (as they might in household wiring).

Solution:

$$\frac{R_{\text{Al}}/L_{\text{Al}}}{R_{\text{Cu}}/L_{\text{Cu}}} = \frac{\rho_{\text{Al}} \frac{1}{\pi \left(\frac{D_{\text{Al}}}{2}\right)^2}}{\rho_{\text{Cu}} \frac{1}{\pi \left(\frac{D_{\text{Cu}}}{2}\right)^2}} = \frac{\rho_{\text{Al}}}{\rho_{\text{Cu}}} \left(\frac{D_{\text{Cu}}}{D_{\text{Al}}}\right)^2 = 1, \quad \frac{D_{\text{Al}}}{D_{\text{Cu}}} = \sqrt{\frac{\rho_{\text{Al}}}{\rho_{\text{Cu}}}}$$

Exercise:**Problem:**

What current flows through a 2.54-cm-diameter rod of pure silicon that is 20.0 cm long, when $1.00 \times 10^3\ \text{V}$ is applied to it? (Such a rod may be used to make nuclear-particle detectors, for example.)

Exercise:**Problem:**

(a) To what temperature must you raise a copper wire, originally at $20.0\ ^\circ\text{C}$, to double its resistance, neglecting any changes in dimensions? (b) Does this happen in household wiring under ordinary circumstances?

Solution:

- a. $R = R_0 (1 + \alpha \Delta T)$, $2 = 1 + \alpha \Delta T$, $\Delta T = 256.4^\circ\text{C}$, $T = 276.4^\circ\text{C}$;
b. Under normal conditions, no it should not occur.

Exercise:**Problem:**

A resistor made of nichrome wire is used in an application where its resistance cannot change more than 1.00% from its value at 20.0°C . Over what temperature range can it be used?

Exercise:**Problem:**

Of what material is a resistor made if its resistance is 40.0% greater at 100.0°C than at 20.0°C ?

Solution:

$$R = R_0 (1 + \alpha \Delta T), \text{ iron}$$
$$\alpha = 0.006^\circ\text{C}^{-1}$$

Exercise:**Problem:**

An electronic device designed to operate at any temperature in the range from -10.0°C to 55.0°C contains pure carbon resistors. By what factor does their resistance increase over this range?

Exercise:**Problem:**

- (a) Of what material is a wire made, if it is 25.0 m long with a diameter of 0.100 mm and has a resistance of $77.7\ \Omega$ at 20.0°C ? (b) What is its resistance at 150.0°C ?

Solution:

a. $R = \rho \frac{L}{A}$, $\rho = 2.44 \times 10^{-8}\ \Omega \cdot \text{m}$, gold;
 $R = \rho \frac{L}{A} (1 + \alpha \Delta T)$

b. $R = 2.44 \times 10^{-8}\ \Omega \cdot \text{m} \left(\frac{25\ \text{m}}{\pi \left(\frac{0.100 \times 10^{-3}\ \text{m}}{2} \right)^2} \right) (1 + 0.0034^\circ\text{C}^{-1} (150^\circ\text{C} - 20^\circ\text{C}))$
 $R = 112\ \Omega$

Exercise:

Problem:

Assuming a constant temperature coefficient of resistivity, what is the maximum percent decrease in the resistance of a constantan wire starting at 20.0°C ?

Exercise:**Problem:**

A copper wire has a resistance of $0.500\ \Omega$ at 20.0°C , and an iron wire has a resistance of $0.525\ \Omega$ at the same temperature. At what temperature are their resistances equal?

Solution:

$$R_{\text{Fe}} = 0.525\ \Omega, \quad R_{\text{Cu}} = 0.500\ \Omega, \quad \alpha_{\text{Fe}} = 0.0065\ ^\circ\text{C}^{-1} \quad \alpha_{\text{Cu}} = 0.0039\ ^\circ\text{C}^{-1}$$

$$R_{\text{Fe}} = R_{\text{Cu}}$$

$$R_{0\text{Fe}} (1 + \alpha_{\text{Fe}} (T - T_0)) = R_{0\text{Cu}} (1 + \alpha_{\text{Cu}} (T - T_0))$$

$$\frac{R_{0\text{Fe}}}{R_{0\text{Cu}}} (1 + \alpha_{\text{Fe}} (T - T_0)) = 1 + \alpha_{\text{Cu}} (T - T_0)$$

$$T = 2.91\ ^\circ\text{C}$$

Glossary

electrical conductivity

measure of a material's ability to conduct or transmit electricity

ohm

(Ω) unit of electrical resistance, $1\ \Omega = 1\ \text{V/A}$

resistance

electric property that impedes current; for ohmic materials, it is the ratio of voltage to current, $R = V/I$

resistivity

intrinsic property of a material, independent of its shape or size, directly proportional to the resistance, denoted by ρ

Ohm's Law

By the end of this section, you will be able to:

- Describe Ohm's law
- Recognize when Ohm's law applies and when it does not

We have been discussing three electrical properties so far in this chapter: current, voltage, and resistance. It turns out that many materials exhibit a simple relationship among the values for these properties, known as Ohm's law. Many other materials do not show this relationship, so despite being called Ohm's law, it is not considered a law of nature, like Newton's laws or the laws of thermodynamics. But it is very useful for calculations involving materials that do obey Ohm's law.

Description of Ohm's Law

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is *directly proportional to the voltage applied*:

Equation:

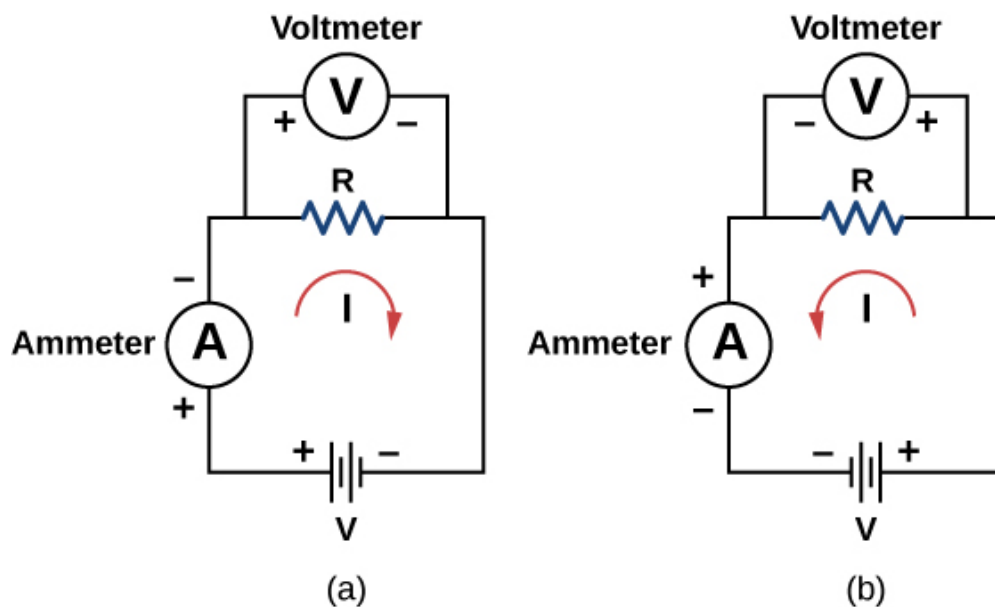
$$I \propto V.$$

This important relationship is the basis for **Ohm's law**. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law, which is to say that it is an experimentally observed phenomenon, like friction. Such a linear relationship doesn't always occur. Any material, component, or device that obeys Ohm's law, where the current through the device is proportional to the voltage applied, is known as an **ohmic** material or ohmic component. Any material or component that does not obey Ohm's law is known as a **nonohmic** material or nonohmic component.

Ohm's Experiment

In a paper published in 1827, Georg Ohm described an experiment in which he measured voltage across and current through various simple electrical circuits containing various lengths of wire. A similar experiment is shown in [\[link\]](#). This experiment is used to observe the current through a resistor that results from an applied voltage. In this simple circuit, a resistor is connected in series with a battery. The voltage is measured with a voltmeter, which must be placed across the resistor

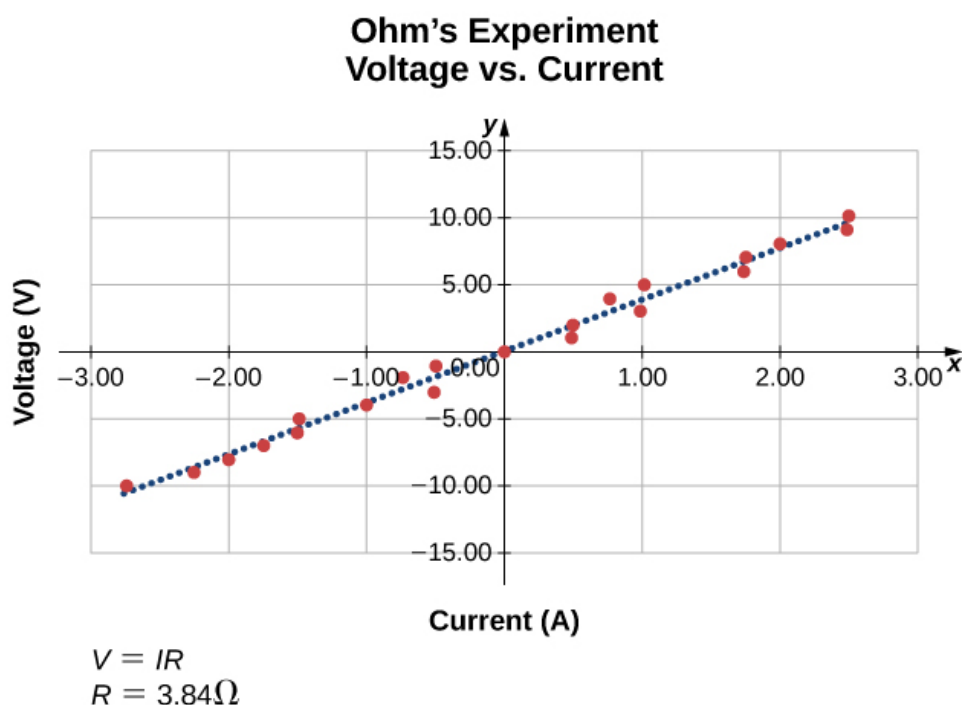
(in parallel with the resistor). The current is measured with an ammeter, which must be in line with the resistor (in series with the resistor).



The experimental set-up used to determine if a resistor is an ohmic or nonohmic device. (a) When the battery is attached, the current flows in the clockwise direction and the voltmeter and ammeter have positive readings. (b) When the leads of the battery are switched, the current flows in the counterclockwise direction and the voltmeter and ammeter have negative readings.

In this updated version of Ohm's original experiment, several measurements of the current were made for several different voltages. When the battery was hooked up as in [\[link\]](#)(a), the current flowed in the clockwise direction and the readings of the voltmeter and ammeter were positive. Does the behavior of the current change if the current flowed in the opposite direction? To get the current to flow in the opposite direction, the leads of the battery can be switched. When the leads of the battery were switched, the readings of the voltmeter and ammeter readings were negative because the current flowed in the opposite direction, in this case, counterclockwise. Results of a similar experiment are shown in [\[link\]](#).

I(A)	V(V)
-2.74	-10.00
-2.25	-9.00
-2.00	-8.00
-1.75	-7.00
-1.50	-6.00
-1.49	-5.00
-1.00	-4.00
-0.51	-3.00
-0.74	-2.00
-0.49	-1.00
+0.00	+0.00
+0.49	+1.00
+0.50	+2.00
+0.99	+3.00
+0.76	+4.00
+1.01	+5.00
+1.74	+6.00
+1.75	+7.00
+2.00	+8.00
+2.49	+9.00
+2.50	+10.00



A resistor is placed in a circuit with a battery. The voltage applied varies from -10.00 V to $+10.00\text{ V}$, increased by 1.00-V increments. A plot shows values of the voltage versus the current typical of what a casual experimenter might find.

In this experiment, the voltage applied across the resistor varies from -10.00 to $+10.00\text{ V}$, by increments of 1.00 V . The current through the resistor and the voltage across the resistor are measured. A plot is made of the voltage versus the current, and the result is approximately linear. The slope of the line is the resistance, or the voltage divided by the current. This result is known as Ohm's law:

Note:

Equation:

$$V = IR,$$

where V is the voltage measured in volts across the object in question, I is the current measured through the object in amps, and R is the resistance in units of ohms. As stated previously, any device that shows a linear relationship between the voltage and the current is known as an ohmic device. A resistor is therefore an ohmic device.

Example:**Measuring Resistance**

A carbon resistor at room temperature (20°C) is attached to a 9.00-V battery and the current measured through the resistor is 3.00 mA. (a) What is the resistance of the resistor measured in ohms? (b) If the temperature of the resistor is increased to 60°C by heating the resistor, what is the current through the resistor?

Strategy

(a) The resistance can be found using Ohm's law. Ohm's law states that $V = IR$, so the resistance can be found using $R = V/I$.

(b) First, the resistance is temperature dependent so the new resistance after the resistor has been heated can be found using $R = R_0(1 + \alpha\Delta T)$. The current can be found using Ohm's law in the form $I = V/R$.

Solution

- a. Using Ohm's law and solving for the resistance yields the resistance at room temperature:

Equation:

$$R = \frac{V}{I} = \frac{9.00 \text{ V}}{3.00 \times 10^{-3} \text{ A}} = 3.00 \times 10^3 \Omega = 3.00 \text{ k}\Omega.$$

- b. The resistance at 60°C can be found using $R = R_0(1 + \alpha\Delta T)$ where the temperature coefficient for carbon is $\alpha = -0.0005$.

$$R = R_0(1 + \alpha\Delta T) = 3.00 \times 10^3 (1 - 0.0005 (60^\circ\text{C} - 20^\circ\text{C})) = 2.94 \text{ k}\Omega$$

.

The current through the heated resistor is

Equation:

$$I = \frac{V}{R} = \frac{9.00 \text{ V}}{2.94 \times 10^3 \Omega} = 3.06 \times 10^{-3} \text{ A} = 3.06 \text{ mA}.$$

Significance

A change in temperature of 40°C resulted in a 2.00% change in current. This may not seem like a very great change, but changing electrical characteristics can have a

strong effect on the circuits. For this reason, many electronic appliances, such as computers, contain fans to remove the heat dissipated by components in the electric circuits.

Note:

Exercise:

Problem:

Check Your Understanding The voltage supplied to your house varies as $V(t) = V_{\max} \sin(2\pi ft)$. If a resistor is connected across this voltage, will Ohm's law $V = IR$ still be valid?

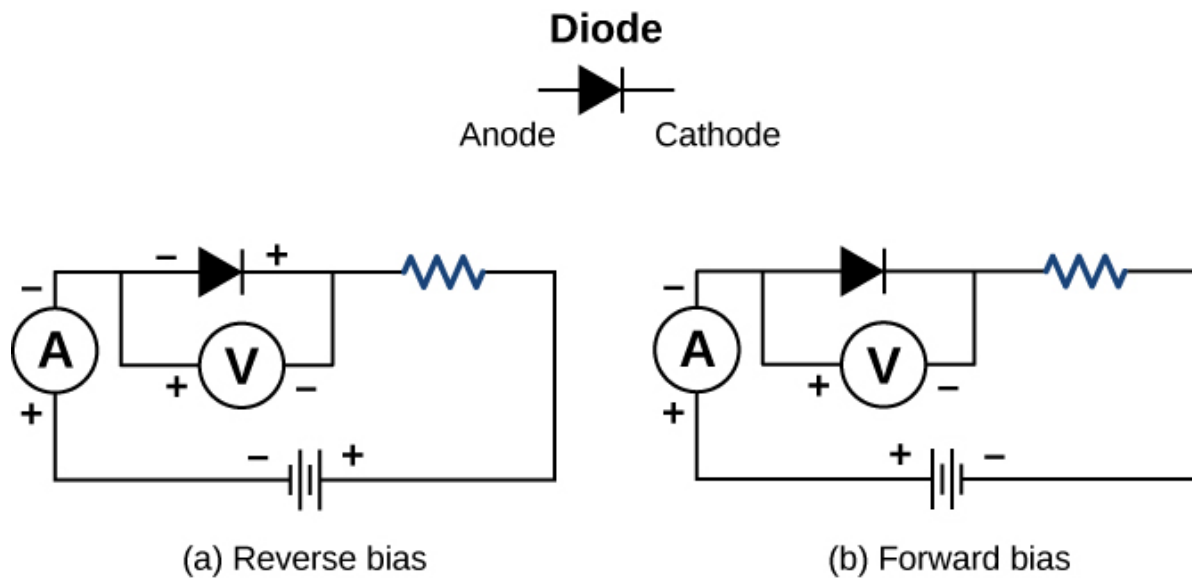
Solution:

Yes, Ohm's law is still valid. At every point in time the current is equal to $I(t) = V(t)/R$, so the current is also a function of time,
$$I(t) = \frac{V_{\max}}{R} \sin(2\pi ft).$$

Note:

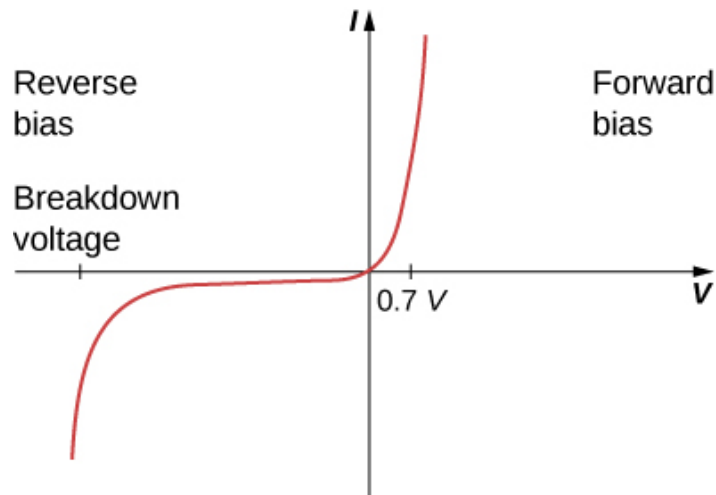
See how the [equation form of Ohm's law](#) relates to a simple circuit. Adjust the voltage and resistance, and see the current change according to Ohm's law. The sizes of the symbols in the equation change to match the circuit diagram.

Nonohmic devices do not exhibit a linear relationship between the voltage and the current. One such device is the semiconducting circuit element known as a diode. A **diode** is a circuit device that allows current flow in only one direction. A diagram of a simple circuit consisting of a battery, a diode, and a resistor is shown in [\[link\]](#). Although we do not cover the theory of the diode in this section, the diode can be tested to see if it is an ohmic or a nonohmic device.



A diode is a semiconducting device that allows current flow only if the diode is forward biased, which means that the anode is positive and the cathode is negative.

A plot of current versus voltage is shown in [\[link\]](#). Note that the behavior of the diode is shown as current versus voltage, whereas the resistor operation was shown as voltage versus current. A diode consists of an anode and a cathode. When the anode is at a negative potential and the cathode is at a positive potential, as shown in part (a), the diode is said to have reverse bias. With reverse bias, the diode has an extremely large resistance and there is very little current flow—essentially zero current—through the diode and the resistor. As the voltage applied to the circuit increases, the current remains essentially zero, until the voltage reaches the breakdown voltage and the diode conducts current, as shown in [\[link\]](#). When the battery and the potential across the diode are reversed, making the anode positive and the cathode negative, the diode conducts and current flows through the diode if the voltage is greater than 0.7 V. The resistance of the diode is close to zero. (This is the reason for the resistor in the circuit; if it were not there, the current would become very large.) You can see from the graph in [\[link\]](#) that the voltage and the current do not have a linear relationship. Thus, the diode is an example of a nonohmic device.



When the voltage across the diode is negative and small, there is very little current flow through the diode. As the voltage reaches the breakdown voltage, the diode conducts. When the voltage across the diode is positive and greater than 0.7 V (the actual voltage value depends on the diode), the diode conducts. As the voltage applied increases, the current through the diode increases, but the voltage across the diode remains approximately 0.7 V.

Ohm's law is commonly stated as $V = IR$, but originally it was stated as a microscopic view, in terms of the current density, the conductivity, and the electrical field. This microscopic view suggests the proportionality $V \propto I$ comes from the drift velocity of the free electrons in the metal that results from an applied electrical field. As stated earlier, the current density is proportional to the applied electrical field. The reformulation of Ohm's law is credited to Gustav Kirchhoff, whose name we will see again in the next chapter.

Summary

- Ohm's law is an empirical relationship for current, voltage, and resistance for some common types of circuit elements, including resistors. It does not apply to other devices, such as diodes.

- One statement of Ohm's law gives the relationship among current I , voltage V , and resistance R in a simple circuit as $V = IR$.
- Another statement of Ohm's law, on a microscopic level, is $J = \sigma E$.

Conceptual Questions

Exercise:

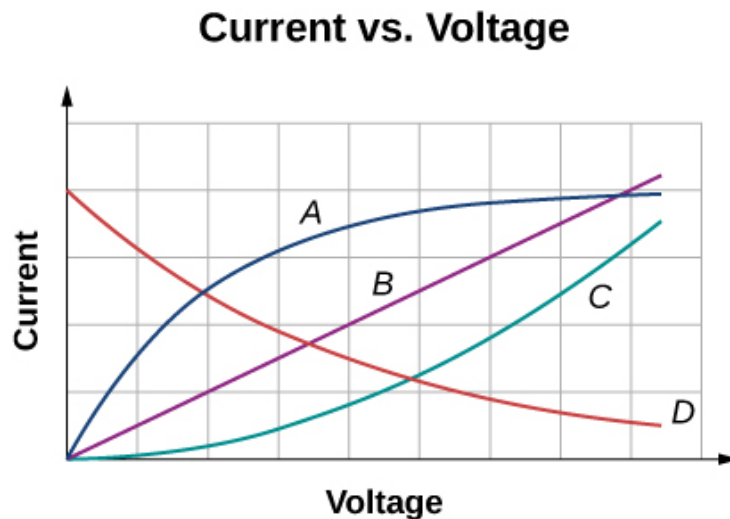
Problem:

In [Determining Field from Potential](#), resistance was defined as $R \equiv \frac{V}{I}$. In this section, we presented Ohm's law, which is commonly expressed as $V = IR$. The equations look exactly alike. What is the difference between Ohm's law and the definition of resistance?

Exercise:

Problem:

Shown below are the results of an experiment where four devices were connected across a variable voltage source. The voltage is increased and the current is measured. Which device, if any, is an ohmic device?



Solution:

Device B shows a linear relationship and the device is ohmic.

Exercise:

Problem:

The current I is measured through a sample of an ohmic material as a voltage V is applied. (a) What is the current when the voltage is doubled to $2V$ (assume the change in temperature of the material is negligible)? (b) What is the voltage applied is the current measured is $0.2I$ (assume the change in temperature of the material is negligible)? What will happen to the current if the material if the voltage remains constant, but the temperature of the material increases significantly?

Problems**Exercise:****Problem:**

A $2.2\text{-k}\Omega$ resistor is connected across a D cell battery (1.5 V). What is the current through the resistor?

Exercise:**Problem:**

A resistor rated at $250\text{ k}\Omega$ is connected across two D cell batteries (each 1.50 V) in series, with a total voltage of 3.00 V . The manufacturer advertises that their resistors are within 5% of the rated value. What are the possible minimum current and maximum current through the resistor?

Solution:

$$R_{\min} = 2.375 \times 10^5 \Omega, \quad I_{\min} = 12.63 \mu\text{ A}$$

$$R_{\max} = 2.625 \times 10^5 \Omega, \quad I_{\max} = 11.43 \mu\text{ A}$$

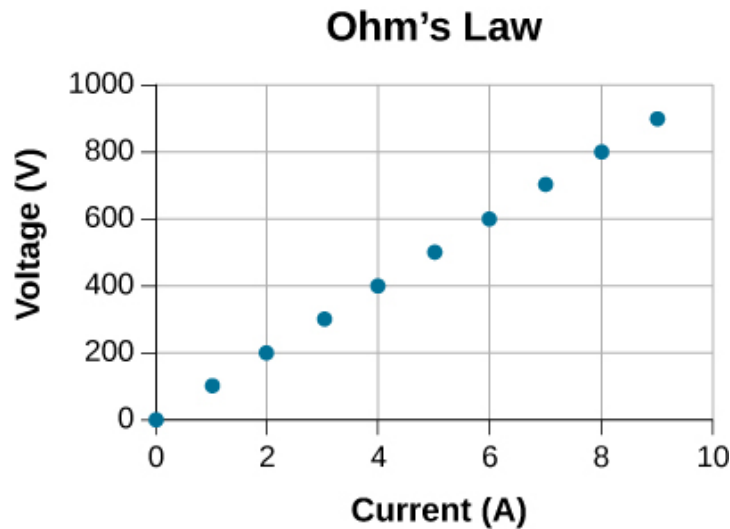
Exercise:**Problem:**

A resistor is connected in series with a power supply of 20.00 V . The current measure is 0.50 A . What is the resistance of the resistor?

Exercise:

Problem:

A resistor is placed in a circuit with an adjustable voltage source. The voltage across and the current through the resistor and the measurements are shown below. Estimate the resistance of the resistor.



Solution:

$$R = 100 \, \Omega$$

Exercise:**Problem:**

The following table show the measurements of a current through and the voltage across a sample of material. Plot the data, and assuming the object is an ohmic device, estimate the resistance.

$I(\text{A})$	$V(\text{V})$
0	3
2	23

$I(\text{A})$	$V(\text{V})$
4	39
6	58
8	77
10	100
12	119
14	142
16	162

Glossary

diode

nonohmic circuit device that allows current flow in only one direction

Ohm's law

empirical relation stating that the current I is proportional to the potential difference V ; it is often written as $V = IR$, where R is the resistance

ohmic

type of a material for which Ohm's law is valid, that is, the voltage drop across the device is equal to the current times the resistance

nonohmic

type of a material for which Ohm's law is not valid

Electrical Energy and Power

By the end of this section, you will be able to:

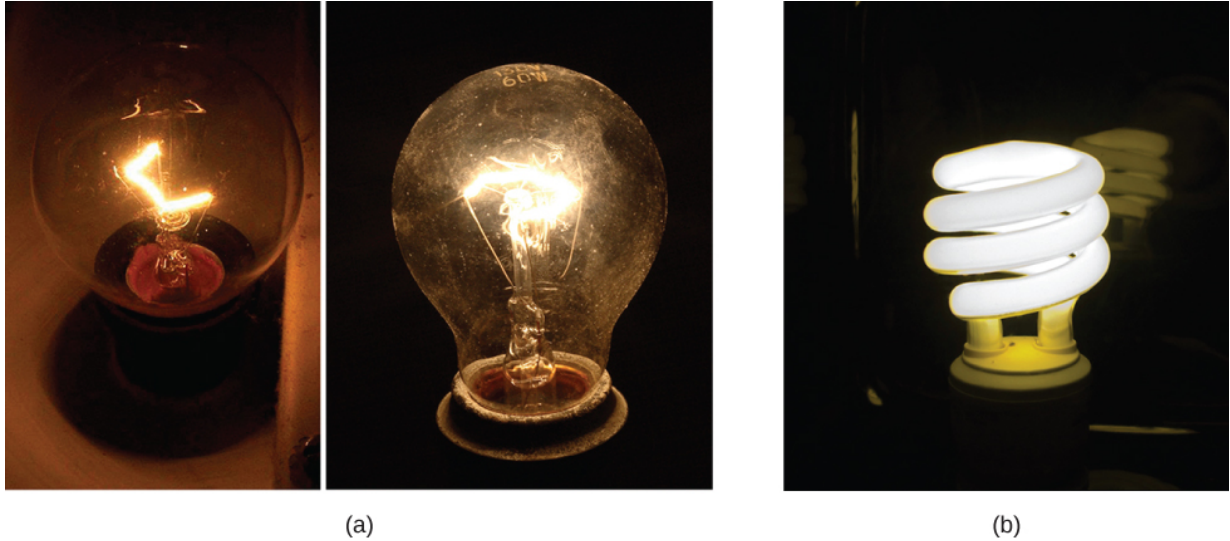
- Express electrical power in terms of the voltage and the current
- Describe the power dissipated by a resistor in an electric circuit
- Calculate the energy efficiency and cost effectiveness of appliances and equipment

In an electric circuit, electrical energy is continuously converted into other forms of energy. For example, when a current flows in a conductor, electrical energy is converted into thermal energy within the conductor. The electrical field, supplied by the voltage source, accelerates the free electrons, increasing their kinetic energy for a short time. This increased kinetic energy is converted into thermal energy through collisions with the ions of the lattice structure of the conductor. In [Work and Kinetic Energy](#), we defined power as the rate at which work is done by a force measured in watts. Power can also be defined as the rate at which energy is transferred. In this section, we discuss the time rate of energy transfer, or power, in an electric circuit.

Power in Electric Circuits

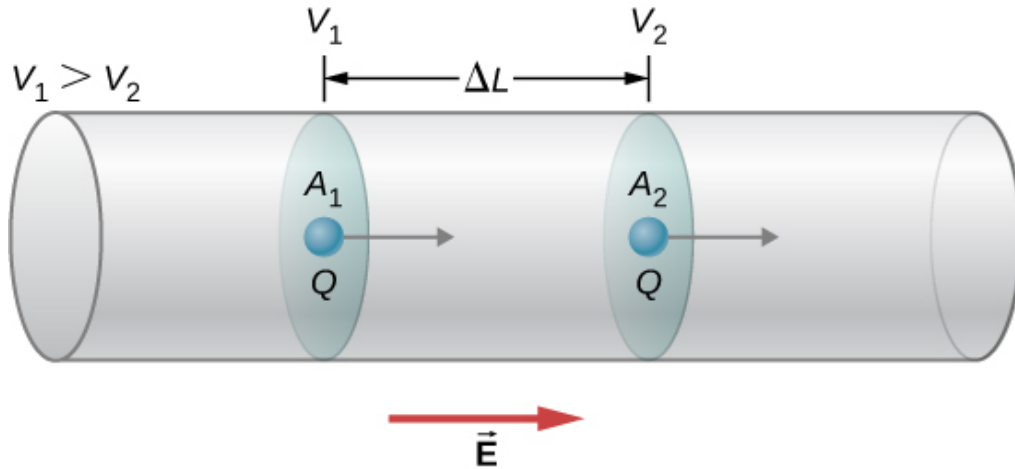
Power is associated by many people with electricity. Power transmission lines might come to mind. We also think of light bulbs in terms of their power ratings in watts. What is the expression for **electric power**?

Let us compare a 25-W bulb with a 60-W bulb ([link](#)(a)). The 60-W bulb glows brighter than the 25-W bulb. Although it is not shown, a 60-W light bulb is also warmer than the 25-W bulb. The heat and light is produced from the conversion of electrical energy. The kinetic energy lost by the electrons in collisions is converted into the internal energy of the conductor and radiation. How are voltage, current, and resistance related to electric power?



(a) Pictured above are two incandescent bulbs: a 25-W bulb (left) and a 60-W bulb (right). The 60-W bulb provides a higher intensity light than the 25-W bulb. The electrical energy supplied to the light bulbs is converted into heat and light. (b) This compact fluorescent light (CFL) bulb puts out the same intensity of light as the 60-W bulb, but at 1/4 to 1/10 the input power. (credit a: modification of works by “Dickbauch”/Wikimedia Commons and Greg Westfall; credit b: modification of work by “dbgg1979”/Flickr)

To calculate electric power, consider a voltage difference existing across a material ([link](#)). The electric potential V_1 is higher than the electric potential at V_2 , and the voltage difference is negative $V = V_2 - V_1$. As discussed in [Electric Potential](#), an electrical field exists between the two potentials, which points from the higher potential to the lower potential. Recall that the electrical potential is defined as the potential energy per charge, $V = U/q$, and the charge Q loses potential energy moving through the potential difference.



When there is a potential difference across a conductor, an electrical field is present that points in the direction from the higher potential to the lower potential.

If the charge is positive, the charge experiences a force due to the electrical field $\vec{F} = m\vec{a} = Q\vec{E}$. This force is necessary to keep the charge moving. This force does not act to accelerate the charge through the entire distance ΔL because of the interactions of the charge with atoms and free electrons in the material. The speed, and therefore the kinetic energy, of the charge do not increase during the entire trip across ΔL , and charge passing through area A_2 has the same drift velocity v_d as the charge that passes through area A_1 . However, work is done on the charge, by the electrical field, which changes the potential energy. Since the change in the electrical potential difference is negative, the electrical field is found to be

Equation:

$$E = -\frac{(V_2 - V_1)}{\Delta L} = \frac{\Delta V}{\Delta L}.$$

The work done on the charge is equal to the electric force times the length at which the force is applied,

Equation:

$$W = F\Delta L = (QE)\Delta L = Q\left(-\frac{\Delta V}{\Delta L}\right)\Delta L = -Q\Delta V = -\Delta U.$$

The charge moves at a drift velocity v_d so the work done on the charge results in a loss of potential energy, but the average kinetic energy remains constant. The lost electrical potential energy appears as thermal energy in the material. On a microscopic scale, the energy transfer is due to collisions between the charge and the molecules of the material, which leads to an increase in temperature in the material. The loss of potential energy results in an increase in the temperature of the material, which is dissipated as radiation. In a resistor, it is dissipated as heat, and in a light bulb, it is dissipated as heat and light.

The power dissipated by the material as heat and light is equal to the time rate of change of the work:

Equation:

$$P = \frac{\Delta U}{\Delta t} = \frac{Q\Delta V}{\Delta t} = IV.$$

With a resistor, the voltage drop across the resistor is dissipated as heat. Ohm's law states that the voltage across the resistor is equal to the current times the resistance, $V = IR$. The power dissipated by the resistor is therefore

Equation:

$$P = IV = I(IR) = I^2R \text{ or } P = IV = \left(\frac{V}{R}\right)V = \frac{V^2}{R}.$$

If a resistor is connected to a battery, the power dissipated as radiant energy by the wires and the resistor is equal to $P = IV = I^2R = \frac{V^2}{R}$. The power supplied from the battery is equal to current times the voltage, $P = IV$.

Note:

Electric Power

The electric power gained or lost by any device has the form

Equation:

$$P = IV.$$

The power dissipated by a resistor has the form

Equation:

$$P = I^2 R = \frac{V^2}{R}.$$

Different insights can be gained from the three different expressions for electric power. For example, $P = V^2/R$ implies that the lower the resistance connected to a given voltage source, the greater the power delivered. Furthermore, since voltage is squared in $P = V^2/R$, the effect of applying a higher voltage is perhaps greater than expected. Thus, when the voltage is doubled to a 25-W bulb, its power nearly quadruples to about 100 W, burning it out. If the bulb's resistance remained constant, its power would be exactly 100 W, but at the higher temperature, its resistance is higher, too.

Example:

Calculating Power in Electric Devices

A DC winch motor is rated at 20.00 A with a voltage of 115 V. When the motor is running at its maximum power, it can lift an object with a weight of 4900.00 N a distance of 10.00 m, in 30.00 s, at a constant speed. (a) What is the power consumed by the motor? (b) What is the power used in lifting the object? Ignore air resistance. (c) Assuming that the difference in the power consumed by the motor and the power used lifting the object are dissipated as heat by the resistance of the motor, estimate the resistance of the motor?

Strategy

(a) The power consumed by the motor can be found using $P = IV$. (b) The power used in lifting the object at a constant speed can be found using $P = Fv$, where the speed is the distance divided by the time. The upward force supplied by the motor is equal to the weight of the object because the acceleration is zero. (c) The resistance of the motor can be found using $P = I^2 R$.

Solution

- a. The power consumed by the motor is equal to $P = IV$ and the current is given as 20.00 A and the voltage is 115.00 V:

Equation:

$$P = IV = (20.00 \text{ A})(115.00 \text{ V}) = 2300.00 \text{ W}.$$

- b. The power used lifting the object is equal to $P = Fv$ where the force is equal to the weight of the object (1960 N) and the magnitude of the velocity is $v = \frac{10.00 \text{ m}}{30.00 \text{ s}} = 0.33 \frac{\text{m}}{\text{s}}$,

Equation:

$$P = Fv = (4900 \text{ N})0.33 \text{ m/s} = 1633.33 \text{ W}.$$

- c. The difference in the power equals $2300.00 \text{ W} - 1633.33 \text{ W} = 666.67 \text{ W}$ and the resistance can be found using $P = I^2 R$:

Equation:

$$R = \frac{P}{I^2} = \frac{666.67 \text{ W}}{(20.00 \text{ A})^2} = 1.67 \Omega .$$

Significance

The resistance of the motor is quite small. The resistance of the motor is due to many windings of copper wire. The power dissipated by the motor can be significant since the thermal power dissipated by the motor would be quite large due to this small resistance; however, due to back emf, the current drawn by the motor is very small.

Note:

Exercise:

Problem:

Check Your Understanding Electric motors have a reasonably high efficiency. A 100-hp motor can have an efficiency of 90% and a 1-hp motor can have an efficiency of 80%. Why is it important to use high-performance motors?

Solution:

Even though electric motors are highly efficient 10–20% of the power consumed is wasted, not being used for doing useful work. Most of the 10–20% of the power lost is transferred into heat dissipated by the copper wires used to make the coils of the motor. This heat adds to the heat of the environment and adds to the demand on power plants providing the power.

The demand on the power plant can lead to increased greenhouse gases, particularly if the power plant uses coal or gas as fuel.

A fuse ([link](#)) is a device that protects a circuit from currents that are too high. A fuse is basically a short piece of wire between two contacts. As we have seen, when a current is running through a conductor, the kinetic energy of the charge carriers is converted into thermal energy in the conductor. The piece of wire in the fuse is under tension and has a low melting point. The wire is designed to heat up and break at the rated current. The fuse is destroyed and must be replaced, but it protects the rest of the circuit. Fuses act quickly, but there is a small time delay while the wire heats up and breaks.



A fuse consists of a piece of wire between two contacts. When a current passes through the wire that is greater than the rated current, the wire melts, breaking the connection. Pictured is a “blown” fuse where the wire broke protecting a circuit (credit: modification of work by “Shardayyy”/Flickr).

Circuit breakers are also rated for a maximum current, and open to protect the circuit, but can be reset. Circuit breakers react much faster. The operation of circuit breakers is not within the scope of this chapter and will be discussed in later chapters. Another method of protecting equipment and people is the ground

fault circuit interrupter (GFCI), which is common in bathrooms and kitchens. The GFCI outlets respond very quickly to changes in current. These outlets open when there is a change in magnetic field produced by current-carrying conductors, which is also beyond the scope of this chapter and is covered in a later chapter.

The Cost of Electricity

The more electric appliances you use and the longer they are left on, the higher your electric bill. This familiar fact is based on the relationship between energy and power. You pay for the energy used. Since $P = \frac{dE}{dt}$, we see that

Equation:

$$E = \int P dt$$

is the energy used by a device using power P for a time interval t . If power is delivered at a constant rate, then the energy can be found by $E = Pt$. For example, the more light bulbs burning, the greater P used; the longer they are on, the greater t is.

The energy unit on electric bills is the kilowatt-hour ($\text{kW} \cdot \text{h}$), consistent with the relationship $E = Pt$. It is easy to estimate the cost of operating electrical appliances if you have some idea of their power consumption rate in watts or kilowatts, the time they are on in hours, and the cost per kilowatt-hour for your electric utility. Kilowatt-hours, like all other specialized energy units such as food calories, can be converted into joules. You can prove to yourself that $1 \text{ kW} \cdot \text{h} = 3.6 \times 10^6 \text{ J}$.

The electrical energy (E) used can be reduced either by reducing the time of use or by reducing the power consumption of that appliance or fixture. This not only reduces the cost but also results in a reduced impact on the environment.

Improvements to lighting are some of the fastest ways to reduce the electrical energy used in a home or business. About 20% of a home's use of energy goes to lighting, and the number for commercial establishments is closer to 40%.

Fluorescent lights are about four times more efficient than incandescent lights—this is true for both the long tubes and the compact fluorescent lights (CFLs). (See [\[link\]](#)(b).) Thus, a 60-W incandescent bulb can be replaced by a 15-W CFL, which has the same brightness and color. CFLs have a bent tube inside a globe or a spiral-shaped tube, all connected to a standard screw-in base that fits standard

incandescent light sockets. (Original problems with color, flicker, shape, and high initial investment for CFLs have been addressed in recent years.)

The heat transfer from these CFLs is less, and they last up to 10 times longer than incandescent bulbs. The significance of an investment in such bulbs is addressed in the next example. New white LED lights (which are clusters of small LED bulbs) are even more efficient (twice that of CFLs) and last five times longer than CFLs.

Example:

Calculating the Cost Effectiveness of LED Bulb

The typical replacement for a 100-W incandescent bulb is a 20-W LED bulb. The 20-W LED bulb can provide the same amount of light output as the 100-W incandescent light bulb. What is the cost savings for using the LED bulb in place of the incandescent bulb for one year, assuming \$0.10 per kilowatt-hour is the average energy rate charged by the power company? Assume that the bulb is turned on for three hours a day.

Strategy

- (a) Calculate the energy used during the year for each bulb, using $E = Pt$.
- (b) Multiply the energy by the cost.

Solution

- a. Calculate the power for each bulb.

Equation:

$$E_{\text{Incandescent}} = Pt = 100 \text{ W} \left(\frac{1 \text{ kW}}{1000 \text{ W}} \right) \left(\frac{3 \text{ h}}{\text{day}} \right) (365 \text{ days}) = 109.5 \text{ kW} \cdot \text{h}$$

$$E_{\text{LED}} = Pt = 20 \text{ W} \left(\frac{1 \text{ kW}}{1000 \text{ W}} \right) \left(\frac{3 \text{ h}}{\text{day}} \right) (365 \text{ days}) = 21.90 \text{ kW} \cdot \text{h}$$

- b. Calculate the cost for each.

Equation:

$$\text{cost}_{\text{Incandescent}} = 109.5 \text{ kW} \cdot \text{h} \left(\frac{\$0.10}{\text{kW} \cdot \text{h}} \right) = \$10.95$$

$$\text{cost}_{\text{LED}} = 21.90 \text{ kW} \cdot \text{h} \left(\frac{\$0.10}{\text{kW} \cdot \text{h}} \right) = \$2.19$$

Significance

A LED bulb uses 80% less energy than the incandescent bulb, saving \$8.76 over the incandescent bulb for one year. The LED bulb can cost \$20.00 and the 100-W incandescent bulb can cost \$0.75, which should be calculated into the computation. A typical lifespan of an incandescent bulb is 1200 hours and is 50,000 hours for the LED bulb. The incandescent bulb would last 1.08 years at 3 hours a day and the LED bulb would last 45.66 years. The initial cost of the LED bulb is high, but the cost to the home owner will be \$0.69 for the incandescent bulbs versus \$0.44 for the LED bulbs per year. (Note that the LED bulbs are coming down in price.) The cost savings per year is approximately \$8.50, and that is just for one bulb.

Note:

Exercise:

Problem:

Check Your Understanding Is the efficiency of the various light bulbs the only consideration when comparing the various light bulbs?

Solution:

No, the efficiency is a very important consideration of the light bulbs, but there are many other considerations. As mentioned above, the cost of the bulbs and the life span of the bulbs are important considerations. For example, CFL bulbs contain mercury, a neurotoxin, and must be disposed of as hazardous waste. When replacing incandescent bulbs that are being controlled by a dimmer switch with LED, the dimmer switch may need to be replaced. The dimmer switches for LED lights are comparably priced to the incandescent light switches, but this is an initial cost which should be considered. The spectrum of light should also be considered, but there is a broad range of color temperatures available, so you should be able to find one that fits your needs. None of these considerations mentioned are meant to discourage the use of LED or CFL light bulbs, but they are considerations.

Changing light bulbs from incandescent bulbs to CFL or LED bulbs is a simple way to reduce energy consumption in homes and commercial sites. CFL bulbs operate with a much different mechanism than do incandescent lights. The

mechanism is complex and beyond the scope of this chapter, but here is a very general description of the mechanism. CFL bulbs contain argon and mercury vapor housed within a spiral-shaped tube. The CFL bulbs use a “ballast” that increases the voltage used by the CFL bulb. The ballast produce an electrical current, which passes through the gas mixture and excites the gas molecules. The excited gas molecules produce ultraviolet (UV) light, which in turn stimulates the fluorescent coating on the inside of the tube. This coating fluoresces in the visible spectrum, emitting visible light. Traditional fluorescent tubes and CFL bulbs had a short time delay of up to a few seconds while the mixture was being “warmed up” and the molecules reached an excited state. It should be noted that these bulbs do contain mercury, which is poisonous, but if the bulb is broken, the mercury is never released. Even if the bulb is broken, the mercury tends to remain in the fluorescent coating. The amount is also quite small and the advantage of the energy saving may outweigh the disadvantage of using mercury.

The CFL light bulbs are being replaced with LED light bulbs, where LED stands for “light-emitting diode.” The diode was briefly discussed as a nonohmic device, made of semiconducting material, which essentially permits current flow in one direction. LEDs are a special type of diode made of semiconducting materials infused with impurities in combinations and concentrations that enable the extra energy from the movement of the electrons during electrical excitation to be converted into visible light. Semiconducting devices will be explained in greater detail in [Condensed Matter Physics](#).

Commercial LEDs are quickly becoming the standard for commercial and residential lighting, replacing incandescent and CFL bulbs. They are designed for the visible spectrum and are constructed from gallium doped with arsenic and phosphorous atoms. The color emitted from an LED depends on the materials used in the semiconductor and the current. In the early years of LED development, small LEDs found on circuit boards were red, green, and yellow, but LED light bulbs can now be programmed to produce millions of colors of light as well as many different hues of white light.

Comparison of Incandescent, CFL, and LED Light Bulbs

The energy savings can be significant when replacing an incandescent light bulb or a CFL light bulb with an LED light. Light bulbs are rated by the amount of power that the bulb consumes, and the amount of light output is measured in lumens. The lumen (lm) is the SI -derived unit of luminous flux and is a measure of the total quantity of visible light emitted by a source. A 60-W incandescent

light bulb can be replaced with a 13- to 15-W CFL bulb or a 6- to 8-W LED bulb, all three of which have a light output of approximately 800 lm. A table of light output for some commonly used light bulbs appears in [\[link\]](#).

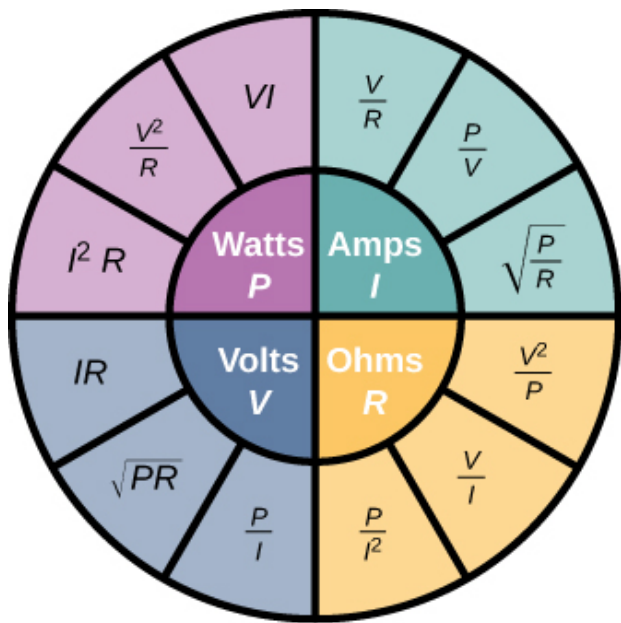
The life spans of the three types of bulbs are significantly different. An LED bulb has a life span of 50,000 hours, whereas the CFL has a lifespan of 8000 hours and the incandescent lasts a mere 1200 hours. The LED bulb is the most durable, easily withstanding rough treatment such as jarring and bumping. The incandescent light bulb has little tolerance to the same treatment since the filament and glass can easily break. The CFL bulb is also less durable than the LED bulb because of its glass construction. The amount of heat emitted is 3.4 btu/h for the 8-W LED bulb, 85 btu/h for the 60-W incandescent bulb, and 30 btu/h for the CFL bulb. As mentioned earlier, a major drawback of the CFL bulb is that it contains mercury, a neurotoxin, and must be disposed of as hazardous waste. From these data, it is easy to understand why the LED light bulb is quickly becoming the standard in lighting.

Light Output (lumens)	LED Light Bulbs (watts)	Incandescent Light Bulbs (watts)	CFL Light Bulbs (watts)
450	4–5	40	9–13
800	6–8	60	13–15
1100	9–13	75	18–25
1600	16–20	100	23–30
2600	25–28	150	30–55

Light Output of LED, Incandescent, and CFL Light Bulbs

Summary of Relationships

In this chapter, we have discussed relationships between voltages, current, resistance, and power. [\[link\]](#) shows a summary of the relationships between these measurable quantities for ohmic devices. (Recall that ohmic devices follow Ohm's law $V = IR$.) For example, if you need to calculate the power, use the pink section, which shows that $P = VI$, $P = \frac{V^2}{R}$, and $P = I^2 R$.



P = Power I = Current
 V = Voltage R = Resistance

This circle shows a summary of the equations for the relationships between power, current, voltage, and resistance.

Which equation you use depends on what values you are given, or you measure. For example if you are given the current and the resistance, use $P = I^2 R$. Although all the possible combinations may seem overwhelming, don't forget that they all are combinations of just two equations, Ohm's law ($V = IR$) and power ($P = IV$).

Summary

- Electric power is the rate at which electric energy is supplied to a circuit or consumed by a load.
- Power dissipated by a resistor depends on the square of the current through the resistor and is equal to $P = I^2 R = \frac{V^2}{R}$.
- The SI unit for electric power is the watt and the SI unit for electric energy is the joule. Another common unit for electric energy, used by power companies, is the kilowatt-hour ($\text{kW} \cdot \text{h}$).
- The total energy used over a time interval can be found by $E = \int P dt$.

Conceptual Questions

Exercise:

Problem:

Common household appliances are rated at 110 V, but power companies deliver voltage in the kilovolt range and then step the voltage down using transformers to 110 V to be used in homes. You will learn in later chapters that transformers consist of many turns of wire, which warm up as current flows through them, wasting some of the energy that is given off as heat. This sounds inefficient. Why do the power companies transport electric power using this method?

Solution:

Although the conductors have a low resistance, the lines from the power company can be kilometers long. Using a high voltage reduces the current that is required to supply the power demand and that reduces line losses.

Exercise:

Problem:

Your electric bill gives your consumption in units of kilowatt-hour ($\text{kW} \cdot \text{h}$). Does this unit represent the amount of charge, current, voltage, power, or energy you buy?

Exercise:

Problem:

Resistors are commonly rated at $\frac{1}{8}$ W, $\frac{1}{4}$ W, $\frac{1}{2}$ W, 1 W and 2 W for use in electrical circuits. If a current of $I = 2.00$ A is accidentally passed through a $R = 1.00\ \Omega$ resistor rated at 1 W, what would be the most probable outcome? Is there anything that can be done to prevent such an accident?

Solution:

The resistor would overheat, possibly to the point of causing the resistor to burn. Fuses are commonly added to circuits to prevent such accidents.

Exercise:**Problem:**

An immersion heater is a small appliance used to heat a cup of water for tea by passing current through a resistor. If the voltage applied to the appliance is doubled, will the time required to heat the water change? By how much? Is this a good idea?

Problems**Exercise:****Problem:**

A 20.00-V battery is used to supply current to a 10-k Ω resistor. Assume the voltage drop across any wires used for connections is negligible. (a) What is the current through the resistor? (b) What is the power dissipated by the resistor? (c) What is the power input from the battery, assuming all the electrical power is dissipated by the resistor? (d) What happens to the energy dissipated by the resistor?

Solution:

a. $I = 2$ mA; b. $P = 0.04$ W; c. $P = 0.04$ W; d. It is converted into heat.

Exercise:

Problem:

What is the maximum voltage that can be applied to a 20-k Ω resistor rated at $\frac{1}{4}$ W?

Exercise:**Problem:**

A heater is being designed that uses a coil of 14-gauge nichrome wire to generate 300 W using a voltage of $V = 110$ V. How long should the engineer make the wire?

Solution:

$$\begin{aligned} A &= 2.08 \text{ mm}^2 \\ P &= \frac{V^2}{R} \quad \rho = 100 \times 10^{-8} \Omega \cdot \text{m} \\ R &= 40 \Omega \quad R = \rho \frac{L}{A} \\ L &= 83 \text{ m} \end{aligned}$$

Exercise:**Problem:**

An alternative to CFL bulbs and incandescent bulbs are light-emitting diode (LED) bulbs. A 100-W incandescent bulb can be replaced by a 16-W LED bulb. Both produce 1600 lumens of light. Assuming the cost of electricity is \$0.10 per kilowatt-hour, how much does it cost to run the bulb for one year if it runs for four hours a day?

Exercise:**Problem:**

The power dissipated by a resistor with a resistance of $R = 100 \Omega$ is $P = 2.0$ W. What are the current through and the voltage drop across the resistor?

Solution:

$$I = 0.14 \text{ A}, \quad V = 14 \text{ V}$$

Exercise:

Problem:

Running late to catch a plane, a driver accidentally leaves the headlights on after parking the car in the airport parking lot. During takeoff, the driver realizes the mistake. Having just replaced the battery, the driver knows that the battery is a 12-V automobile battery, rated at 100 A · h. The driver, knowing there is nothing that can be done, estimates how long the lights will shine, assuming there are two 12-V headlights, each rated at 40 W. What did the driver conclude?

Exercise:**Problem:**

A physics student has a single-occupancy dorm room. The student has a small refrigerator that runs with a current of 3.00 A and a voltage of 110 V, a lamp that contains a 100-W bulb, an overhead light with a 60-W bulb, and various other small devices adding up to 3.00 W. (a) Assuming the power plant that supplies 110 V electricity to the dorm is 10 km away and the two aluminum transmission cables use 0-gauge wire with a diameter of 8.252 mm, estimate the percentage of the total power supplied by the power company that is lost in the transmission. (b) What would be the result is the power company delivered the electric power at 110 kV?

Solution:

$$I \approx 3.00 \text{ A} + \frac{100 \text{ W}}{110 \text{ V}} + \frac{60 \text{ W}}{110 \text{ V}} + \frac{3.00 \text{ W}}{110 \text{ V}} = 4.48 \text{ A}$$

$$P = 493 \text{ W}$$

a. $R = 9.91 \, \Omega$,

$$P_{\text{loss}} = 200. \text{ W}$$

$$\% \text{loss} = 40\%$$

$$P = 493 \text{ W}$$

$$I = 0.0045 \text{ A}$$

b. $R = 9.91 \, \Omega$

$$P_{\text{loss}} = 201 \mu \text{ W}$$

$$\% \text{loss} = 0.00004\%$$

Exercise:

Problem:

A 0.50-W, 220- Ω resistor carries the maximum current possible without damaging the resistor. If the current were reduced to half the value, what would be the power consumed?

Glossary

electrical power

time rate of change of energy in an electric circuit

Superconductors

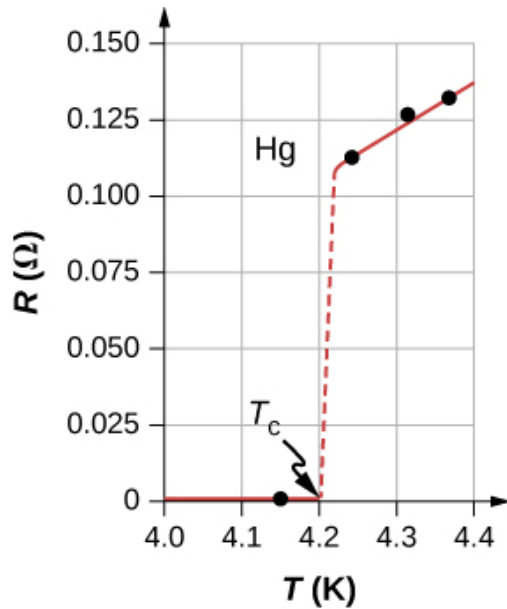
By the end of this section, you will be able to:

- Describe the phenomenon of superconductivity
- List applications of superconductivity

Touch the power supply of your laptop computer or some other device. It probably feels slightly warm. That heat is an unwanted byproduct of the process of converting household electric power into a current that can be used by your device. Although electric power is reasonably efficient, other losses are associated with it. As discussed in the section on power and energy, transmission of electric power produces I^2R line losses. These line losses exist whether the power is generated from conventional power plants (using coal, oil, or gas), nuclear plants, solar plants, hydroelectric plants, or wind farms. These losses can be reduced, but not eliminated, by transmitting using a higher voltage. It would be wonderful if these line losses could be eliminated, but that would require transmission lines that have zero resistance. In a world that has a global interest in not wasting energy, the reduction or elimination of this unwanted thermal energy would be a significant achievement. Is this possible?

The Resistance of Mercury

In 1911, Heike Kamerlingh Onnes of Leiden University, a Dutch physicist, was looking at the temperature dependence of the resistance of the element mercury. He cooled the sample of mercury and noticed the familiar behavior of a linear dependence of resistance on temperature; as the temperature decreased, the resistance decreased. Kamerlingh Onnes continued to cool the sample of mercury, using liquid helium. As the temperature approached 4.2 K (-269.2°C), the resistance abruptly went to zero ([link](#)). This temperature is known as the **critical temperature** T_c for mercury. The sample of mercury entered into a phase where the resistance was absolutely zero. This phenomenon is known as **superconductivity**. (*Note:* If you connect the leads of a three-digit ohmmeter across a conductor, the reading commonly shows up as $0.00\ \Omega$. The resistance of the conductor is not actually zero, it is less than $0.01\ \Omega$.) There are various methods to measure very small resistances, such as the four-point method, but an ohmmeter is not an acceptable method to use for testing resistance in superconductivity.



The resistance of a sample of mercury is zero at very low temperatures—it is a superconductor up to the temperature of about 4.2 K.

Above that critical temperature, its resistance makes a sudden jump and then increases nearly linearly with temperature.

Other Superconducting Materials

As research continued, several other materials were found to enter a superconducting phase, when the temperature reached near absolute zero. In 1941, an alloy of niobium-nitride was found that could become superconducting at $T_c = 16$ K (-257°C) and in 1953, vanadium-silicon was found to become superconductive at $T_c = 17.5$ K (-255.7°C). The temperatures for the transition into superconductivity were slowly creeping higher. Strangely, many materials that make good conductors, such as copper, silver, and gold, do not exhibit superconductivity. Imagine the energy savings if transmission lines for electric power-generating stations could be made to be superconducting at temperatures near room temperature! A resistance of zero ohms means no I^2R losses and a great boost to reducing energy consumption. The problem

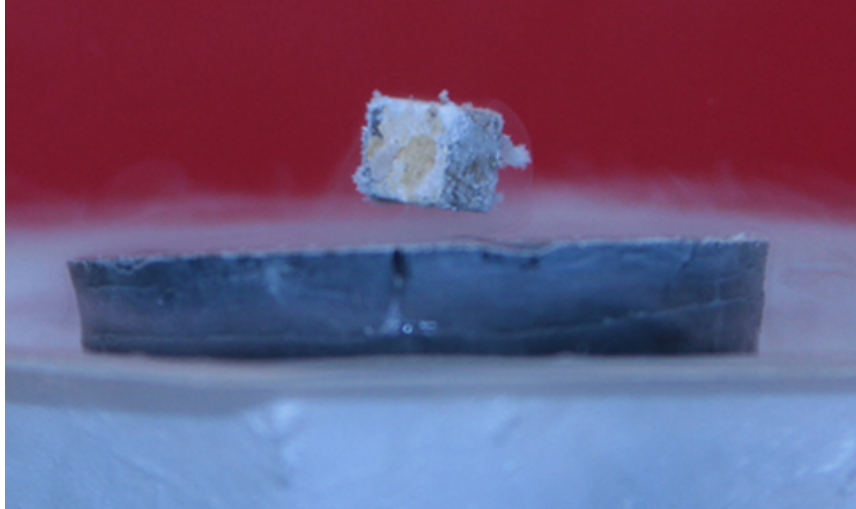
is that $T_c = 17.5 \text{ K}$ is still very cold and in the range of liquid helium temperatures. At this temperature, it is not cost effective to transmit electrical energy because of the cooling requirements.

A large jump was seen in 1986, when a team of researchers, headed by Dr. Ching Wu Chu of Houston University, fabricated a brittle, ceramic compound with a transition temperature of $T_c = 92 \text{ K}$ (-181°C). The ceramic material, composed of yttrium barium copper oxide (YBCO), was an insulator at room temperature. Although this temperature still seems quite cold, it is near the boiling point of liquid nitrogen, a liquid commonly used in refrigeration. You may have noticed refrigerated trucks traveling down the highway labeled as “Liquid Nitrogen Cooled.”

YBCO ceramic is a material that could be useful for transmitting electrical energy because the cost saving of reducing the I^2R losses are larger than the cost of cooling the superconducting cable, making it financially feasible. There were and are many engineering problems to overcome. For example, unlike traditional electrical cables, which are flexible and have a decent tensile strength, ceramics are brittle and would break rather than stretch under pressure. Processes that are rather simple with traditional cables, such as making connections, become difficult when working with ceramics. The problems are difficult and complex, and material scientists and engineers are coming up with innovative solutions.

An interesting consequence of the resistance going to zero is that once a current is established in a superconductor, it persists without an applied voltage source. Current loops in a superconductor have been set up and the current loops have been observed to persist for years without decaying.

Zero resistance is not the only interesting phenomenon that occurs as the materials reach their transition temperatures. A second effect is the exclusion of magnetic fields. This is known as the **Meissner effect** ([\[link\]](#)). A light, permanent magnet placed over a superconducting sample will levitate in a stable position above the superconductor. High-speed trains have been developed that levitate on strong superconducting magnets, eliminating the friction normally experienced between the train and the tracks. In Japan, the Yamanashi Maglev test line opened on April 3, 1997. In April 2015, the MLX01 test vehicle attained a speed of 374 mph (603 km/h).



A small, strong magnet levitates over a superconductor cooled to liquid nitrogen temperature. The magnet levitates because the superconductor excludes magnetic fields. (credit: Joseph J. Trout)

[\[link\]](#) shows a select list of elements, compounds, and high-temperature superconductors, along with the critical temperatures for which they become superconducting. Each section is sorted from the highest critical temperature to the lowest. Also listed is the critical magnetic field for some of the materials. This is the strength of the magnetic field that destroys superconductivity. Finally, the type of the superconductor is listed.

There are two types of superconductors. There are 30 pure metals that exhibit zero resistivity below their critical temperature and exhibit the Meissner effect, the property of excluding magnetic fields from the interior of the superconductor while the superconductor is at a temperature below the critical temperature. These metals are called Type I superconductors. The superconductivity exists only below their critical temperatures and below a critical magnetic field strength. Type I superconductors are well described by the BCS theory (described next). Type I superconductors have limited practical applications because the strength of the critical magnetic field needed to destroy the superconductivity is quite low.

Type II superconductors are found to have much higher critical magnetic fields and therefore can carry much higher current densities while remaining in the superconducting state. A collection of various ceramics containing barium-copper-oxide have much higher critical temperatures for the transition into a superconducting

state. Superconducting materials that belong to this subcategory of the Type II superconductors are often categorized as high-temperature superconductors.

Introduction to BCS Theory

Type I superconductors, along with some Type II superconductors can be modeled using the BCS theory, proposed by John Bardeen, Leon Cooper, and Robert Schrieffer. Although the theory is beyond the scope of this chapter, a short summary of the theory is provided here. (More detail is provided in [Condensed Matter Physics](#).) The theory considers pairs of electrons and how they are coupled together through lattice-vibration interactions. Through the interactions with the crystalline lattice, electrons near the Fermi energy level feel a small attractive force and form pairs (Cooper pairs), and the coupling is known as a phonon interaction. Single electrons are fermions, which are particles that obey the Pauli exclusion principle. The Pauli exclusion principle in quantum mechanics states that two identical fermions (particles with half-integer spin) cannot occupy the same quantum state simultaneously. Each electron has four quantum numbers (n, l, m_l, m_s). The principal quantum number (n) describes the energy of the electron, the orbital angular momentum quantum number (l) indicates the most probable distance from the nucleus, the magnetic quantum number (m_l) describes the energy levels in the subshell, and the electron spin quantum number (m_s) describes the orientation of the spin of the electron, either up or down. As the material enters a superconducting state, pairs of electrons act more like bosons, which can condense into the same energy level and need not obey the Pauli exclusion principle. The electron pairs have a slightly lower energy and leave an energy gap above them on the order of 0.001 eV. This energy gap inhibits collision interactions that lead to ordinary resistivity. When the material is below the critical temperature, the thermal energy is less than the band gap and the material exhibits zero resistivity.

Material	Symbol or Formula	Critical Temperature T_c (K)	Critical Magnetic Field H_c (T)	Type
Elements				

Material	Symbol or Formula	Critical Temperature T_c (K)	Critical Magnetic Field H_c (T)	Type
Lead	Pb	7.19	0.08	I
Lanthanum	La	$(\alpha) 4.90 - (\beta) 6.30$		I
Tantalum	Ta	4.48	0.09	I
Mercury	Hg	$(\alpha) 4.15 - (\beta) 3.95$	0.04	I
Tin	Sn	3.72	0.03	I
Indium	In	3.40	0.03	I
Thallium	Tl	2.39	0.03	I
Rhenium	Re	2.40	0.03	I
Thorium	Th	1.37	0.013	I
Protactinium	Pa	1.40		I
Aluminum	Al	1.20	0.01	I
Gallium	Ga	1.10	0.005	I
Zinc	Zn	0.86	0.014	I
Titanium	Ti	0.39	0.01	I
Uranium	U	$(\alpha) 0.68 - (\beta) 1.80$		I
Cadmium	Cd	11.4	4.00	I
Compounds				

Material	Symbol or Formula	Critical Temperature T_c (K)	Critical Magnetic Field H_c (T)	Type
Niobium-germanium	Nb_3Ge	23.20	37.00	II
Niobium-tin	Nb_3Sn	18.30	30.00	II
Niobium-nitride	NbN	16.00		II
Niobium-titanium	$NbTi$	10.00	15.00	II
High-Temperature Oxides				
	$HgBa_2CaCu_2O_8$	134.00		II
	$Tl_2Ba_2Ca_2Cu_3O_{10}$	125.00		II
	$YBa_2Cu_3O_7$	92.00	120.00	II

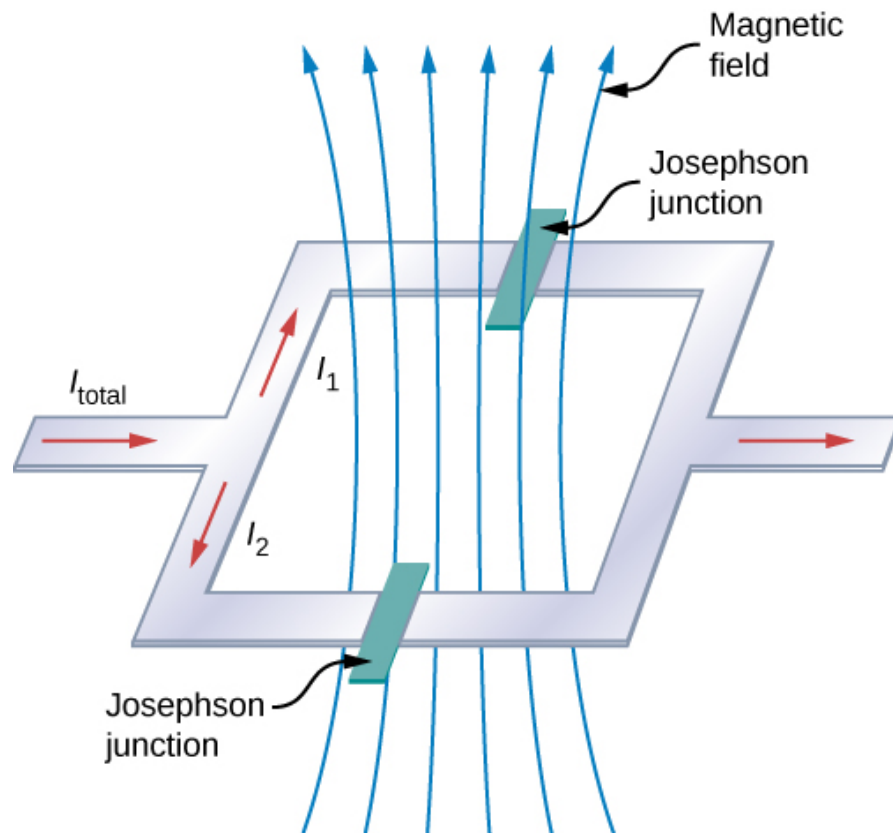
Superconductor Critical Temperatures

Applications of Superconductors

Superconductors can be used to make superconducting magnets. These magnets are 10 times stronger than the strongest electromagnets. These magnets are currently in use in magnetic resonance imaging (MRI), which produces high-quality images of the body interior without dangerous radiation.

Another interesting application of superconductivity is the **SQUID** (superconducting quantum interference device). A SQUID is a very sensitive magnetometer used to measure extremely subtle magnetic fields. The operation of the SQUID is based on superconducting loops containing Josephson junctions. A **Josephson junction** is the result of a theoretical prediction made by B. D. Josephson in an article published in 1962. In the article, Josephson described how a supercurrent can flow between two pieces of superconductor separated by a thin layer of insulator. This phenomenon is now called the Josephson effect. The SQUID consists of a superconducting current

loop containing two Josephson junctions, as shown in [\[link\]](#). When the loop is placed in even a very weak magnetic field, there is an interference effect that depends on the strength of the magnetic field.



The SQUID (superconducting quantum interference device) uses a superconducting current loop and two Josephson junctions to detect magnetic fields as low as 10^{-14} T (Earth's magnet field is on the order of 0.3×10^{-5} T).

Superconductivity is a fascinating and useful phenomenon. At critical temperatures near the boiling point of liquid nitrogen, superconductivity has special applications in MRIs, particle accelerators, and high-speed trains. Will we reach a state where we can have materials enter the superconducting phase at near room temperatures? It seems a long way off, but if scientists in 1911 were asked if we would reach liquid-nitrogen temperatures with a ceramic, they might have thought it implausible.

Summary

- Superconductivity is a phenomenon that occurs in some materials when cooled to very low critical temperatures, resulting in a resistance of exactly zero and the expulsion of all magnetic fields.
- Materials that are normally good conductors (such as copper, gold, and silver) do not experience superconductivity.
- Superconductivity was first observed in mercury by Heike Kamerlingh Onnes in 1911. In 1986, Dr. Ching Wu Chu of Houston University fabricated a brittle, ceramic compound with a critical temperature close to the temperature of liquid nitrogen.
- Superconductivity can be used in the manufacture of superconducting magnets for use in MRIs and high-speed, levitated trains.

Key Equations

Average electrical current	$I_{\text{ave}} = \frac{\Delta Q}{\Delta t}$
Definition of an ampere	$1 \text{ A} = 1 \text{ C/s}$
Electrical current	$I = \frac{dQ}{dt}$
Drift velocity	$v_d = \frac{I}{nqA}$
Current density	$I = \iint_{\text{area}} \vec{\mathbf{J}} \cdot d\vec{\mathbf{A}}$
Resistivity	$\rho = \frac{E}{J}$
Common expression of Ohm's law	$V = IR$
Resistivity as a function of temperature	$\rho = \rho_0 [1 + \alpha (T - T_0)]$
Definition of resistance	$R \equiv \frac{V}{I}$

Resistance of a cylinder of material	$R = \rho \frac{L}{A}$
Temperature dependence of resistance	$R = R_0 (1 + \alpha \Delta T)$
Electric power	$P = IV$
Power dissipated by a resistor	$P = I^2 R = \frac{V^2}{R}$

Conceptual Questions

Exercise:

Problem:

What requirement for superconductivity makes current superconducting devices expensive to operate?

Solution:

Very low temperatures necessitate refrigeration. Some materials require liquid nitrogen to cool them below their critical temperatures. Other materials may need liquid helium, which is even more costly.

Exercise:

Problem:

Name two applications for superconductivity listed in this section and explain how superconductivity is used in the application. Can you think of a use for superconductivity that is not listed?

Problems

Exercise:

Problem:

Consider a power plant is located 60 km away from a residential area uses 0-gauge ($A = 42.40 \text{ mm}^2$) wire of copper to transmit power at a current of $I = 100.00 \text{ A}$. How much more power is dissipated in the copper wires than it would be in superconducting wires?

Solution:

$$R_{\text{copper}} = 23.77 \, \Omega$$

$$P = 2.377 \times 10^5 \text{ W}$$

Exercise:

Problem:

A wire is drawn through a die, stretching it to four times its original length. By what factor does its resistance increase?

Exercise:

Problem:

Digital medical thermometers determine temperature by measuring the resistance of a semiconductor device called a thermistor (which has $\alpha = -0.06/^\circ\text{C}$) when it is at the same temperature as the patient. What is a patient's temperature if the thermistor's resistance at that temperature is 82.0% of its value at 37°C (normal body temperature)?

Solution:

$$R = R_0 (1 + \alpha (T - T_0))$$

$$0.82R_0 = R_0 (1 + \alpha (T - T_0)), \quad 0.82 = 1 - 0.06 (T - 37^\circ\text{C}), \quad T = 40^\circ\text{C}$$

Exercise:

Problem:

Electrical power generators are sometimes “load tested” by passing current through a large vat of water. A similar method can be used to test the heat output of a resistor. A $R = 30 \, \Omega$ resistor is connected to a 9.0-V battery and the resistor leads are waterproofed and the resistor is placed in 1.0 kg of room temperature water ($T = 20^\circ\text{C}$). Current runs through the resistor for 20 minutes. Assuming all the electrical energy dissipated by the resistor is converted to heat, what is the final temperature of the water?

Exercise:

Problem:

A 12-gauge gold wire has a length of 1 meter. (a) What would be the length of a silver 12-gauge wire with the same resistance? (b) What are their respective resistances at the temperature of boiling water?

Solution:

$$\text{a. } R_{\text{Au}} = R_{\text{Ag}}, \quad \rho_{\text{Au}} \frac{L_{\text{Au}}}{A_{\text{Au}}} = \rho_{\text{Ag}} \frac{L_{\text{Ag}}}{A_{\text{Ag}}}, \quad L_{\text{Ag}} = 1.53 \text{ m};$$

$$\text{b. } R_{\text{Au}, 20^\circ \text{C}} = 0.0074 \, \Omega, \quad R_{\text{Au}, 100^\circ \text{C}} = 0.0094 \, \Omega, \quad R_{\text{Ag}, 100^\circ \text{C}} = 0.0096 \, \Omega$$

Exercise:**Problem:**

What is the change in temperature required to decrease the resistance for a carbon resistor by 10%?

Additional Problems**Exercise:****Problem:**

A coaxial cable consists of an inner conductor with radius $r_i = 0.25 \text{ cm}$ and an outer radius of $r_o = 0.5 \text{ cm}$ and has a length of 10 meters. Plastic, with a resistivity of $\rho = 2.00 \times 10^{13} \, \Omega \cdot \text{m}$, separates the two conductors. What is the resistance of the cable?

Solution:

$$dR = \frac{\rho}{2\pi r L} dr$$

$$R = \frac{\rho}{2\pi L} \ln \frac{r_o}{r_i}$$

$$R = 2.21 \times 10^{11} \, \Omega$$

Exercise:**Problem:**

A 10.00-meter long wire cable that is made of copper has a resistance of 0.051 ohms. (a) What is the weight if the wire was made of copper? (b) What is the weight of a 10.00-meter-long wire of the same gauge made of aluminum? (c) What is the resistance of the aluminum wire? The density of copper is 8960 kg/m^3 and the density of aluminum is 2760 kg/m^3 .

Exercise:

Problem:

A nichrome rod that is 3.00 mm long with a cross-sectional area of 1.00 mm^2 is used for a digital thermometer. (a) What is the resistance at room temperature? (b) What is the resistance at body temperature?

Solution:

a.

$$R_0 = 0.003 \, \Omega; \text{ b.}$$

$$T_c = 37.0 \, ^\circ\text{C}$$

$$R = 0.00302 \, \Omega$$

Exercise:**Problem:**

The temperature in Philadelphia, PA can vary between $68.00 \, ^\circ\text{F}$ and $100.00 \, ^\circ\text{F}$ in one summer day. By what percentage will an aluminum wire's resistance change during the day?

Exercise:**Problem:**

When 100.0 V is applied across a 5-gauge (diameter 4.621 mm) wire that is 10 m long, the magnitude of the current density is $2.0 \times 10^8 \text{ A/m}^2$. What is the resistivity of the wire?

Solution:

$$\rho = 5.00 \times 10^{-8} \, \Omega \cdot \text{m}$$

Exercise:**Problem:**

A wire with a resistance of $5.0 \, \Omega$ is drawn out through a die so that its new length is twice times its original length. Find the resistance of the longer wire. You may assume that the resistivity and density of the material are unchanged.

Exercise:**Problem:**

What is the resistivity of a wire of 5-gauge wire ($A = 16.8 \times 10^{-6} \text{ m}^2$), 5.00 m length, and $5.10 \text{ m} \, \Omega$ resistance?

Solution:

$$\rho = 1.71 \times 10^{-8} \Omega \cdot \text{m}$$

Exercise:**Problem:**

Coils are often used in electrical and electronic circuits. Consider a coil which is formed by winding 1000 turns of insulated 20-gauge copper wire (area 0.52 mm^2) in a single layer on a cylindrical non-conducting core of radius 2.0 mm. What is the resistance of the coil? Neglect the thickness of the insulation.

Exercise:**Problem:**

Currents of approximately 0.06 A can be potentially fatal. Currents in that range can make the heart fibrillate (beat in an uncontrolled manner). The resistance of a dry human body can be approximately $100 \text{ k}\Omega$. (a) What voltage can cause 0.06 A through a dry human body? (b) When a human body is wet, the resistance can fall to 100Ω . What voltage can cause harm to a wet body?

Solution:

$$\text{a. } V = 6000 \text{ V; b. } V = 6 \text{ V}$$

Exercise:**Problem:**

A 20.00-ohm, 5.00-watt resistor is placed in series with a power supply. (a) What is the maximum voltage that can be applied to the resistor without harming the resistor? (b) What would be the current through the resistor?

Exercise:**Problem:**

A battery with an emf of 24.00 V delivers a constant current of 2.00 mA to an appliance. How much work does the battery do in three minutes?

Solution:

$$P = \frac{W}{t}, \quad W = 8.64 \text{ J}$$

Exercise:

Problem:

A 12.00-V battery has an internal resistance of a tenth of an ohm. (a) What is the current if the battery terminals are momentarily shorted together? (b) What is the terminal voltage if the battery delivers 0.25 amps to a circuit?

Challenge Problems**Exercise:****Problem:**

A 10-gauge copper wire has a cross-sectional area $A = 5.26 \text{ mm}^2$ and carries a current of $I = 5.00 \text{ A}$. The density of copper is $\rho = 8.95 \text{ g/cm}^3$. One mole of copper atoms (6.02×10^{23} atoms) has a mass of approximately 63.50 g. What is the magnitude of the drift velocity of the electrons, assuming that each copper atom contributes one free electron to the current?

Solution:

$$V = 7.09 \text{ cm}^3$$

$$n = 8.49 \times 10^{28} \frac{\text{electrons}}{\text{m}^3}$$

$$v_d = 7.00 \times 10^{-5} \frac{\text{m}}{\text{s}}$$

Exercise:**Problem:**

The current through a 12-gauge wire is given as

$I(t) = (5.00 \text{ A}) \sin(2\pi 60 \text{ Hz } t)$. What is the current density at time 15.00 ms?

Exercise:**Problem:**

A particle accelerator produces a beam with a radius of 1.25 mm with a current of 2.00 mA. Each proton has a kinetic energy of 10.00 MeV. (a) What is the velocity of the protons? (b) What is the number (n) of protons per unit volume? (b) How many electrons pass a cross sectional area each second?

Solution:

a. $4.38 \times 10^7 \text{ m/s}$ b. $v = 5.81 \times 10^{13} \frac{\text{protons}}{\text{m}^3}$ c. $1.25 \frac{\text{electrons}}{\text{m}^3}$

Exercise:

Problem:

In this chapter, most examples and problems involved direct current (DC). DC circuits have the current flowing in one direction, from positive to negative. When the current was changing, it was changed linearly from $I = -I_{\text{max}}$ to $I = +I_{\text{max}}$ and the voltage changed linearly from $V = -V_{\text{max}}$ to $V = +V_{\text{max}}$, where $V_{\text{max}} = I_{\text{max}}R$. Suppose a voltage source is placed in series with a resistor of $R = 10 \Omega$ that supplied a current that alternated as a sine wave, for example, $I(t) = (3.00 \text{ A}) \sin\left(\frac{2\pi}{4.00 \text{ s}}t\right)$. (a) What would a graph of the voltage drop across the resistor $V(t)$ versus time look like? (b) What would a plot of $V(t)$ versus $I(t)$ for one period look like? (*Hint:* If you are not sure, try plotting $V(t)$ versus $I(t)$ using a spreadsheet.)

Exercise:

Problem:

A current of $I = 25 \text{ A}$ is drawn from a 100-V battery for 30 seconds. By how much is the chemical energy reduced?

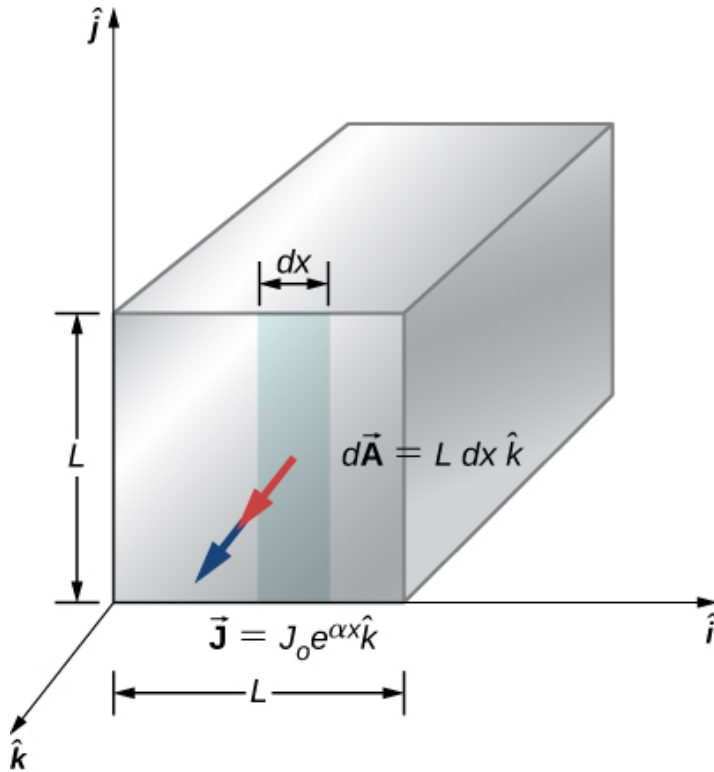
Solution:

$$E = 75 \text{ kJ}$$

Exercise:

Problem:

Consider a square rod of material with sides of length $L = 3.00 \text{ cm}$ with a current density of $\vec{J} = J_0 e^{\alpha x} \hat{k} = \left(0.35 \frac{\text{A}}{\text{m}^2}\right) e^{(2.1 \times 10^{-3} \text{ m}^{-1})x} \hat{k}$ as shown below. Find the current that passes through the face of the rod.



Exercise:

Problem:

A resistor of an unknown resistance is placed in an insulated container filled with 0.75 kg of water. A voltage source is connected in series with the resistor and a current of 1.2 amps flows through the resistor for 10 minutes. During this time, the temperature of the water is measured and the temperature change during this time is $\Delta T = 10.00^\circ \text{C}$. (a) What is the resistance of the resistor? (b) What is the voltage supplied by the power supply?

Solution:

a. $P = 52 \text{ W}$; b. $V = 43.54 \text{ V}$
 $R = 36 \Omega$

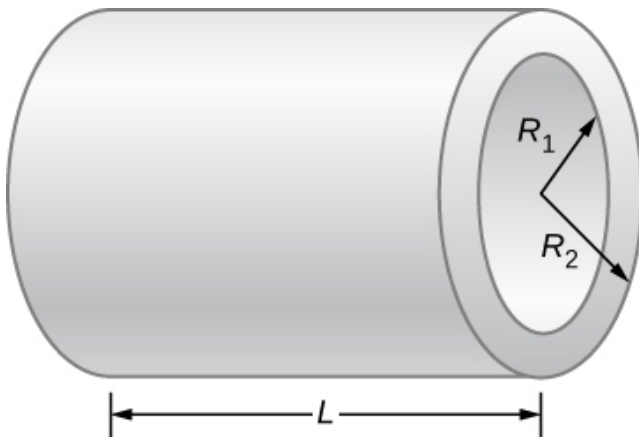
Exercise:

Problem:

The charge that flows through a point in a wire as a function of time is modeled as $q(t) = q_0 e^{-t/T} = 10.0 \text{ C} e^{-t/5 \text{ s}}$. (a) What is the initial current through the wire at time $t = 0.00 \text{ s}$? (b) Find the current at time $t = \frac{1}{2}T$. (c) At what time t will the current be reduced by one-half $I = \frac{1}{2}I_0$?

Exercise:**Problem:**

Consider a resistor made from a hollow cylinder of carbon as shown below. The inner radius of the cylinder is $R_i = 0.20 \text{ mm}$ and the outer radius is $R_0 = 0.30 \text{ mm}$. The length of the resistor is $L = 0.90 \text{ mm}$. The resistivity of the carbon is $\rho = 3.5 \times 10^{-5} \Omega \cdot \text{m}$. (a) Prove that the resistance perpendicular from the axis is $R = \frac{\rho}{2\pi L} \ln\left(\frac{R_0}{R_i}\right)$. (b) What is the resistance?

**Solution:**

a. $R = \frac{\rho}{2\pi L} \ln\left(\frac{R_0}{R_i}\right)$; b. $R = 2.5 \text{ m}\Omega$

Exercise:**Problem:**

What is the current through a cylindrical wire of radius $R = 0.1 \text{ mm}$ if the current density is $J = \frac{J_0}{R} r$, where $J_0 = 32000 \frac{\text{A}}{\text{m}^2}$?

Exercise:

Problem:

A student uses a 100.00-W, 115.00-V radiant heater to heat the student's dorm room, during the hours between sunset and sunrise, 6:00 p.m. to 7:00 a.m. (a) What current does the heater operate at? (b) How many electrons move through the heater? (c) What is the resistance of the heater? (d) How much heat was added to the dorm room?

Solution:

- (a) 0.870 A
- (b) #electrons = 2.54×10^{23} electrons
- (c) 132 ohms
- (d) $q = 4.68 \times 10^6$ J

Exercise:**Problem:**

A 12-V car battery is used to power a 20.00-W, 12.00-V lamp during the physics club camping trip/star party. The cable to the lamp is 2.00 meters long, 14-gauge copper wire with a charge density of $n = 9.50 \times 10^{28} \text{m}^{-3}$. (a) What is the current draw by the lamp? (b) How long would it take an electron to get from the battery to the lamp?

Exercise:**Problem:**

A physics student uses a 115.00-V immersion heater to heat 400.00 grams (almost two cups) of water for herbal tea. During the two minutes it takes the water to heat, the physics student becomes bored and decides to figure out the resistance of the heater. The student starts with the assumption that the water is initially at the temperature of the room $T_i = 25.00^\circ \text{C}$ and reaches $T_f = 100.00^\circ \text{C}$. The specific heat of the water is $c = 4180 \frac{\text{J}}{\text{kg} \cdot \text{K}}$. What is the resistance of the heater?

Solution:

$$P = 1045 \text{ W}, \quad P = \frac{V^2}{R}, \quad R = 12.27 \Omega$$

Glossary

critical temperature

temperature at which a material reaches superconductivity

Josephson junction

junction of two pieces of superconducting material separated by a thin layer of insulating material, which can carry a supercurrent

Meissner effect

phenomenon that occurs in a superconducting material where all magnetic fields are expelled

SQUID

(Superconducting Quantum Interference Device) device that is a very sensitive magnetometer, used to measure extremely subtle magnetic fields

superconductivity

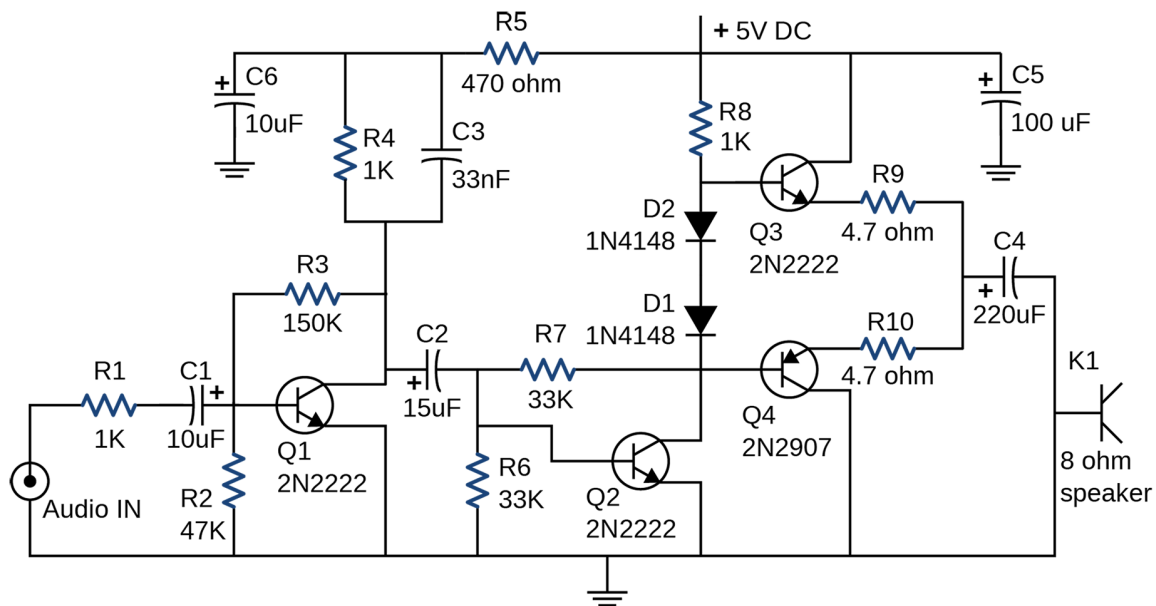
phenomenon that occurs in some materials where the resistance goes to exactly zero and all magnetic fields are expelled, which occurs dramatically at some low critical temperature (T_C)

Introduction

class="introduction"

This circuit shown is used to amplify small signals and power the earbud speakers attached to a cellular phone. This circuit's components include resistors, capacitors, and diodes, all of which have been covered in previous chapters, as well as transistors, which are semi-conducting devices covered in [Condensed Matter Physics](#).
Circuits

using
similar
components
are found in
all types of
equipment
and
appliances
you
encounter in
everyday
life, such as
alarm
clocks,
televisions,
computers,
and
refrigerators



In the preceding few chapters, we discussed electric components, including capacitors, resistors, and diodes. In this chapter, we use these electric components in circuits. A circuit is a collection of electrical components connected to accomplish a specific task. [\[link\]](#) shows an amplifier circuit, which takes a small-amplitude signal and amplifies it to power the speakers in earbuds. Although the circuit looks complex, it actually consists of a set of series, parallel, and series-parallel circuits. The second section of this chapter covers the analysis of series and parallel circuits that consist of resistors. Later in this chapter, we introduce the basic equations and techniques to analyze any circuit, including those that are not reducible through simplifying parallel and series elements. But first, we need to understand how to power a circuit.

Electromotive Force

By the end of the section, you will be able to:

- Describe the electromotive force (emf) and the internal resistance of a battery
- Explain the basic operation of a battery

If you forget to turn off your car lights, they slowly dim as the battery runs down. Why don't they suddenly blink off when the battery's energy is gone? Their gradual dimming implies that the battery output voltage decreases as the battery is depleted. The reason for the decrease in output voltage for depleted batteries is that all voltage sources have two fundamental parts—a source of electrical energy and an internal resistance. In this section, we examine the energy source and the internal resistance.

Introduction to Electromotive Force

Voltage has many sources, a few of which are shown in [\[link\]](#). All such devices create a **potential difference** and can supply current if connected to a circuit. A special type of potential difference is known as **electromotive force (emf)**. The emf is not a force at all, but the term 'electromotive force' is used for historical reasons. It was coined by Alessandro Volta in the 1800s, when he invented the first battery, also known as the voltaic pile. Because the electromotive force is not a force, it is common to refer to these sources simply as sources of emf (pronounced as the letters "ee-em-eff"), instead of sources of electromotive force.



(a)



(b)



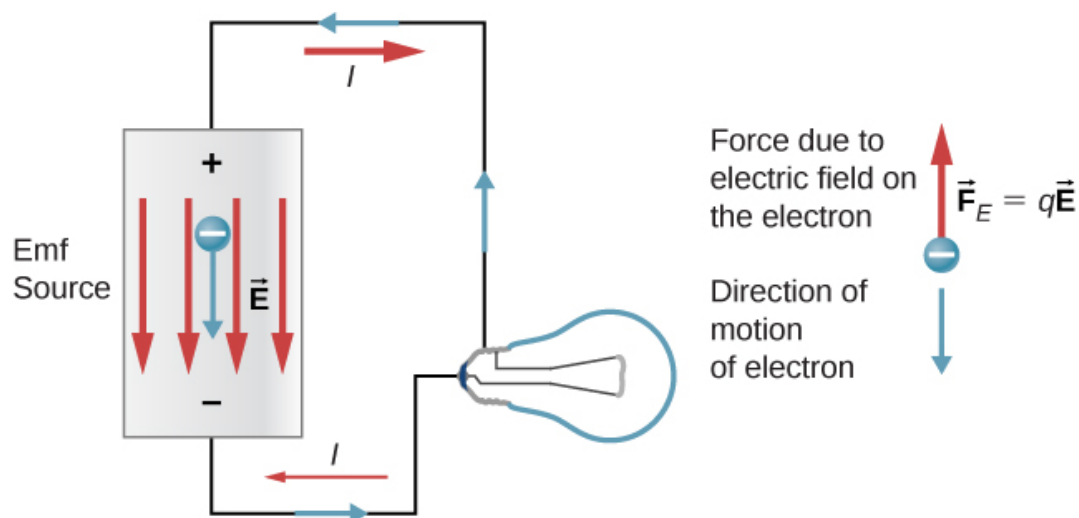
(c)



(d)

A variety of voltage sources. (a) The Brazos Wind Farm in Fluvanna, Texas; (b) the Krasnoyarsk Dam in Russia; (c) a solar farm; (d) a group of nickel metal hydride batteries. The voltage output of each device depends on its construction and load. The voltage output equals emf only if there is no load. (credit a: modification of work by Stig Nygaard; credit b: modification of work by "vadimpl"/Wikimedia Commons; credit c: modification of work by "The tdog"/Wikimedia Commons; credit d: modification of work by "Itrados"/Wikimedia Commons)

If the electromotive force is not a force at all, then what is the emf and what is a source of emf? To answer these questions, consider a simple circuit of a 12-V lamp attached to a 12-V battery, as shown in [\[link\]](#). The battery can be modeled as a two-terminal device that keeps one terminal at a higher electric potential than the second terminal. The higher electric potential is sometimes called the positive terminal and is labeled with a plus sign. The lower-potential terminal is sometimes called the negative terminal and labeled with a minus sign. This is the source of the emf.



A source of emf maintains one terminal at a higher electric potential than the other terminal, acting as a source of current in a circuit.

When the emf source is not connected to the lamp, there is no net flow of charge within the emf source. Once the battery is connected to the lamp, charges flow from one terminal of the battery, through the lamp (causing the lamp to light), and back to the other terminal of the battery. If we consider positive (conventional) current flow, positive charges leave the positive terminal, travel through the lamp, and enter the negative terminal.

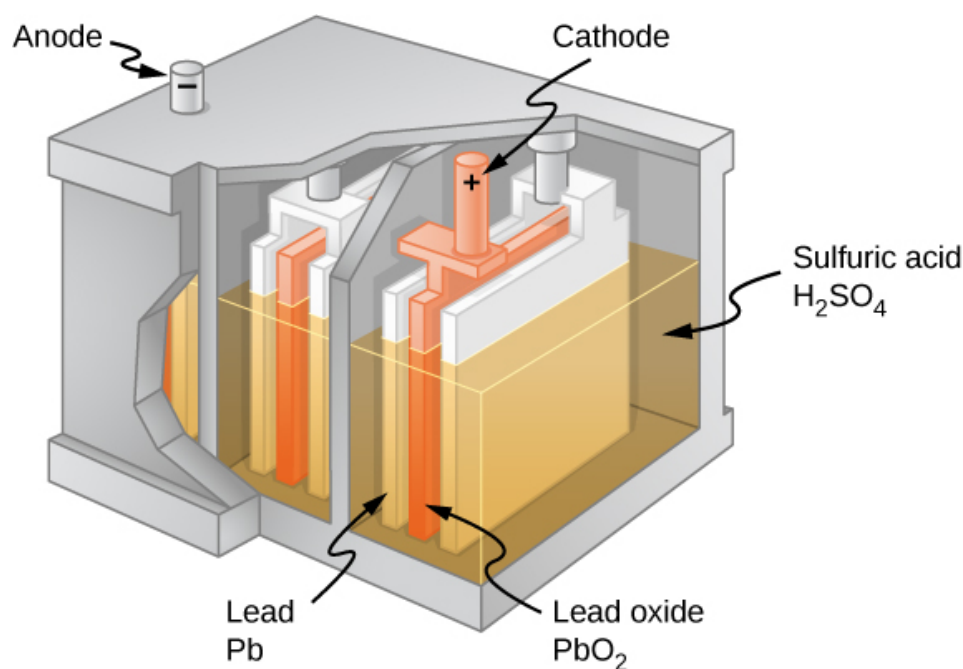
Positive current flow is useful for most of the circuit analysis in this chapter, but in metallic wires and resistors, electrons contribute the most to current, flowing in the opposite direction of positive current flow. Therefore, it is more realistic to consider the movement of electrons for the analysis of the circuit in [\[link\]](#). The electrons leave the negative terminal, travel through the lamp, and return to the positive terminal. In order for the emf source to maintain the potential difference between the two terminals, negative charges (electrons) must be moved from the positive terminal to the negative terminal. The emf source acts as a charge pump, moving negative charges from the positive terminal to the negative terminal to maintain the potential difference. This increases the potential energy of the charges and, therefore, the electric potential of the charges.

The force on the negative charge from the electric field is in the opposite direction of the electric field, as shown in [\[link\]](#). In order for the negative charges to be moved to the negative terminal, work must be done on the negative charges. This requires energy, which comes from chemical reactions in the battery. The potential is kept high on the positive terminal and low on the negative terminal to maintain the potential difference between the two terminals. The emf is equal to the work done on the charge per unit charge ($\mathcal{E} = \frac{dW}{dq}$) when there is no current flowing. Since the unit for work is the joule and the unit for charge is the coulomb, the unit for emf is the volt ($1 \text{ V} = 1 \text{ J/C}$).

The **terminal voltage** V_{terminal} of a battery is voltage measured across the terminals of the battery. An ideal battery is an emf source that maintains a constant terminal voltage, independent of the current between the two terminals. An ideal battery has no internal resistance, and the terminal voltage is equal to the emf of the battery. In the next section, we will show that a real battery does have internal resistance and the terminal voltage is always less than the emf of the battery.

The Origin of Battery Potential

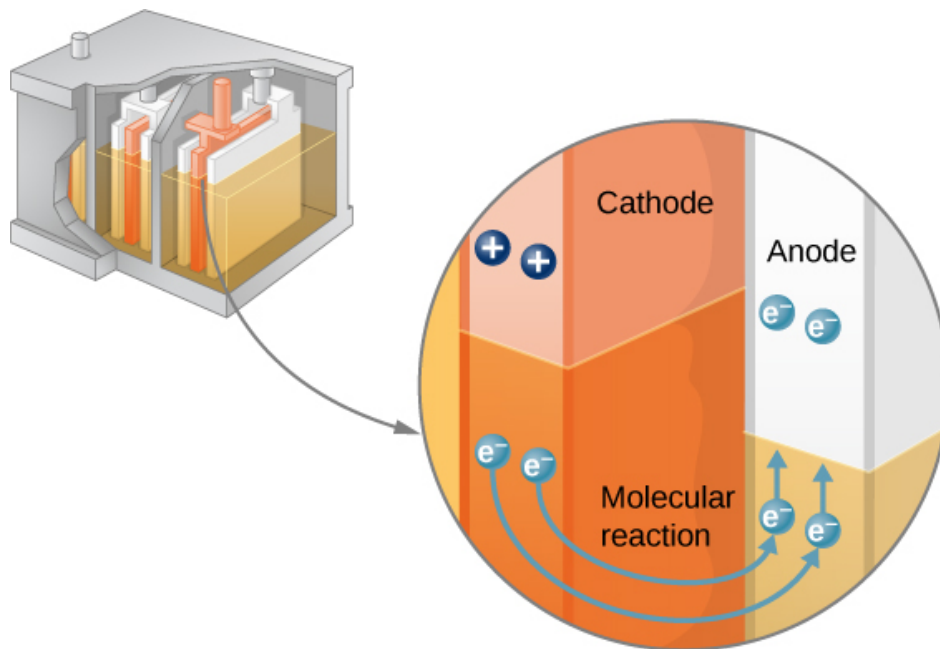
The combination of chemicals and the makeup of the terminals in a battery determine its emf. The lead acid battery used in cars and other vehicles is one of the most common combinations of chemicals. [\[link\]](#) shows a single cell (one of six) of this battery. The cathode (positive) terminal of the cell is connected to a lead oxide plate, whereas the anode (negative) terminal is connected to a lead plate. Both plates are immersed in sulfuric acid, the electrolyte for the system.



Chemical reactions in a lead-acid cell separate charge, sending negative charge to the anode, which is connected to the lead plates. The lead oxide plates are connected to the positive or cathode terminal of the cell. Sulfuric acid conducts the charge, as well as participates in the chemical reaction.

Knowing a little about how the chemicals in a lead-acid battery interact helps in understanding the potential created by the battery. [\[link\]](#) shows the result of a single chemical reaction. Two electrons are placed on the anode, making it negative, provided that the cathode supplies two electrons. This leaves the cathode positively charged, because it has lost two electrons. In short, a separation of charge has been driven by a chemical reaction.

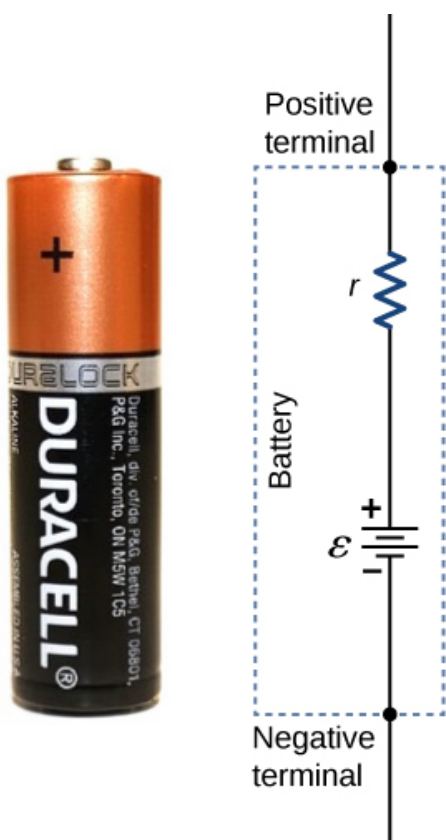
Note that the reaction does not take place unless there is a complete circuit to allow two electrons to be supplied to the cathode. Under many circumstances, these electrons come from the anode, flow through a resistance, and return to the cathode. Note also that since the chemical reactions involve substances with resistance, it is not possible to create the emf without an internal resistance.



In a lead-acid battery, two electrons are forced onto the anode of a cell, and two electrons are removed from the cathode of the cell. The chemical reaction in a lead-acid battery places two electrons on the anode and removes two from the cathode. It requires a closed circuit to proceed, since the two electrons must be supplied to the cathode.

Internal Resistance and Terminal Voltage

The amount of resistance to the flow of current within the voltage source is called the **internal resistance**. The internal resistance r of a battery can behave in complex ways. It generally increases as a battery is depleted, due to the oxidation of the plates or the reduction of the acidity of the electrolyte. However, internal resistance may also depend on the magnitude and direction of the current through a voltage source, its temperature, and even its history. The internal resistance of rechargeable nickel-cadmium cells, for example, depends on how many times and how deeply they have been depleted. A simple model for a battery consists of an idealized emf source ε and an internal resistance r ([link](#)).



A battery can be modeled as an idealized emf (ε) with an internal resistance (r). The terminal voltage of the battery is $V_{\text{terminal}} = \varepsilon - Ir$

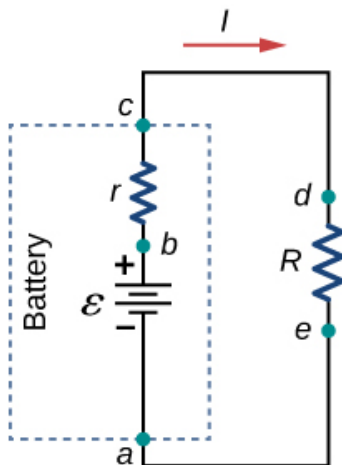
Suppose an external resistor, known as the load resistance R , is connected to a voltage source such as a battery, as in [link](#). The figure shows a model of a battery with an emf ε ,

an internal resistance r , and a load resistor R connected across its terminals. Using conventional current flow, positive charges leave the positive terminal of the battery, travel through the resistor, and return to the negative terminal of the battery. The terminal voltage of the battery depends on the emf, the internal resistance, and the current, and is equal to

Note:
Equation:

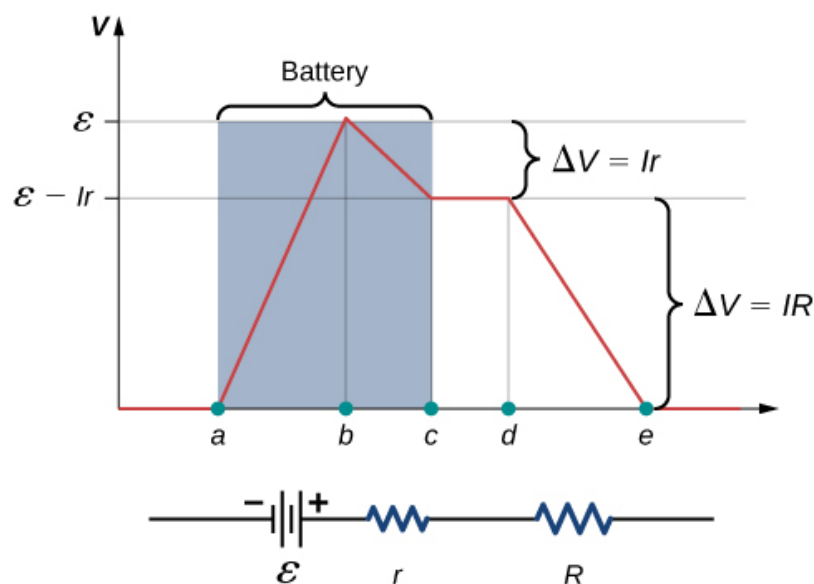
$$V_{\text{terminal}} = \varepsilon - Ir.$$

For a given emf and internal resistance, the terminal voltage decreases as the current increases due to the potential drop Ir of the internal resistance.



Schematic of a voltage source and its load resistor R . Since the internal resistance r is in series with the load, it can significantly affect the terminal voltage and the current delivered to the load.

A graph of the potential difference across each element the circuit is shown in [\[link\]](#). A current I runs through the circuit, and the potential drop across the internal resistor is equal to Ir . The terminal voltage is equal to $\varepsilon - Ir$, which is equal to the **potential drop** across the load resistor $IR = \varepsilon - Ir$. As with potential energy, it is the change in voltage that is important. When the term “voltage” is used, we assume that it is actually the change in the potential, or ΔV . However, Δ is often omitted for convenience.



A graph of the voltage through the circuit of a battery and a load resistance. The electric potential increases the emf of the battery due to the chemical reactions doing work on the charges. There is a decrease in the electric potential in the battery due to the internal resistance. The potential decreases due to the internal resistance ($-Ir$), making the terminal voltage of the battery equal to $(\varepsilon - Ir)$. The voltage then decreases by (IR) . The current is equal to $I = \frac{\varepsilon}{r+R}$.

The current through the load resistor is $I = \frac{\varepsilon}{r+R}$. We see from this expression that the smaller the internal resistance r , the greater the current the voltage source supplies to its load R . As batteries are depleted, r increases. If r becomes a significant fraction of the load resistance, then the current is significantly reduced, as the following example illustrates.

Example:

Analyzing a Circuit with a Battery and a Load

A given battery has a 12.00-V emf and an internal resistance of $0.100\ \Omega$. (a) Calculate its terminal voltage when connected to a $10.00\text{-}\Omega$ load. (b) What is the terminal voltage when connected to a $0.500\text{-}\Omega$ load? (c) What power does the $0.500\text{-}\Omega$ load dissipate? (d) If the internal resistance grows to $0.500\ \Omega$, find the current, terminal voltage, and power dissipated by a $0.500\text{-}\Omega$ load.

Strategy

The analysis above gave an expression for current when internal resistance is taken into account. Once the current is found, the terminal voltage can be calculated by using the equation $V_{\text{terminal}} = \varepsilon - Ir$. Once current is found, we can also find the power dissipated by the resistor.

Solution

- a. Entering the given values for the emf, load resistance, and internal resistance into the expression above yields

Equation:

$$I = \frac{\varepsilon}{R + r} = \frac{12.00\ \text{V}}{10.10\ \Omega} = 1.188\ \text{A}.$$

Enter the known values into the equation $V_{\text{terminal}} = \varepsilon - Ir$ to get the terminal voltage:

Equation:

$$V_{\text{terminal}} = \varepsilon - Ir = 12.00\ \text{V} - (1.188\ \text{A})(0.100\ \Omega) = 11.90\ \text{V}.$$

The terminal voltage here is only slightly lower than the emf, implying that the current drawn by this light load is not significant.

- b. Similarly, with $R_{\text{load}} = 0.500\ \Omega$, the current is

Equation:

$$I = \frac{\varepsilon}{R + r} = \frac{12.00\ \text{V}}{0.600\ \Omega} = 20.00\ \text{A}.$$

The terminal voltage is no

Equation:

$$V_{\text{terminal}} = \varepsilon - Ir = 12.00\ \text{V} - (20.00\ \text{A})(0.100\ \Omega) = 10.00\ \text{V}.$$

The terminal voltage exhibits a more significant reduction compared with emf, implying $0.500\ \Omega$ is a heavy load for this battery. A “heavy load” signifies a larger draw of current from the source but not a larger resistance.

- c. The power dissipated by the $0.500\text{-}\Omega$ load can be found using the formula $P = I^2R$. Entering the known values gives

Equation:

$$P = I^2 R = (20.0 \text{ A})^2 (0.500 \Omega) = 2.00 \times 10^2 \text{ W}.$$

Note that this power can also be obtained using the expression $\frac{V^2}{R}$ or IV , where V is the terminal voltage (10.0 V in this case).

- d. Here, the internal resistance has increased, perhaps due to the depletion of the battery, to the point where it is as great as the load resistance. As before, we first find the current by entering the known values into the expression, yielding

Equation:

$$I = \frac{\varepsilon}{R + r} = \frac{12.00 \text{ V}}{1.00 \Omega} = 12.00 \text{ A}.$$

Now the terminal voltage is

Equation:

$$V_{\text{terminal}} = \varepsilon - Ir = 12.00 \text{ V} - (12.00 \text{ A})(0.500 \Omega) = 6.00 \text{ V},$$

and the power dissipated by the load is

Equation:

$$P = I^2 R = (12.00 \text{ A})^2 (0.500 \Omega) = 72.00 \text{ W}.$$

We see that the increased internal resistance has significantly decreased the terminal voltage, current, and power delivered to a load.

Significance

The internal resistance of a battery can increase for many reasons. For example, the internal resistance of a rechargeable battery increases as the number of times the battery is recharged increases. The increased internal resistance may have two effects on the battery. First, the terminal voltage will decrease. Second, the battery may overheat due to the increased power dissipated by the internal resistance.

Note:

Exercise:

Problem:

Check Your Understanding If you place a wire directly across the two terminal of a battery, effectively shorting out the terminals, the battery will begin to get hot. Why do you suppose this happens?

Solution:

If a wire is connected across the terminals, the load resistance is close to zero, or at least considerably less than the internal resistance of the battery. Since the internal resistance is small, the current through the circuit will be large,
$$I = \frac{\varepsilon}{R+r} = \frac{\varepsilon}{0+r} = \frac{\varepsilon}{r}.$$
The large current causes a high power to be dissipated by the internal resistance ($P = I^2r$). The power is dissipated as heat.

Battery Testers

Battery testers, such as those in [\[link\]](#), use small load resistors to intentionally draw current to determine whether the terminal potential drops below an acceptable level. Although it is difficult to measure the internal resistance of a battery, battery testers can provide a measurement of the internal resistance of the battery. If internal resistance is high, the battery is weak, as evidenced by its low terminal voltage.



(a)

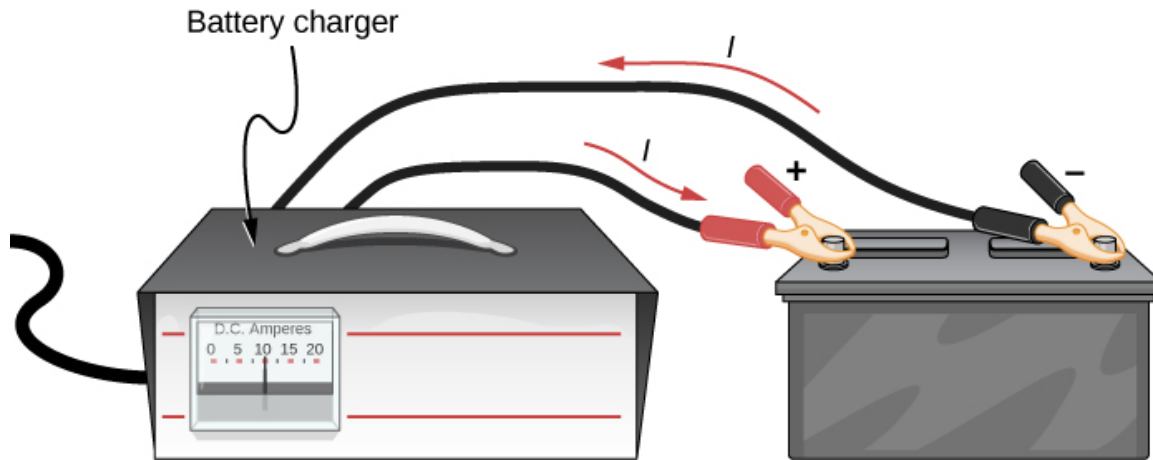


(b)

Battery testers measure terminal voltage under a load to determine the condition of a battery. (a) A US Navy electronics technician uses a battery tester to test large batteries aboard the aircraft carrier USS *Nimitz*. The battery tester she uses has a small resistance that can dissipate large amounts of power. (b) The small device shown is used on small batteries and has a digital display to indicate the acceptability of the terminal voltage. (credit a: modification of work by Jason A. Johnston; credit b: modification of work by Keith Williamson)

Some batteries can be recharged by passing a current through them in the direction opposite to the current they supply to an appliance. This is done routinely in cars and in batteries for small electrical appliances and electronic devices ([\[link\]](#)). The voltage output of the battery charger must be greater than the emf of the battery to reverse the current through it. This

causes the terminal voltage of the battery to be greater than the emf, since $V = \varepsilon - Ir$ and I is now negative.



A car battery charger reverses the normal direction of current through a battery, reversing its chemical reaction and replenishing its chemical potential.

It is important to understand the consequences of the internal resistance of emf sources, such as batteries and solar cells, but often, the analysis of circuits is done with the terminal voltage of the battery, as we have done in the previous sections. The terminal voltage is referred to as simply as V , dropping the subscript “terminal.” This is because the internal resistance of the battery is difficult to measure directly and can change over time.

Summary

- All voltage sources have two fundamental parts: a source of electrical energy that has a characteristic electromotive force (emf), and an internal resistance r . The emf is the work done per charge to keep the potential difference of a source constant. The emf is equal to the potential difference across the terminals when no current is flowing. The internal resistance r of a voltage source affects the output voltage when a current flows.
- The voltage output of a device is called its terminal voltage V_{terminal} and is given by $V_{\text{terminal}} = \varepsilon - Ir$, where I is the electric current and is positive when flowing away from the positive terminal of the voltage source and r is the internal resistance.

Conceptual Questions

Exercise:

Problem:

What effect will the internal resistance of a rechargeable battery have on the energy being used to recharge the battery?

Solution:

Some of the energy being used to recharge the battery will be dissipated as heat by the internal resistance.

Exercise:**Problem:**

A battery with an internal resistance of r and an emf of 10.00 V is connected to a load resistor $R = r$. As the battery ages, the internal resistance triples. How much is the current through the load resistor reduced?

Exercise:**Problem:**

Show that the power dissipated by the load resistor is maximum when the resistance of the load resistor is equal to the internal resistance of the battery.

Solution:

$$P = I^2 R = \left(\frac{\varepsilon}{r+R} \right)^2 R = \varepsilon^2 R (r+R)^{-2}, \quad \frac{dP}{dR} = \varepsilon^2 \left[(r+R)^{-2} - 2R(r+R)^{-3} \right] = 0,$$
$$\left[\frac{(r+R)-2R}{(r+R)^3} \right] = 0, \quad r = R$$

Problems**Exercise:****Problem:**

A car battery with a 12-V emf and an internal resistance of $0.050 \, \Omega$ is being charged with a current of 60 A. Note that in this process, the battery is being charged. (a) What is the potential difference across its terminals? (b) At what rate is thermal energy being dissipated in the battery? (c) At what rate is electric energy being converted into chemical energy?

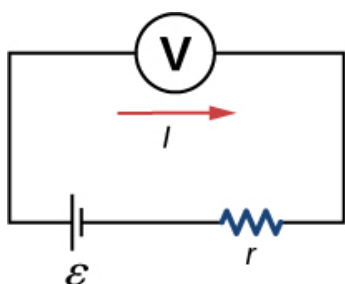
Exercise:

Problem:

The label on a battery-powered radio recommends the use of a rechargeable nickel-cadmium cell (nicads), although it has a 1.25-V emf, whereas an alkaline cell has a 1.58-V emf. The radio has a $3.20\ \Omega$ resistance. (a) Draw a circuit diagram of the radio and its battery. Now, calculate the power delivered to the radio (b) when using a nicad cells, each having an internal resistance of $0.0400\ \Omega$, and (c) when using an alkaline cell, having an internal resistance of $0.200\ \Omega$. (d) Does this difference seem significant, considering that the radio's effective resistance is lowered when its volume is turned up?

Solution:

a.



b. 0.476 W ; c. 0.691 W ; d. As R_L is lowered, the power difference decreases; therefore, at higher volumes, there is no significant difference.

Exercise:**Problem:**

An automobile starter motor has an equivalent resistance of $0.0500\ \Omega$ and is supplied by a 12.0-V battery with a $0.0100\text{-}\Omega$ internal resistance. (a) What is the current to the motor? (b) What voltage is applied to it? (c) What power is supplied to the motor? (d) Repeat these calculations for when the battery connections are corroded and add $0.0900\ \Omega$ to the circuit. (Significant problems are caused by even small amounts of unwanted resistance in low-voltage, high-current applications.)

Exercise:**Problem:**

(a) What is the internal resistance of a voltage source if its terminal potential drops by 2.00 V when the current supplied increases by 5.00 A? (b) Can the emf of the voltage source be found with the information supplied?

Solution:

a. $0.400\ \Omega$; b. No, there is only one independent equation, so only r can be found.

Exercise:

Problem:

A person with body resistance between his hands of $10.0\ \text{k}\Omega$ accidentally grasps the terminals of a 20.0-kV power supply. (Do NOT do this!) (a) Draw a circuit diagram to represent the situation. (b) If the internal resistance of the power supply is $2000\ \Omega$, what is the current through his body? (c) What is the power dissipated in his body? (d) If the power supply is to be made safe by increasing its internal resistance, what should the internal resistance be for the maximum current in this situation to be $1.00\ \text{mA}$ or less? (e) Will this modification compromise the effectiveness of the power supply for driving low-resistance devices? Explain your reasoning.

Exercise:

Problem:

A 12.0-V emf automobile battery has a terminal voltage of $16.0\ \text{V}$ when being charged by a current of $10.0\ \text{A}$. (a) What is the battery's internal resistance? (b) What power is dissipated inside the battery? (c) At what rate (in $^{\circ}\text{C}/\text{min}$) will its temperature increase if its mass is $20.0\ \text{kg}$ and it has a specific heat of $0.300\ \text{kcal}/\text{kg} \cdot ^{\circ}\text{C}$, assuming no heat escapes?

Solution:

a. $0.400\ \Omega$; b. $40.0\ \text{W}$; c. $0.0956\ ^{\circ}\text{C}/\text{min}$

Glossary

electromotive force (emf)

energy produced per unit charge, drawn from a source that produces an electrical current

internal resistance

amount of resistance to the flow of current within the voltage source

potential difference

difference in electric potential between two points in an electric circuit, measured in volts

potential drop

loss of electric potential energy as a current travels across a resistor, wire, or other component

terminal voltage

potential difference measured across the terminals of a source when there is no load attached

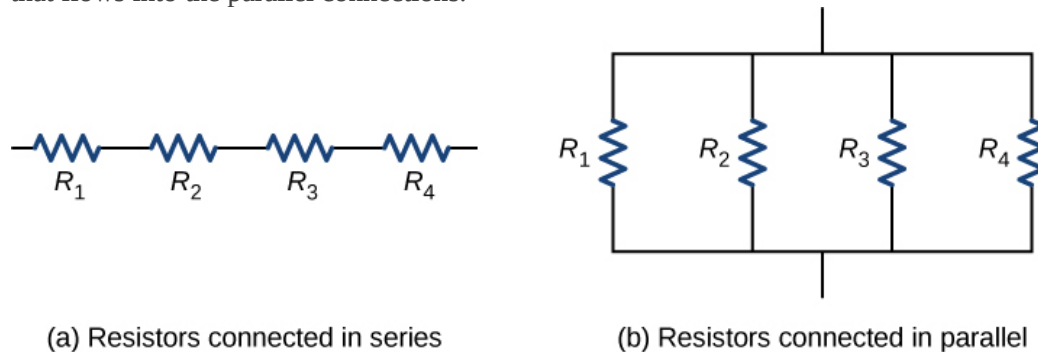
Resistors in Series and Parallel

By the end of the section, you will be able to:

- Define the term equivalent resistance
- Calculate the equivalent resistance of resistors connected in series
- Calculate the equivalent resistance of resistors connected in parallel

In [Current and Resistance](#), we described the term ‘resistance’ and explained the basic design of a resistor. Basically, a resistor limits the flow of charge in a circuit and is an ohmic device where $V = IR$. Most circuits have more than one resistor. If several resistors are connected together and connected to a battery, the current supplied by the battery depends on the **equivalent resistance** of the circuit.

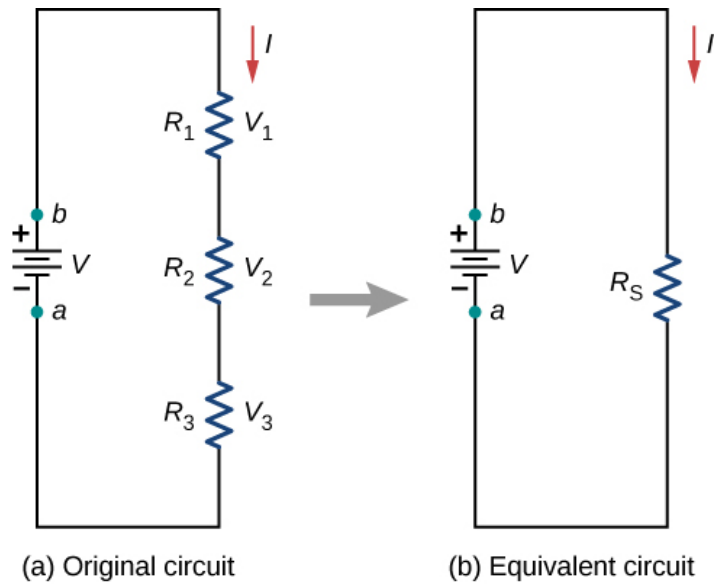
The equivalent resistance of a combination of resistors depends on both their individual values and how they are connected. The simplest combinations of resistors are series and parallel connections ([link](#)). In a series circuit, the output current of the first resistor flows into the input of the second resistor; therefore, the current is the same in each resistor. In a parallel circuit, all of the resistor leads on one side of the resistors are connected together and all the leads on the other side are connected together. In the case of a parallel configuration, each resistor has the same potential drop across it, and the currents through each resistor may be different, depending on the resistor. The sum of the individual currents equals the current that flows into the parallel connections.



(a) For a series connection of resistors, the current is the same in each resistor. (b) For a parallel connection of resistors, the voltage is the same across each resistor.

Resistors in Series

Resistors are said to be in series whenever the current flows through the resistors sequentially. Consider [link](#), which shows three resistors in series with an applied voltage equal to V_{ab} . Since there is only one path for the charges to flow through, the current is the same through each resistor. The equivalent resistance of a set of resistors in a series connection is equal to the algebraic sum of the individual resistances.



- (a) Three resistors connected in series to a voltage source. (b) The original circuit is reduced to an equivalent resistance and a voltage source.

In [\[link\]](#), the current coming from the voltage source flows through each resistor, so the current through each resistor is the same. The current through the circuit depends on the voltage supplied by the voltage source and the resistance of the resistors. For each resistor, a potential drop occurs that is equal to the loss of electric potential energy as a current travels through each resistor. According to Ohm's law, the potential drop V across a resistor when a current flows through it is calculated using the equation $V = IR$, where I is the current in amps (A) and R is the resistance in ohms (Ω). Since energy is conserved, and the voltage is equal to the potential energy per charge, the sum of the voltage applied to the circuit by the source and the potential drops across the individual resistors around a loop should be equal to zero:

Equation:

$$\sum_{i=1}^N V_i = 0.$$

This equation is often referred to as Kirchhoff's loop law, which we will look at in more detail later in this chapter. For [\[link\]](#), the sum of the potential drop of each resistor and the voltage supplied by the voltage source should equal zero:

Equation:

$$\begin{aligned} V - V_1 - V_2 - V_3 &= 0, \\ V &= V_1 + V_2 + V_3, \\ &= IR_1 + IR_2 + IR_3, \\ I &= \frac{V}{R_1 + R_2 + R_3} = \frac{V}{R_S}. \end{aligned}$$

Since the current through each component is the same, the equality can be simplified to an equivalent resistance, which is just the sum of the resistances of the individual resistors.

Any number of resistors can be connected in series. If N resistors are connected in series, the equivalent resistance is

Note:

Equation:

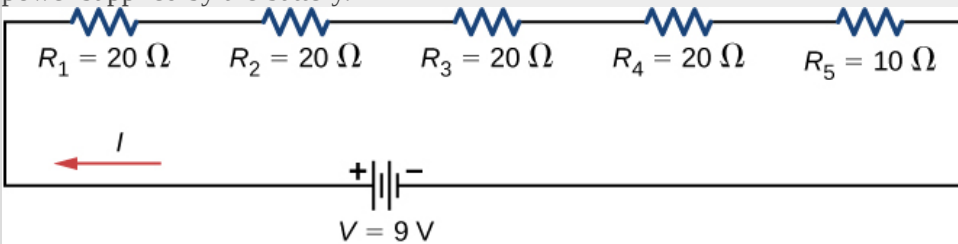
$$R_S = R_1 + R_2 + R_3 + \cdots + R_{N-1} + R_N = \sum_{i=1}^N R_i.$$

One result of components connected in a series circuit is that if something happens to one component, it affects all the other components. For example, if several lamps are connected in series and one bulb burns out, all the other lamps go dark.

Example:

Equivalent Resistance, Current, and Power in a Series Circuit

A battery with a terminal voltage of 9 V is connected to a circuit consisting of four 20- Ω and one 10- Ω resistors all in series ([link](#)). Assume the battery has negligible internal resistance. (a) Calculate the equivalent resistance of the circuit. (b) Calculate the current through each resistor. (c) Calculate the potential drop across each resistor. (d) Determine the total power dissipated by the resistors and the power supplied by the battery.



A simple series circuit with five resistors.

Strategy

In a series circuit, the equivalent resistance is the algebraic sum of the resistances. The current through the circuit can be found from Ohm's law and is equal to the voltage divided by the equivalent resistance. The potential drop across each resistor can be found using Ohm's law. The power dissipated by each resistor can be found using $P = I^2 R$, and the total power dissipated by the resistors is equal to the sum of the power dissipated by each resistor. The power supplied by the battery can be found using $P = I\varepsilon$.

Solution

- The equivalent resistance is the algebraic sum of the resistances:

Equation:

$$R_S = R_1 + R_2 + R_3 + R_4 + R_5 = 20\ \Omega + 20\ \Omega + 20\ \Omega + 20\ \Omega + 10\ \Omega = 90\ \Omega.$$

- b. The current through the circuit is the same for each resistor in a series circuit and is equal to the applied voltage divided by the equivalent resistance:

Equation:

$$I = \frac{V}{R_S} = \frac{9\ \text{V}}{90\ \Omega} = 0.1\ \text{A}.$$

- c. The potential drop across each resistor can be found using Ohm's law:

Equation:

$$V_1 = V_2 = V_3 = V_4 = (0.1\ \text{A})20\ \Omega = 2\ \text{V},$$

$$V_5 = (0.1\ \text{A})10\ \Omega = 1\ \text{V},$$

$$V_1 + V_2 + V_3 + V_4 + V_5 = 9\ \text{V}.$$

Note that the sum of the potential drops across each resistor is equal to the voltage supplied by the battery.

- d. The power dissipated by a resistor is equal to $P = I^2 R$, and the power supplied by the battery is equal to $P = I\varepsilon$:

Equation:

$$P_1 = P_2 = P_3 = P_4 = (0.1\ \text{A})^2 (20\ \Omega) = 0.2\ \text{W},$$

$$P_5 = (0.1\ \text{A})^2 (10\ \Omega) = 0.1\ \text{W},$$

$$P_{\text{dissipated}} = 0.2\ \text{W} + 0.2\ \text{W} + 0.2\ \text{W} + 0.2\ \text{W} + 0.1\ \text{W} = 0.9\ \text{W},$$

$$P_{\text{source}} = I\varepsilon = (0.1\ \text{A})(9\ \text{V}) = 0.9\ \text{W}.$$

Significance

There are several reasons why we would use multiple resistors instead of just one resistor with a resistance equal to the equivalent resistance of the circuit. Perhaps a resistor of the required size is not available, or we need to dissipate the heat generated, or we want to minimize the cost of resistors. Each resistor may cost a few cents to a few dollars, but when multiplied by thousands of units, the cost saving may be appreciable.

Note:

Exercise:

Problem:

Check Your Understanding Some strings of miniature holiday lights are made to short out when a bulb burns out. The device that causes the short is called a shunt, which allows current to flow around the open circuit. A “short” is like putting a piece of wire across the component. The bulbs are usually grouped in series of nine bulbs. If too many bulbs burn out, the shunts eventually open. What causes this?

Solution:

The equivalent resistance of nine bulbs connected in series is $9R$. The current is $I = V/9R$. If one bulb burns out, the equivalent resistance is $8R$, and the voltage does not change, but the current

increases ($I = V/8 R$). As more bulbs burn out, the current becomes even higher. Eventually, the current becomes too high, burning out the shunt.

Let's briefly summarize the major features of resistors in series:

1. Series resistances add together to get the equivalent resistance:

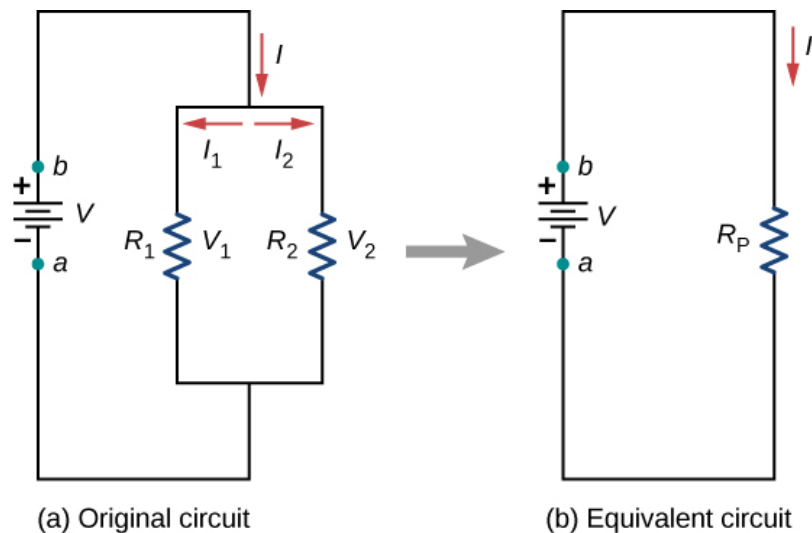
Equation:

$$R_S = R_1 + R_2 + R_3 + \cdots + R_{N-1} + R_N = \sum_{i=1}^N R_i.$$

2. The same current flows through each resistor in series.
3. Individual resistors in series do not get the total source voltage, but divide it. The total potential drop across a series configuration of resistors is equal to the sum of the potential drops across each resistor.

Resistors in Parallel

[\[link\]](#) shows resistors in parallel, wired to a voltage source. Resistors are in parallel when one end of all the resistors are connected by a continuous wire of negligible resistance and the other end of all the resistors are also connected to one another through a continuous wire of negligible resistance. The potential drop across each resistor is the same. Current through each resistor can be found using Ohm's law $I = V/R$, where the voltage is constant across each resistor. For example, an automobile's headlights, radio, and other systems are wired in parallel, so that each subsystem utilizes the full voltage of the source and can operate completely independently. The same is true of the wiring in your house or any building.



(a) Two resistors connected in parallel to a voltage source. (b) The original circuit is reduced to an equivalent resistance and a voltage source.

The current flowing from the voltage source in [\[link\]](#) depends on the voltage supplied by the voltage source and the equivalent resistance of the circuit. In this case, the current flows from the voltage source and enters a junction, or node, where the circuit splits flowing through resistors R_1 and R_2 . As the charges flow from the battery, some go through resistor R_1 and some flow through resistor R_2 . The sum of the currents flowing into a junction must be equal to the sum of the currents flowing out of the junction:

Equation:

$$\sum I_{\text{in}} = \sum I_{\text{out}}.$$

This equation is referred to as Kirchhoff's junction rule and will be discussed in detail in the next section. In [\[link\]](#), the junction rule gives $I = I_1 + I_2$. There are two loops in this circuit, which leads to the equations $V = I_1 R_1$ and $I_1 R_1 = I_2 R_2$. Note the voltage across the resistors in parallel are the same ($V = V_1 = V_2$) and the current is additive:

Equation:

$$\begin{aligned} I &= I_1 + I_2 \\ &= \frac{V_1}{R_1} + \frac{V_2}{R_2} \\ &= \frac{V}{R_1} + \frac{V}{R_2} \\ &= V \left(\frac{1}{R_1} + \frac{1}{R_2} \right) = \frac{V}{R_P} \\ R_P &= \left(\frac{1}{R_1} + \frac{1}{R_2} \right)^{-1}. \end{aligned}$$

Generalizing to any number of N resistors, the equivalent resistance R_P of a parallel connection is related to the individual resistances by

Note:

Equation:

$$R_P = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_{N-1}} + \frac{1}{R_N} \right)^{-1} = \left(\sum_{i=1}^N \frac{1}{R_i} \right)^{-1}.$$

This relationship results in an equivalent resistance R_P that is less than the smallest of the individual resistances. When resistors are connected in parallel, more current flows from the source than would flow for any of them individually, so the total resistance is lower.

Example:

Analysis of a Parallel Circuit

Three resistors $R_1 = 1.00\ \Omega$, $R_2 = 2.00\ \Omega$, and $R_3 = 2.00\ \Omega$, are connected in parallel. The parallel connection is attached to a $V = 3.00\ \text{V}$ voltage source. (a) What is the equivalent resistance? (b) Find the current supplied by the source to the parallel circuit. (c) Calculate the currents in each resistor and show that these add together to equal the current output of the source. (d) Calculate the power dissipated by each resistor. (e) Find the power output of the source and show that it equals the total power dissipated by the resistors.

Strategy

(a) The total resistance for a parallel combination of resistors is found using $R_P = \left(\sum_i \frac{1}{R_i} \right)^{-1}$.

(Note that in these calculations, each intermediate answer is shown with an extra digit.)

(b) The current supplied by the source can be found from Ohm's law, substituting R_P for the total resistance $I = \frac{V}{R_P}$.

(c) The individual currents are easily calculated from Ohm's law $\left(I_i = \frac{V_i}{R_i} \right)$, since each resistor gets the full voltage. The total current is the sum of the individual currents: $I = \sum_i I_i$.

(d) The power dissipated by each resistor can be found using any of the equations relating power to current, voltage, and resistance, since all three are known. Let us use $P_i = V^2/R_i$, since each resistor gets full voltage.

(e) The total power can also be calculated in several ways, use $P = IV$.

Solution

a. The total resistance for a parallel combination of resistors is found using [\[link\]](#). Entering known values gives

Equation:

$$R_P = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)^{-1} = \left(\frac{1}{1.00\ \Omega} + \frac{1}{2.00\ \Omega} + \frac{1}{2.00\ \Omega} \right)^{-1} = 0.50\ \Omega.$$

The total resistance with the correct number of significant digits is $R_P = 0.50\ \Omega$. As predicted, R_P is less than the smallest individual resistance.

b. The total current can be found from Ohm's law, substituting R_P for the total resistance. This gives
Equation:

$$I = \frac{V}{R_P} = \frac{3.00\ \text{V}}{0.50\ \Omega} = 6.00\ \text{A}.$$

Current I for each device is much larger than for the same devices connected in series (see the previous example). A circuit with parallel connections has a smaller total resistance than the resistors connected in series.

c. The individual currents are easily calculated from Ohm's law, since each resistor gets the full voltage. Thus,

Equation:

$$I_1 = \frac{V}{R_1} = \frac{3.00\ \text{V}}{1.00\ \Omega} = 3.00\ \text{A}.$$

Similarly,

Equation:

$$I_2 = \frac{V}{R_2} = \frac{3.00 \text{ V}}{2.00 \Omega} = 1.50 \text{ A}$$

and

Equation:

$$I_3 = \frac{V}{R_3} = \frac{3.00 \text{ V}}{2.00 \Omega} = 1.50 \text{ A}.$$

The total current is the sum of the individual currents:

Equation:

$$I_1 + I_2 + I_3 = 6.00 \text{ A}.$$

- d. The power dissipated by each resistor can be found using any of the equations relating power to current, voltage, and resistance, since all three are known. Let us use $P = V^2/R$, since each resistor gets full voltage. Thus,

Equation:

$$P_1 = \frac{V^2}{R_1} = \frac{(3.00 \text{ V})^2}{1.00 \Omega} = 9.00 \text{ W}.$$

Similarly,

Equation:

$$P_2 = \frac{V^2}{R_2} = \frac{(3.00 \text{ V})^2}{2.00 \Omega} = 4.50 \text{ W}$$

and

Equation:

$$P_3 = \frac{V^2}{R_3} = \frac{(3.00 \text{ V})^2}{2.00 \Omega} = 4.50 \text{ W}.$$

- e. The total power can also be calculated in several ways. Choosing $P = IV$ and entering the total current yields

Equation:

$$P = IV = (6.00 \text{ A})(3.00 \text{ V}) = 18.00 \text{ W}.$$

Significance

Total power dissipated by the resistors is also 18.00 W:

Equation:

$$P_1 + P_2 + P_3 = 9.00 \text{ W} + 4.50 \text{ W} + 4.50 \text{ W} = 18.00 \text{ W}.$$

Notice that the total power dissipated by the resistors equals the power supplied by the source.

Note:

Exercise:

Problem:

Check Your Understanding Consider the same potential difference ($V = 3.00 \text{ V}$) applied to the same three resistors connected in series. Would the equivalent resistance of the series circuit be higher, lower, or equal to the three resistor in parallel? Would the current through the series circuit be higher, lower, or equal to the current provided by the same voltage applied to the parallel circuit? How would the power dissipated by the resistor in series compare to the power dissipated by the resistors in parallel?

Solution:

The equivalent of the series circuit would be $R_{\text{eq}} = 1.00 \, \Omega + 2.00 \, \Omega + 2.00 \, \Omega = 5.00 \, \Omega$, which is higher than the equivalent resistance of the parallel circuit $R_{\text{eq}} = 0.50 \, \Omega$. The equivalent resistor of any number of resistors is always higher than the equivalent resistance of the same resistors connected in parallel. The current through for the series circuit would be $I = \frac{3.00 \text{ V}}{5.00 \, \Omega} = 0.60 \text{ A}$, which is lower than the sum of the currents through each resistor in the parallel circuit, $I = 6.00 \text{ A}$. This is not surprising since the equivalent resistance of the series circuit is higher. The current through a series connection of any number of resistors will always be lower than the current into a parallel connection of the same resistors, since the equivalent resistance of the series circuit will be higher than the parallel circuit. The power dissipated by the resistors in series would be $P = 1.80 \text{ W}$, which is lower than the power dissipated in the parallel circuit $P = 18.00 \text{ W}$.

Note:**Exercise:****Problem:**

Check Your Understanding How would you use a river and two waterfalls to model a parallel configuration of two resistors? How does this analogy break down?

Solution:

A river, flowing horizontally at a constant rate, splits in two and flows over two waterfalls. The water molecules are analogous to the electrons in the parallel circuits. The number of water molecules that flow in the river and falls must be equal to the number of molecules that flow over each waterfall, just like sum of the current through each resistor must be equal to the current flowing into the parallel circuit. The water molecules in the river have energy due to their motion and height. The potential energy of the water molecules in the river is constant due to their equal heights. This is analogous to the constant change in voltage across a parallel circuit. Voltage is the potential energy across each resistor.

The analogy quickly breaks down when considering the energy. In the waterfall, the potential energy is converted into kinetic energy of the water molecules. In the case of electrons flowing through a resistor, the potential drop is converted into heat and light, not into the kinetic energy of the electrons.

Let us summarize the major features of resistors in parallel:

1. Equivalent resistance is found from

Equation:

$$R_P = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_{N-1}} + \frac{1}{R_N} \right)^{-1} = \left(\sum_{i=1}^N \frac{1}{R_i} \right)^{-1},$$

and is smaller than any individual resistance in the combination.

2. The potential drop across each resistor in parallel is the same.
3. Parallel resistors do not each get the total current; they divide it. The current entering a parallel combination of resistors is equal to the sum of the current through each resistor in parallel.

In this chapter, we introduced the equivalent resistance of resistors connect in series and resistors connected in parallel. You may recall that in [Capacitance](#), we introduced the equivalent capacitance of capacitors connected in series and parallel. Circuits often contain both capacitors and resistors. [\[link\]](#) summarizes the equations used for the equivalent resistance and equivalent capacitance for series and parallel connections.

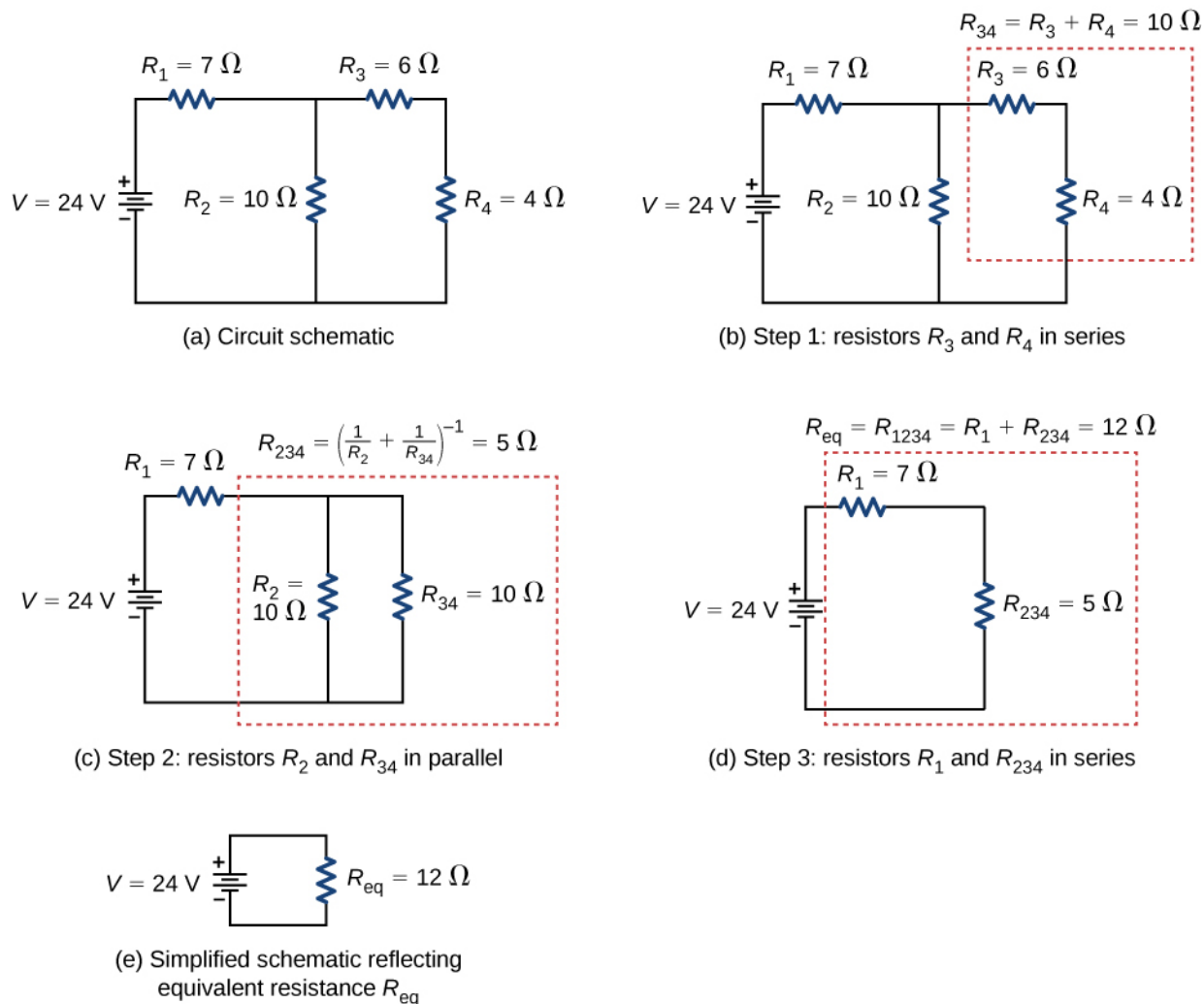
	Series combination	Parallel combination
Equivalent capacitance	$\frac{1}{C_S} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots$	$C_P = C_1 + C_2 + C_3 + \cdots$
Equivalent resistance	$R_S = R_1 + R_2 + R_3 + \cdots = \sum_{i=1}^N R_i$	$\frac{1}{R_P} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots$

Summary for Equivalent Resistance and Capacitance in Series and Parallel Combinations

Combinations of Series and Parallel

More complex connections of resistors are often just combinations of series and parallel connections. Such combinations are common, especially when wire resistance is considered. In that case, wire resistance is in series with other resistances that are in parallel.

Combinations of series and parallel can be reduced to a single equivalent resistance using the technique illustrated in [\[link\]](#). Various parts can be identified as either series or parallel connections, reduced to their equivalent resistances, and then further reduced until a single equivalent resistance is left. The process is more time consuming than difficult. Here, we note the equivalent resistance as R_{eq} .



(a) The original circuit of four resistors. (b) Step 1: The resistors R_3 and R_4 are in series and the equivalent resistance is $R_{34} = 10\ \Omega$. (c) Step 2: The reduced circuit shows resistors R_2 and R_{34} are in parallel, with an equivalent resistance of $R_{234} = 5\ \Omega$. (d) Step 3: The reduced circuit shows that R_1 and R_{234} are in series with an equivalent resistance of $R_{1234} = 12\ \Omega$, which is the equivalent resistance R_{eq} . (e) The reduced circuit with a voltage source of $V = 24\text{ V}$ with an equivalent resistance of $R_{eq} = 12\ \Omega$. This results in a current of $I = 2\text{ A}$ from the voltage source.

Notice that resistors R_3 and R_4 are in series. They can be combined into a single equivalent resistance. One method of keeping track of the process is to include the resistors as subscripts. Here the equivalent resistance of R_3 and R_4 is

Equation:

$$R_{34} = R_3 + R_4 = 6\ \Omega + 4\ \Omega = 10\ \Omega.$$

The circuit now reduces to three resistors, shown in [\[link\]](#)(c). Redrawing, we now see that resistors R_2 and R_{34} constitute a parallel circuit. Those two resistors can be reduced to an equivalent resistance:

Equation:

$$R_{234} = \left(\frac{1}{R_2} + \frac{1}{R_{34}} \right)^{-1} = \left(\frac{1}{10\ \Omega} + \frac{1}{10\ \Omega} \right)^{-1} = 5\ \Omega.$$

This step of the process reduces the circuit to two resistors, shown in in [\[link\]](#)(d). Here, the circuit reduces to two resistors, which in this case are in series. These two resistors can be reduced to an equivalent resistance, which is the equivalent resistance of the circuit:

Equation:

$$R_{\text{eq}} = R_{1234} = R_1 + R_{234} = 7\ \Omega + 5\ \Omega = 12\ \Omega.$$

The main goal of this circuit analysis is reached, and the circuit is now reduced to a single resistor and single voltage source.

Now we can analyze the circuit. The current provided by the voltage source is $I = \frac{V}{R_{\text{eq}}} = \frac{24\text{ V}}{12\ \Omega} = 2\text{ A}$.

This current runs through resistor R_1 and is designated as I_1 . The potential drop across R_1 can be found using Ohm's law:

Equation:

$$V_1 = I_1 R_1 = (2\text{ A})(7\ \Omega) = 14\text{ V}.$$

Looking at [\[link\]](#)(c), this leaves $24\text{ V} - 14\text{ V} = 10\text{ V}$ to be dropped across the parallel combination of R_2 and R_{34} . The current through R_2 can be found using Ohm's law:

Equation:

$$I_2 = \frac{V_2}{R_2} = \frac{10\text{ V}}{10\ \Omega} = 1\text{ A}.$$

The resistors R_3 and R_4 are in series so the currents I_3 and I_4 are equal to

Equation:

$$I_3 = I_4 = I - I_2 = 2\text{ A} - 1\text{ A} = 1\text{ A}.$$

Using Ohm's law, we can find the potential drop across the last two resistors. The potential drops are $V_3 = I_3 R_3 = 6\text{ V}$ and $V_4 = I_4 R_4 = 4\text{ V}$. The final analysis is to look at the power supplied by the voltage source and the power dissipated by the resistors. The power dissipated by the resistors is

Equation:

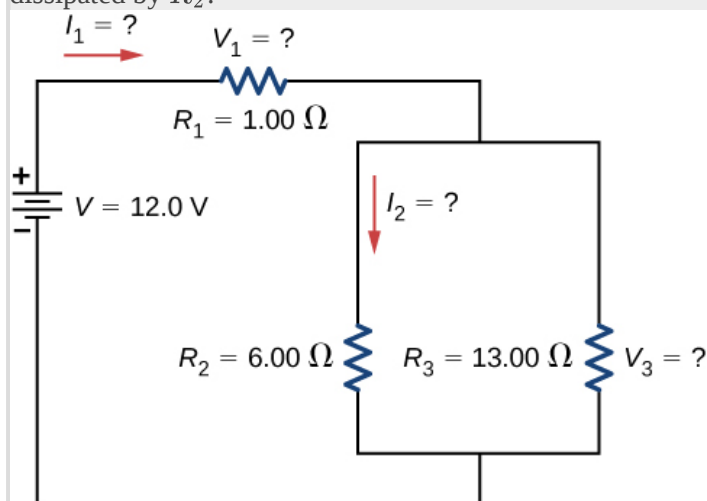
$$\begin{aligned} P_1 &= I_1^2 R_1 = (2\text{ A})^2 (7\ \Omega) = 28\text{ W}, \\ P_2 &= I_2^2 R_2 = (1\text{ A})^2 (10\ \Omega) = 10\text{ W}, \\ P_3 &= I_3^2 R_3 = (1\text{ A})^2 (6\ \Omega) = 6\text{ W}, \\ P_4 &= I_4^2 R_4 = (1\text{ A})^2 (4\ \Omega) = 4\text{ W}, \\ P_{\text{dissipated}} &= P_1 + P_2 + P_3 + P_4 = 48\text{ W}. \end{aligned}$$

The total energy is constant in any process. Therefore, the power supplied by the voltage source is $P_s = IV = (2 \text{ A})(24 \text{ V}) = 48 \text{ W}$. Analyzing the power supplied to the circuit and the power dissipated by the resistors is a good check for the validity of the analysis; they should be equal.

Example:

Combining Series and Parallel Circuits

[link](#) shows resistors wired in a combination of series and parallel. We can consider R_1 to be the resistance of wires leading to R_2 and R_3 . (a) Find the equivalent resistance of the circuit. (b) What is the potential drop V_1 across resistor R_1 ? (c) Find the current I_2 through resistor R_2 . (d) What power is dissipated by R_2 ?



These three resistors are connected to a voltage source so that R_2 and R_3 are in parallel with one another and that combination is in series with R_1 .

Strategy

- (a) To find the equivalent resistance, first find the equivalent resistance of the parallel connection of R_2 and R_3 . Then use this result to find the equivalent resistance of the series connection with R_1 .
- (b) The current through R_1 can be found using Ohm's law and the voltage applied. The current through R_1 is equal to the current from the battery. The potential drop V_1 across the resistor R_1 (which represents the resistance in the connecting wires) can be found using Ohm's law.
- (c) The current through R_2 can be found using Ohm's law $I_2 = \frac{V_2}{R_2}$. The voltage across R_2 can be found using $V_2 = V - V_1$.
- (d) Using Ohm's law ($V_2 = I_2 R_2$), the power dissipated by the resistor can also be found using $P_2 = I_2^2 R_2 = \frac{V_2^2}{R_2}$.

Solution

- a. To find the equivalent resistance of the circuit, notice that the parallel connection of R_2 and R_3 is in series with R_1 , so the equivalent resistance is

Equation:

$$R_{\text{eq}} = R_1 + \left(\frac{1}{R_2} + \frac{1}{R_3} \right)^{-1} = 1.00 \, \Omega + \left(\frac{1}{6.00 \, \Omega} + \frac{1}{13.00 \, \Omega} \right)^{-1} = 5.10 \, \Omega.$$

The total resistance of this combination is intermediate between the pure series and pure parallel values ($20.0 \, \Omega$ and $0.804 \, \Omega$, respectively).

- b. The current through R_1 is equal to the current supplied by the battery:

Equation:

$$I_1 = I = \frac{V}{R_{\text{eq}}} = \frac{12.0 \, \text{V}}{5.10 \, \Omega} = 2.35 \, \text{A}.$$

The voltage across R_1 is

Equation:

$$V_1 = I_1 R_1 = (2.35 \, \text{A})(1 \, \Omega) = 2.35 \, \text{V}.$$

The voltage applied to R_2 and R_3 is less than the voltage supplied by the battery by an amount V_1 . When wire resistance is large, it can significantly affect the operation of the devices represented by R_2 and R_3 .

- c. To find the current through R_2 , we must first find the voltage applied to it. The voltage across the two resistors in parallel is the same:

Equation:

$$V_2 = V_3 = V - V_1 = 12.0 \, \text{V} - 2.35 \, \text{V} = 9.65 \, \text{V}.$$

Now we can find the current I_2 through resistance R_2 using Ohm's law:

Equation:

$$I_2 = \frac{V_2}{R_2} = \frac{9.65 \, \text{V}}{6.00 \, \Omega} = 1.61 \, \text{A}.$$

The current is less than the $2.00 \, \text{A}$ that flowed through R_2 when it was connected in parallel to the battery in the previous parallel circuit example.

- d. The power dissipated by R_2 is given by

Equation:

$$P_2 = I_2^2 R_2 = (1.61 \, \text{A})^2 (6.00 \, \Omega) = 15.5 \, \text{W}.$$

Significance

The analysis of complex circuits can often be simplified by reducing the circuit to a voltage source and an equivalent resistance. Even if the entire circuit cannot be reduced to a single voltage source and a single equivalent resistance, portions of the circuit may be reduced, greatly simplifying the analysis.

Note:

Exercise:

Problem:

Check Your Understanding Consider the electrical circuits in your home. Give at least two examples of circuits that must use a combination of series and parallel circuits to operate efficiently.

Solution:

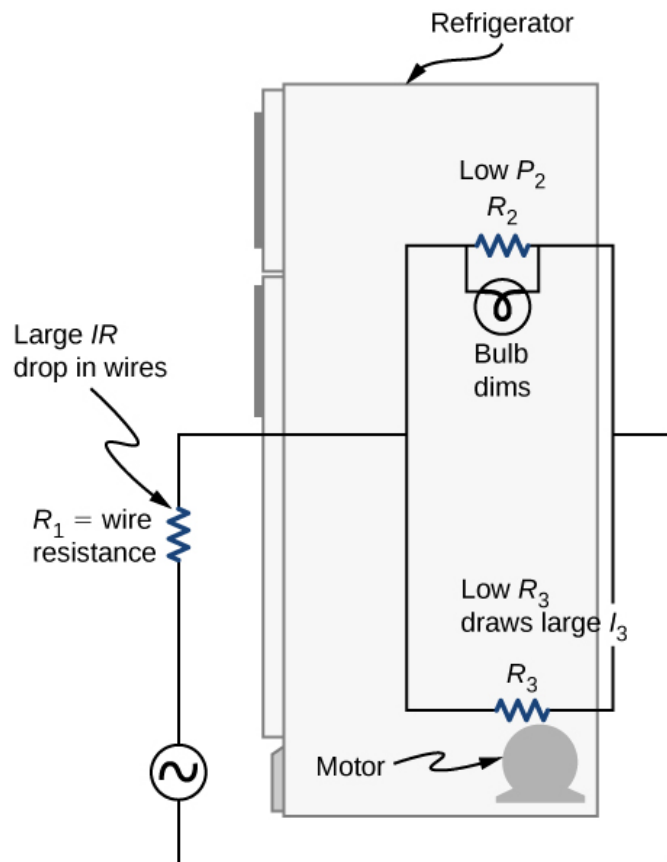
1. All the overhead lighting circuits are in parallel and connected to the main supply line, so when one bulb burns out, all the overhead lighting does not go dark. Each overhead light will have at least one switch in series with the light, so you can turn it on and off. 2. A refrigerator has a compressor and a light that goes on when the door opens. There is usually only one cord for the refrigerator to plug into the wall. The circuit containing the compressor and the circuit containing the lighting circuit are in parallel, but there is a switch in series with the light. A thermostat controls a switch that is in series with the compressor to control the temperature of the refrigerator.

Practical Implications

One implication of this last example is that resistance in wires reduces the current and power delivered to a resistor. If wire resistance is relatively large, as in a worn (or a very long) extension cord, then this loss can be significant. If a large current is drawn, the IR drop in the wires can also be significant and may become apparent from the heat generated in the cord.

For example, when you are rummaging in the refrigerator and the motor comes on, the refrigerator light dims momentarily. Similarly, you can see the passenger compartment light dim when you start the engine of your car (although this may be due to resistance inside the battery itself).

What is happening in these high-current situations is illustrated in [\[link\]](#). The device represented by R_3 has a very low resistance, so when it is switched on, a large current flows. This increased current causes a larger IR drop in the wires represented by R_1 , reducing the voltage across the light bulb (which is R_2), which then dims noticeably.



Why do lights dim when a large appliance is switched on? The answer is that the large current the appliance motor draws causes a significant IR drop in the wires and reduces the voltage across the light.

Note:

Series and Parallel Resistors

1. Draw a clear circuit diagram, labeling all resistors and voltage sources. This step includes a list of the known values for the problem, since they are labeled in your circuit diagram.
2. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.
3. Determine whether resistors are in series, parallel, or a combination of both series and parallel. Examine the circuit diagram to make this assessment. Resistors are in series if the same current must pass sequentially through them.
4. Use the appropriate list of major features for series or parallel connections to solve for the unknowns. There is one list for series and another for parallel.
5. Check to see whether the answers are reasonable and consistent.

Example:**Combining Series and Parallel Circuits**

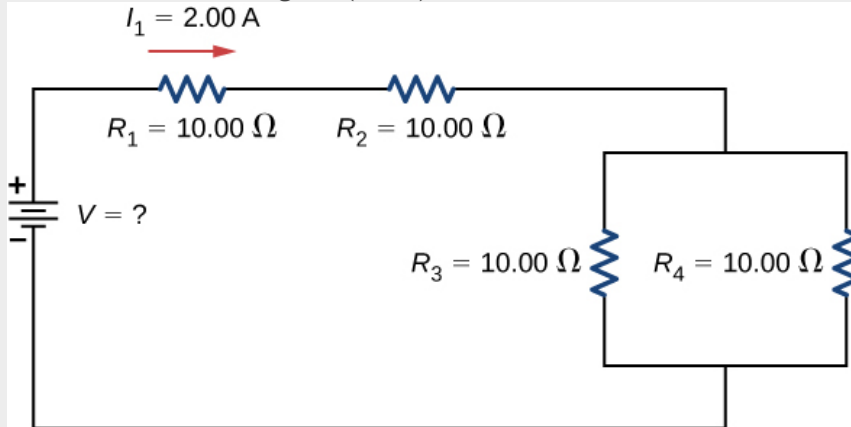
Two resistors connected in series (R_1, R_2) are connected to two resistors that are connected in parallel (R_3, R_4). The series-parallel combination is connected to a battery. Each resistor has a resistance of 10.00 Ohms. The wires connecting the resistors and battery have negligible resistance. A current of 2.00 Amps runs through resistor R_1 . What is the voltage supplied by the voltage source?

Strategy

Use the steps in the preceding problem-solving strategy to find the solution for this example.

Solution

1. Draw a clear circuit diagram ([link](#)).



To find the unknown voltage, we must first find the equivalent resistance of the circuit.

2. The unknown is the voltage of the battery. In order to find the voltage supplied by the battery, the equivalent resistance must be found.
3. In this circuit, we already know that the resistors R_1 and R_2 are in series and the resistors R_3 and R_4 are in parallel. The equivalent resistance of the parallel configuration of the resistors R_3 and R_4 is in series with the series configuration of resistors R_1 and R_2 .
4. The voltage supplied by the battery can be found by multiplying the current from the battery and the equivalent resistance of the circuit. The current from the battery is equal to the current through R_1 and is equal to 2.00 A. We need to find the equivalent resistance by reducing the circuit. To reduce the circuit, first consider the two resistors in parallel. The equivalent resistance is

$R_{34} = \left(\frac{1}{10.00\ \Omega} + \frac{1}{10.00\ \Omega} \right)^{-1} = 5.00\ \Omega$. This parallel combination is in series with the other two resistors, so the equivalent resistance of the circuit is $R_{eq} = R_1 + R_2 + R_{34} = 25.00\ \Omega$. The voltage supplied by the battery is therefore $V = IR_{eq} = 2.00\ \text{A} (25.00\ \Omega) = 50.00\ \text{V}$.

5. One way to check the consistency of your results is to calculate the power supplied by the battery and the power dissipated by the resistors. The power supplied by the battery is $P_{\text{batt}} = IV = 100.00\ \text{W}$.

Since they are in series, the current through R_2 equals the current through R_1 . Since $R_3 = R_4$, the current through each will be 1.00 Amps. The power dissipated by the resistors is equal to the sum of the power dissipated by each resistor:

Equation:

$$P = I_1^2 R_1 + I_2^2 R_2 + I_3^2 R_3 + I_4^2 R_4 = 40.00\ \text{W} + 40.00\ \text{W} + 10.00\ \text{W} + 10.00\ \text{W} = 100.00\ \text{W}.$$

Since the power dissipated by the resistors equals the power supplied by the battery, our solution seems consistent.

Significance

If a problem has a combination of series and parallel, as in this example, it can be reduced in steps by using the preceding problem-solving strategy and by considering individual groups of series or parallel connections. When finding R_{eq} for a parallel connection, the reciprocal must be taken with care. In addition, units and numerical results must be reasonable. Equivalent series resistance should be greater, whereas equivalent parallel resistance should be smaller, for example. Power should be greater for the same devices in parallel compared with series, and so on.

Summary

- The equivalent resistance of an electrical circuit with resistors wired in a series is the sum of the individual resistances: $R_s = R_1 + R_2 + R_3 + \cdots = \sum_{i=1}^N R_i$.
- Each resistor in a series circuit has the same amount of current flowing through it.
- The potential drop, or power dissipation, across each individual resistor in a series is different, and their combined total is the power source input.
- The equivalent resistance of an electrical circuit with resistors wired in parallel is less than the lowest resistance of any of the components and can be determined using the formula

Equation:

$$R_{\text{eq}} = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \right)^{-1} = \left(\sum_{i=1}^N \frac{1}{R_i} \right)^{-1}.$$

- Each resistor in a parallel circuit has the same full voltage of the source applied to it.
- The current flowing through each resistor in a parallel circuit is different, depending on the resistance.
- If a more complex connection of resistors is a combination of series and parallel, it can be reduced to a single equivalent resistance by identifying its various parts as series or parallel, reducing each to its equivalent, and continuing until a single resistance is eventually reached.

Conceptual Questions

Exercise:

Problem: A voltage occurs across an open switch. What is the power dissipated by the open switch?

Exercise:

Problem:

The severity of a shock depends on the magnitude of the current through your body. Would you prefer to be in series or in parallel with a resistance, such as the heating element of a toaster, if you were shocked by it? Explain.

Solution:

It would probably be better to be in series because the current will be less than if it were in parallel.

Exercise:

Problem:

Suppose you are doing a physics lab that asks you to put a resistor into a circuit, but all the resistors supplied have a larger resistance than the requested value. How would you connect the available resistances to attempt to get the smaller value asked for?

Exercise:

Problem:

Some light bulbs have three power settings (not including zero), obtained from multiple filaments that are individually switched and wired in parallel. What is the minimum number of filaments needed for three power settings?

Solution:

two filaments, a low resistance and a high resistance, connected in parallel

Problems

Exercise:

Problem:

(a) What is the resistance of a $1.00 \times 10^2\text{-}\Omega$, a $2.50\text{-k}\Omega$, and a $4.00\text{-k}\Omega$ resistor connected in series? (b) In parallel?

Exercise:

Problem:

What are the largest and smallest resistances you can obtain by connecting a $36.0\text{-}\Omega$, a $50.0\text{-}\Omega$, and a $700\text{-}\Omega$ resistor together?

Solution:

largest, $786\text{ }\Omega$, smallest, $20.32\text{ }\Omega$

Exercise:

Problem:

An 1800-W toaster, a 1400-W speaker, and a 75-W lamp are plugged into the same outlet in a 15-A fuse and 120-V circuit. (The three devices are in parallel when plugged into the same socket.) (a) What current is drawn by each device? (b) Will this combination blow the 15-A fuse?

Exercise:

Problem:

Your car's 30.0-W headlight and 2.40-kW starter are ordinarily connected in parallel in a 12.0-V system. What power would one headlight and the starter consume if connected in series to a 12.0-V battery? (Neglect any other resistance in the circuit and any change in resistance in the two devices.)

Solution:

29.6 W

Exercise:**Problem:**

(a) Given a 48.0-V battery and 24.0- Ω and 96.0- Ω resistors, find the current and power for each when connected in series. (b) Repeat when the resistances are in parallel.

Exercise:**Problem:**

Referring to the example combining series and parallel circuits and [\[link\]](#), calculate I_3 in the following two different ways: (a) from the known values of I and I_2 ; (b) using Ohm's law for R_3 . In both parts, explicitly show how you follow the steps in the [\[link\]](#).

Solution:

a. 0.74 A; b. 0.742 A

Exercise:**Problem:**

Referring to [\[link\]](#), (a) Calculate P_3 and note how it compares with P_3 found in the first two example problems in this module. (b) Find the total power supplied by the source and compare it with the sum of the powers dissipated by the resistors.

Exercise:**Problem:**

Refer to [\[link\]](#) and the discussion of lights dimming when a heavy appliance comes on. (a) Given the voltage source is 120 V, the wire resistance is 0.800 Ω , and the bulb is nominally 75.0 W, what power will the bulb dissipate if a total of 15.0 A passes through the wires when the motor comes on? Assume negligible change in bulb resistance. (b) What power is consumed by the motor?

Solution:

a. 60.8 W; b. 1.56 kW

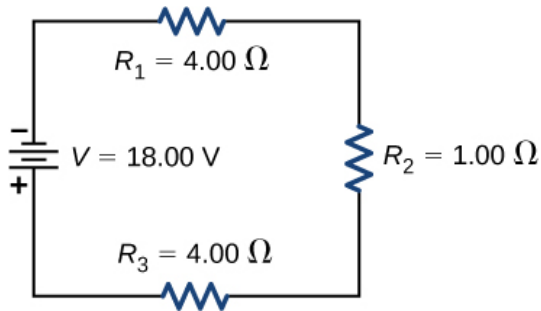
Exercise:**Problem:**

Show that if two resistors R_1 and R_2 are combined and one is much greater than the other ($R_1 \gg R_2$), (a) their series resistance is very nearly equal to the greater resistance R_1 and (b) their parallel resistance is very nearly equal to the smaller resistance R_2 .

Exercise:

Problem:

Consider the circuit shown below. The terminal voltage of the battery is $V = 18.00 \text{ V}$. (a) Find the equivalent resistance of the circuit. (b) Find the current through each resistor. (c) Find the potential drop across each resistor. (d) Find the power dissipated by each resistor. (e) Find the power supplied by the battery.



Solution:

- a. $R_s = 9.00 \Omega$; b. $I_1 = I_2 = I_3 = 2.00 \text{ A}$;
c. $V_1 = 8.00 \text{ V}$, $V_2 = 2.00 \text{ V}$, $V_3 = 8.00 \text{ V}$; d. $P_1 = 16.00 \text{ W}$, $P_2 = 4.00 \text{ W}$, $P_3 = 16.00 \text{ W}$; e.
 $P = 36.00 \text{ W}$

Glossary

equivalent resistance

resistance of a combination of resistors; it can be thought of as the resistance of a single resistor that can replace a combination of resistors in a series and/or parallel circuit

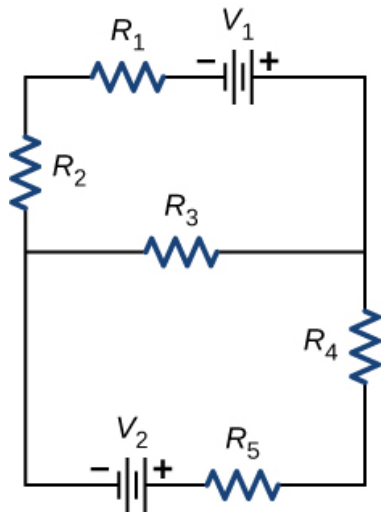
Kirchhoff's Rules

By the end of the section, you will be able to:

- State Kirchhoff's junction rule
- State Kirchhoff's loop rule
- Analyze complex circuits using Kirchhoff's rules

We have just seen that some circuits may be analyzed by reducing a circuit to a single voltage source and an equivalent resistance. Many complex circuits cannot be analyzed with the series-parallel techniques developed in the preceding sections. In this section, we elaborate on the use of Kirchhoff's rules to analyze more complex circuits. For example, the circuit in [\[link\]](#) is known as a multi-loop circuit, which consists of junctions. A junction, also known as a node, is a connection of three or more wires. In this circuit, the previous methods cannot be used, because not all the resistors are in clear series or parallel configurations that can be reduced. Give it a try. The resistors R_1 and R_2 are in series and can be reduced to an equivalent resistance. The same is true of resistors R_4 and R_5 . But what do you do then?

Even though this circuit cannot be analyzed using the methods already learned, two circuit analysis rules can be used to analyze any circuit, simple or complex. The rules are known as **Kirchhoff's rules**, after their inventor Gustav Kirchhoff (1824–1887).



This circuit cannot be reduced to a combination of series and parallel connections. However, we can use Kirchhoff's rules to analyze it.

Note:**Kirchhoff's Rules**

- Kirchhoff's first rule—the junction rule. The sum of all currents entering a junction must equal the sum of all currents leaving the junction:

Equation:

$$\sum I_{\text{in}} = \sum I_{\text{out}}.$$

- Kirchhoff's second rule—the loop rule. The algebraic sum of changes in potential around any closed circuit path (loop) must be zero:

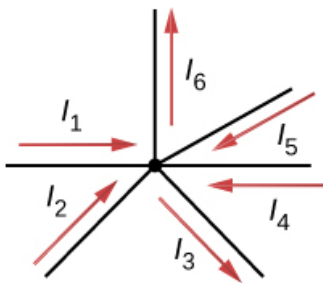
Equation:

$$\sum V = 0.$$

We now provide explanations of these two rules, followed by problem-solving hints for applying them and a worked example that uses them.

Kirchhoff's First Rule

Kirchhoff's first rule (the **junction rule**) applies to the charge entering and leaving a junction ([link](#)). As stated earlier, a junction, or node, is a connection of three or more wires. Current is the flow of charge, and charge is conserved; thus, whatever charge flows into the junction must flow out.



$$\sum I_{\text{in}} = \sum I_{\text{out}} \\ I_1 + I_2 + I_4 + I_5 = I_3 + I_6$$

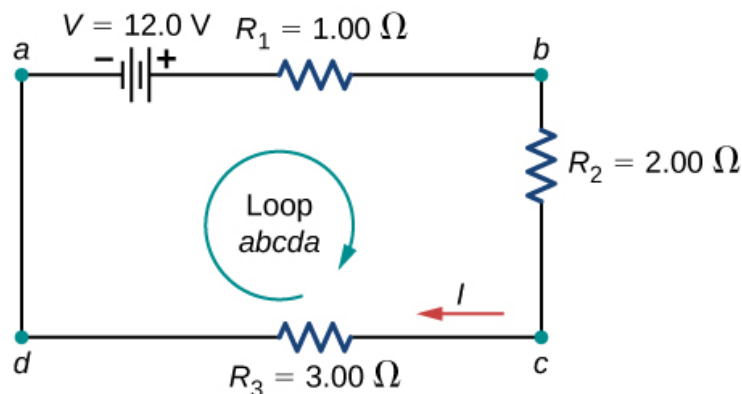
Charge must be conserved, so the sum of currents into a junction must be

equal to the sum of
currents out of the
junction.

Although it is an over-simplification, an analogy can be made with water pipes connected in a plumbing junction. If the wires in [\[link\]](#) were replaced by water pipes, and the water was assumed to be incompressible, the volume of water flowing into the junction must equal the volume of water flowing out of the junction.

Kirchhoff's Second Rule

Kirchhoff's second rule (the **loop rule**) applies to potential differences. The loop rule is stated in terms of potential V rather than potential energy, but the two are related since $U = qV$. In a closed loop, whatever energy is supplied by a voltage source, the energy must be transferred into other forms by the devices in the loop, since there are no other ways in which energy can be transferred into or out of the circuit. Kirchhoff's loop rule states that the algebraic sum of potential differences, including voltage supplied by the voltage sources and resistive elements, in any loop must be equal to zero. For example, consider a simple loop with no junctions, as in [\[link\]](#).

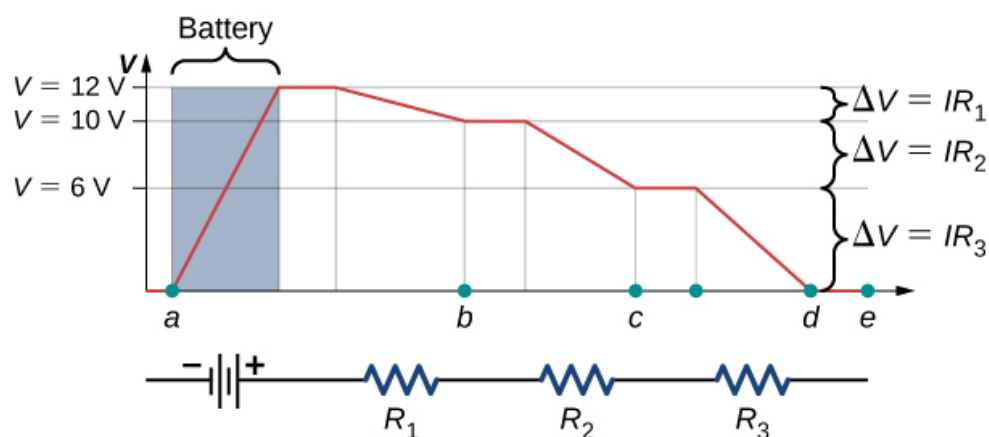


A simple loop with no junctions. Kirchhoff's loop rule states that the algebraic sum of the voltage differences is equal to zero.

The circuit consists of a voltage source and three external load resistors. The labels a , b , c , and d serve as references, and have no other significance. The usefulness of these labels will become apparent soon. The loop is designated as Loop $abcda$, and the labels help keep track of the voltage differences as we travel around the circuit. Start at point a and travel to point b .

The voltage of the voltage source is added to the equation and the potential drop of the resistor R_1 is subtracted. From point b to c , the potential drop across R_2 is subtracted. From c to d , the potential drop across R_3 is subtracted. From points d to a , nothing is done because there are no components.

[\[link\]](#) shows a graph of the voltage as we travel around the loop. Voltage increases as we cross the battery, whereas voltage decreases as we travel across a resistor. The potential drop, or change in the electric potential, is equal to the current through the resistor times the resistance of the resistor. Since the wires have negligible resistance, the voltage remains constant as we cross the wires connecting the components.



A voltage graph as we travel around the circuit. The voltage increases as we cross the battery and decreases as we cross each resistor. Since the resistance of the wire is quite small, we assume that the voltage remains constant as we cross the wires connecting the components.

Then Kirchhoff's loop rule states

Equation:

$$V - IR_1 - IR_2 - IR_3 = 0.$$

The loop equation can be used to find the current through the loop:

Equation:

$$I = \frac{V}{R_1 + R_2 + R_3} = \frac{12.00 \text{ V}}{1.00 \Omega + 2.00 \Omega + 3.00 \Omega} = 2.00 \text{ A}.$$

This loop could have been analyzed using the previous methods, but we will demonstrate the power of Kirchhoff's method in the next section.

Applying Kirchhoff's Rules

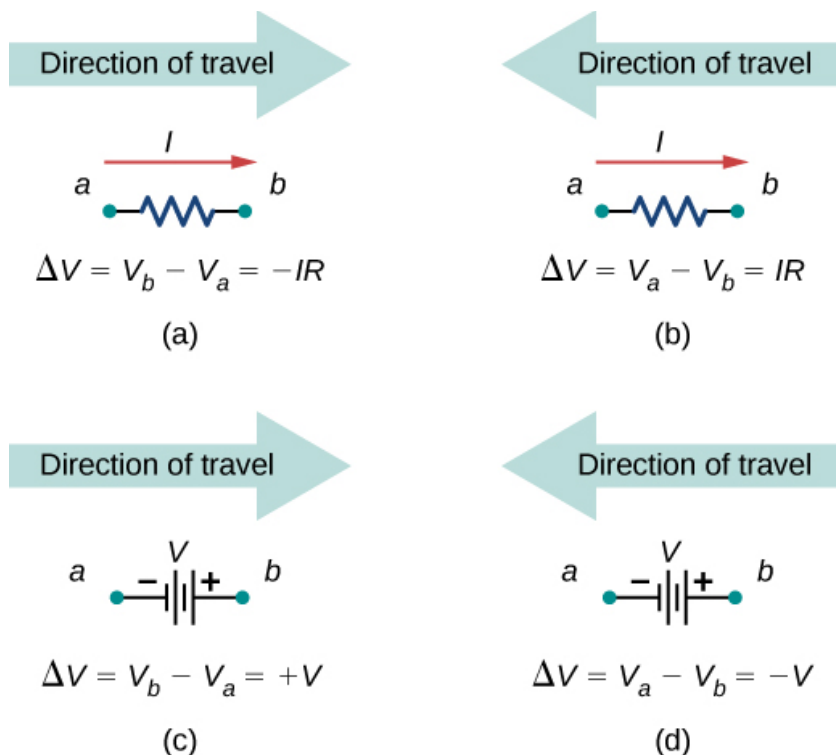
By applying Kirchhoff's rules, we generate a set of linear equations that allow us to find the unknown values in circuits. These may be currents, voltages, or resistances. Each time a rule is applied, it produces an equation. If there are as many independent equations as unknowns, then the problem can be solved.

Using Kirchhoff's method of analysis requires several steps, as listed in the following procedure.

Note:

Kirchhoff's Rules

1. Label points in the circuit diagram using lowercase letters a, b, c, \dots . These labels simply help with orientation.
2. Locate the junctions in the circuit. The junctions are points where three or more wires connect. Label each junction with the currents and directions into and out of it. Make sure at least one current points into the junction and at least one current points out of the junction.
3. Choose the loops in the circuit. Every component must be contained in at least one loop, but a component may be contained in more than one loop.
4. Apply the junction rule. Again, some junctions should not be included in the analysis. You need only use enough nodes to include every current.
5. Apply the loop rule. Use the map in [\[link\]](#).



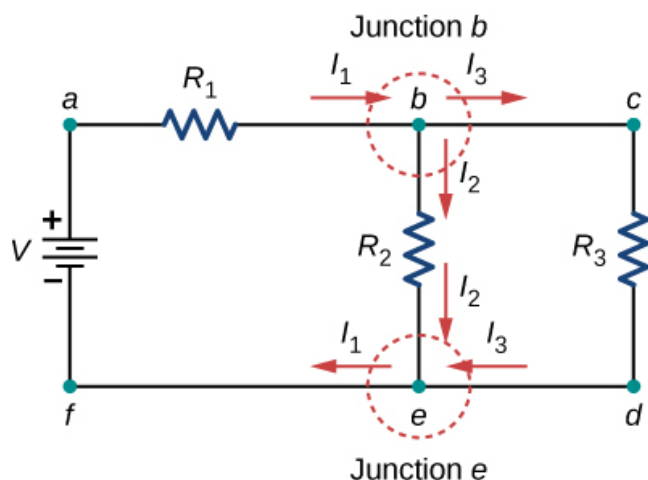
Each of these resistors and voltage sources is traversed from a to b . (a) When moving across a resistor in the same direction as the current flow, subtract the potential drop. (b) When moving across a resistor in the opposite direction as the current flow, add the potential drop. (c) When moving across a voltage source from the negative terminal to the positive terminal, add the potential drop. (d) When moving across a voltage source from the positive terminal to the negative terminal, subtract the potential drop.

Let's examine some steps in this procedure more closely. When locating the junctions in the circuit, do not be concerned about the direction of the currents. If the direction of current flow is not obvious, choosing any direction is sufficient as long as at least one current points into the junction and at least one current points out of the junction. If the arrow is in the opposite direction of the conventional current flow, the result for the current in question will be negative but the answer will still be correct.

The number of nodes depends on the circuit. Each current should be included in a node and thus included in at least one junction equation. Do not include nodes that are not linearly independent, meaning nodes that contain the same information.

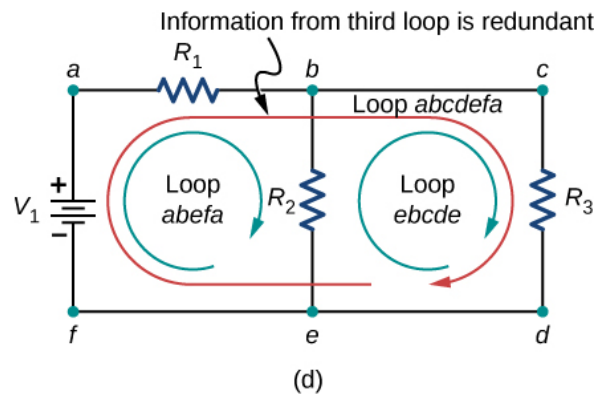
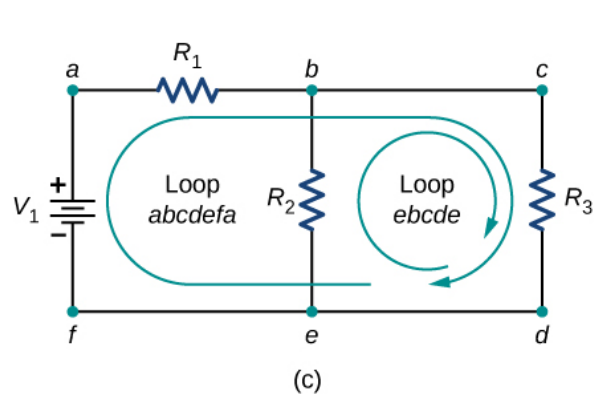
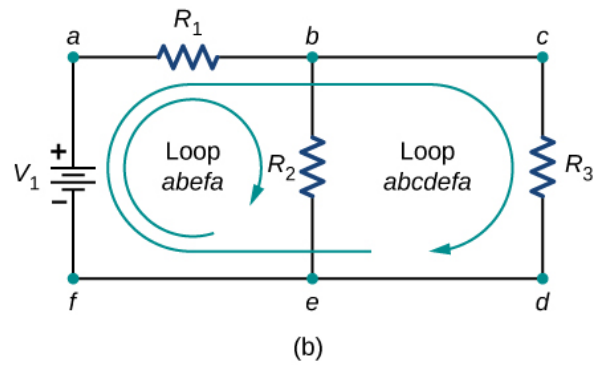
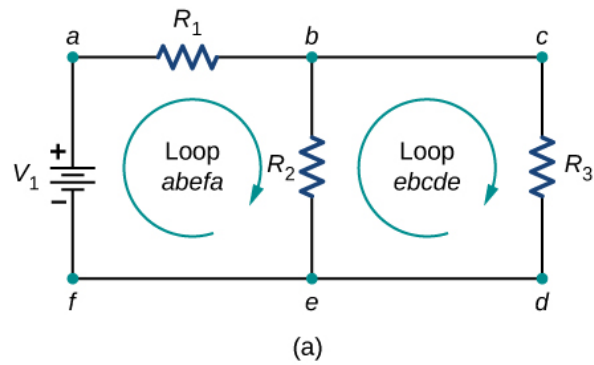
Consider [\[link\]](#). There are two junctions in this circuit: Junction b and Junction e . Points a , c , d , and f are not junctions, because a junction must have three or more connections. The

equation for Junction b is $I_1 = I_2 + I_3$, and the equation for Junction e is $I_2 + I_3 = I_1$. These are equivalent equations, so it is necessary to keep only one of them.



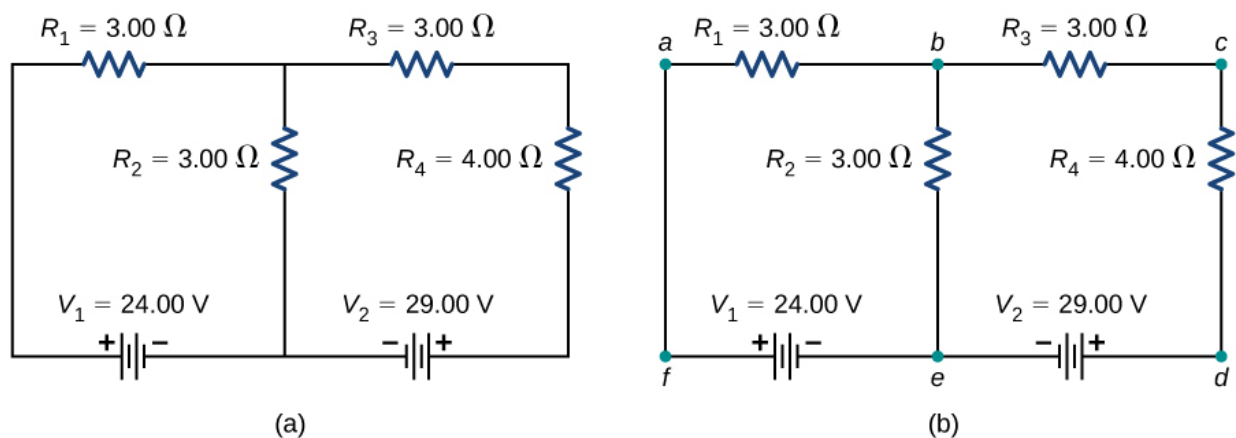
At first glance, this circuit contains two junctions, Junction b and Junction e , but only one should be considered because their junction equations are equivalent.

When choosing the loops in the circuit, you need enough loops so that each component is covered once, without repeating loops. [\[link\]](#) shows four choices for loops to solve a sample circuit; choices (a), (b), and (c) have a sufficient amount of loops to solve the circuit completely. Option (d) reflects more loops than necessary to solve the circuit.



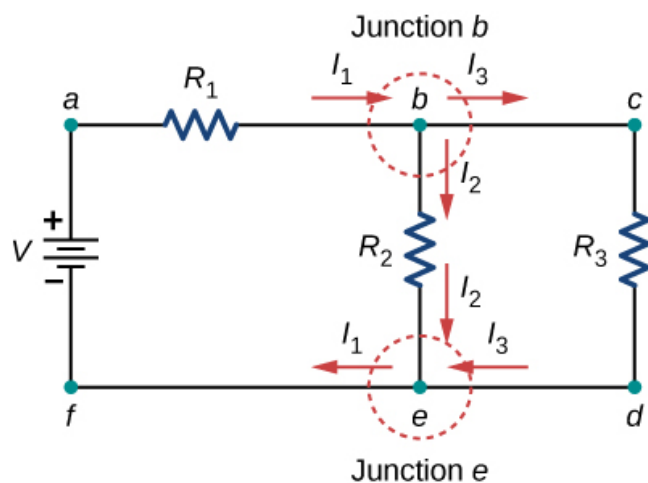
Panels (a)–(c) are sufficient for the analysis of the circuit. In each case, the two loops shown contain all the circuit elements necessary to solve the circuit completely. Panel (d) shows three loops used, which is more than necessary. Any two loops in the system will contain all information needed to solve the circuit. Adding the third loop provides redundant information.

Consider the circuit in [\[link\]](#)(a). Let us analyze this circuit to find the current through each resistor. First, label the circuit as shown in part (b).

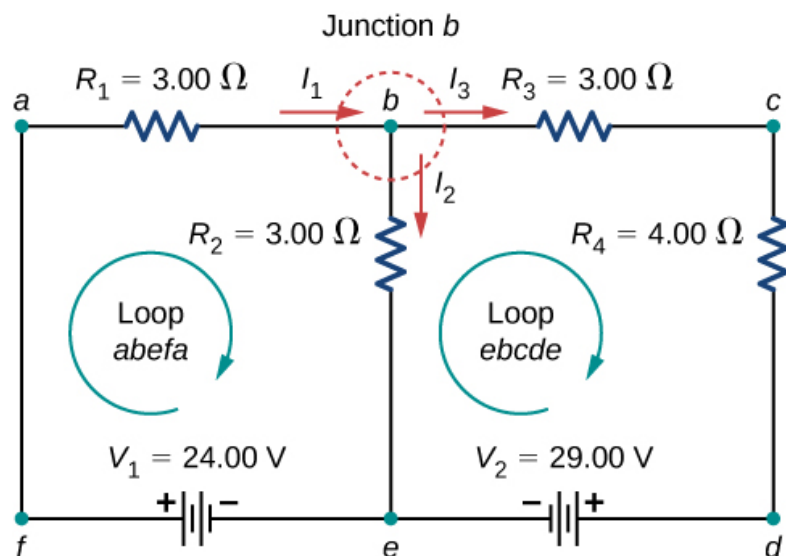


(a) A multi-loop circuit. (b) Label the circuit to help with orientation.

Next, determine the junctions. In this circuit, points b and e each have three wires connected, making them junctions. Start to apply Kirchhoff's junction rule ($\sum I_{\text{in}} = \sum I_{\text{out}}$) by drawing arrows representing the currents and labeling each arrow, as shown in [link](#)(b). Junction b shows that $I_1 = I_2 + I_3$ and Junction e shows that $I_2 + I_3 = I_1$. Since Junction e gives the same information of Junction b , it can be disregarded. This circuit has three unknowns, so we need three linearly independent equations to analyze it.



Next we need to choose the loops. In [\[link\]](#), Loop *abefa* includes the voltage source V_1 and resistors R_1 and R_2 . The loop starts at point *a*, then travels through points *b*, *e*, and *f*, and then back to point *a*. The second loop, Loop *ebcde*, starts at point *e* and includes resistors R_2 and R_3 , and the voltage source V_2 .



Choose the loops in the circuit.

Now we can apply Kirchhoff's loop rule, using the map in [\[link\]](#). Starting at point *a* and moving to point *b*, the resistor R_1 is crossed in the same direction as the current flow I_1 , so the potential drop $I_1 R_1$ is subtracted. Moving from point *b* to point *e*, the resistor R_2 is crossed in the same direction as the current flow I_2 so the potential drop $I_2 R_2$ is subtracted. Moving from point *e* to point *f*, the voltage source V_1 is crossed from the negative terminal to the positive terminal, so V_1 is added. There are no components between points *f* and *a*. The sum of the voltage differences must equal zero:

Equation:

$$\text{Loop } abefa : -I_1 R_1 - I_2 R_2 + V_1 = 0 \text{ or } V_1 = I_1 R_1 + I_2 R_2.$$

Finally, we check loop *ebcde*. We start at point *e* and move to point *b*, crossing R_2 in the opposite direction as the current flow I_2 . The potential drop $I_2 R_2$ is added. Next, we cross R_3 and R_4 in the same direction as the current flow I_3 and subtract the potential drops $I_3 R_3$ and $I_3 R_4$. Note that the current is the same through resistors R_3 and R_4 , because they are connected in series. Finally, the voltage source is crossed from the positive terminal to the negative terminal, and the voltage source V_2 is subtracted. The sum of these voltage differences equals zero and yields the loop equation

Equation:

$$\text{Loop } ebcde : I_2 R_2 - I_3 (R_3 + R_4) - V_2 = 0.$$

We now have three equations, which we can solve for the three unknowns.

Equation:

$$(1) \text{ Junction } b : I_1 - I_2 - I_3 = 0.$$

$$(2) \text{ Loop } abefa : I_1 R_1 + I_2 R_2 = V_1.$$

$$(3) \text{ Loop } ebcde : I_2 R_2 - I_3 (R_3 + R_4) = V_2.$$

To solve the three equations for the three unknown currents, start by eliminating current I_2 . First add Eq. (1) times R_2 to Eq. (2). The result is labeled as Eq. (4):

Equation:

$$(R_1 + R_2)I_1 - R_2 I_3 = V_1.$$

$$(4) 6 \Omega I_1 - 3 \Omega I_3 = 24 \text{ V}.$$

Next, subtract Eq. (3) from Eq. (2). The result is labeled as Eq. (5):

Equation:

$$I_1 R_1 + I_3 (R_3 + R_4) = V_1 - V_2.$$

$$(5) 3 \Omega I_1 + 7 \Omega I_3 = -5 \text{ V}.$$

We can solve Eqs. (4) and (5) for current I_1 . Adding seven times Eq. (4) and three times Eq. (5) results in $51 \Omega I_1 = 153 \text{ V}$, or $I_1 = 3.00 \text{ A}$. Using Eq. (4) results in $I_3 = -2.00 \text{ A}$. Finally, Eq. (1) yields $I_2 = I_1 - I_3 = 5.00 \text{ A}$. One way to check that the solutions are consistent is to check the power supplied by the voltage sources and the power dissipated by the resistors:

Equation:

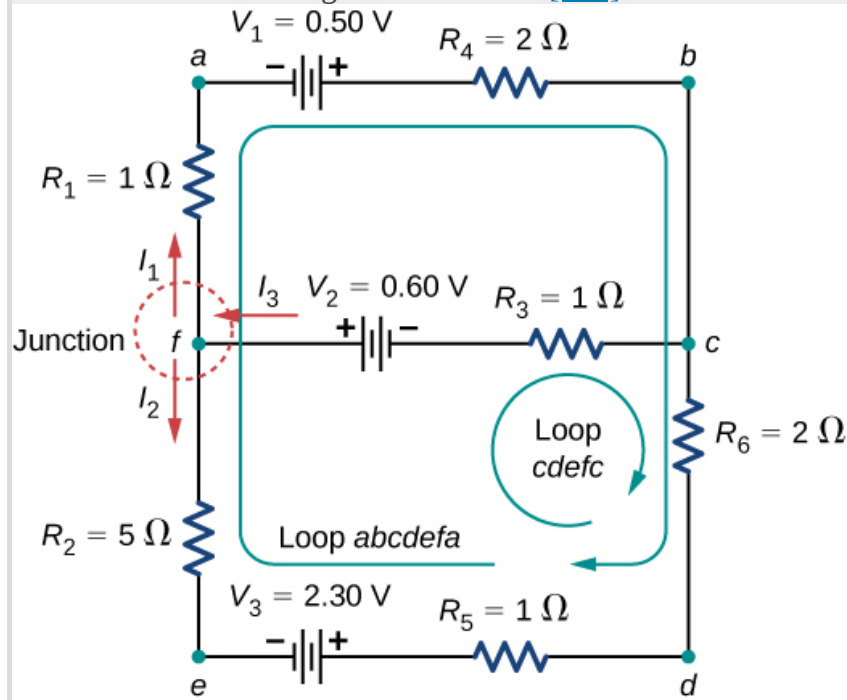
$$P_{\text{in}} = I_1 V_1 + I_3 V_2 = 130 \text{ W},$$

$$P_{\text{out}} = I_1^2 R_1 + I_2^2 R_2 + I_3^2 R_3 + I_3^2 R_4 = 130 \text{ W}.$$

Note that the solution for the current I_3 is negative. This is the correct answer, but suggests that the arrow originally drawn in the junction analysis is the direction opposite of conventional current flow. The power supplied by the second voltage source is 58 W and not -58 W.

Example:**Calculating Current by Using Kirchhoff's Rules**

Find the currents flowing in the circuit in [\[link\]](#).



This circuit is combination of series and parallel configurations of resistors and voltage sources. This circuit cannot be analyzed using the techniques discussed in [Electromotive Force](#) but can be analyzed using Kirchhoff's rules.

Strategy

This circuit is sufficiently complex that the currents cannot be found using Ohm's law and the series-parallel techniques—it is necessary to use Kirchhoff's rules. Currents have been labeled I_1 , I_2 , and I_3 in the figure, and assumptions have been made about their directions. Locations on the diagram have been labeled with letters a through h . In the solution, we apply the junction and loop rules, seeking three independent equations to allow us to solve for the three unknown currents.

Solution

Applying the junction and loop rules yields the following three equations. We have three unknowns, so three equations are required.

Equation:

$$\text{Junction } c : I_1 + I_2 = I_3.$$

$$\text{Loop } abcdefa : I_1 (R_1 + R_4) - I_2 (R_2 + R_5 + R_6) = V_1 - V_3.$$

$$\text{Loop } cdefc : I_2 (R_2 + R_5 + R_6) + I_3 R_3 = V_2 + V_3.$$

Simplify the equations by placing the unknowns on one side of the equations.

Equation:

$$\text{Junction } c : I_1 + I_2 - I_3 = 0.$$

$$\text{Loop } abcdefa : I_1 (3 \Omega) - I_2 (8 \Omega) = 0.5 \text{ V} - 2.30 \text{ V}.$$

$$\text{Loop } cdefc : I_2 (8 \Omega) + I_3 (1 \Omega) = 0.6 \text{ V} + 2.30 \text{ V}.$$

Simplify the equations. The first loop equation can be simplified by dividing both sides by 3.00. The second loop equation can be simplified by dividing both sides by 6.00.

Equation:

$$\text{Junction } c : I_1 + I_2 - I_3 = 0.$$

$$\text{Loop } abcdefa : I_1 (3 \Omega) - I_2 (8 \Omega) = -1.8 \text{ V}.$$

$$\text{Loop } cdefc : I_2 (8 \Omega) + I_3 (1 \Omega) = 2.9 \text{ V}.$$

The results are

Equation:

$$I_1 = 0.20 \text{ A}, \quad I_2 = 0.30 \text{ A}, \quad I_3 = 0.50 \text{ A}.$$

Significance

A method to check the calculations is to compute the power dissipated by the resistors and the power supplied by the voltage sources:

Equation:

$$P_{R_1} = I_1^2 R_1 = 0.04 \text{ W}.$$

$$P_{R_2} = I_2^2 R_2 = 0.45 \text{ W}.$$

$$P_{R_3} = I_3^2 R_3 = 0.25 \text{ W}.$$

$$P_{R_4} = I_1^2 R_4 = 0.08 \text{ W}.$$

$$P_{R_5} = I_2^2 R_5 = 0.09 \text{ W}.$$

$$P_{R_6} = I_2^2 R_6 = 0.18 \text{ W}.$$

$$P_{\text{dissipated}} = 1.09 \text{ W}.$$

$$P_{\text{source}} = I_1 V_1 + I_2 V_3 + I_3 V_2 = 0.10 \text{ W} + 0.69 \text{ W} + 0.30 \text{ W} = 1.09 \text{ W}.$$

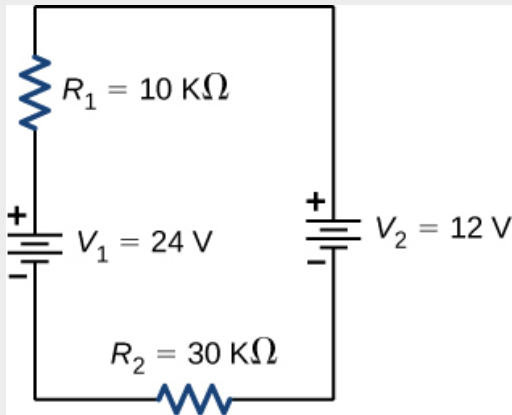
The power supplied equals the power dissipated by the resistors.

Note:

Exercise:

Problem:

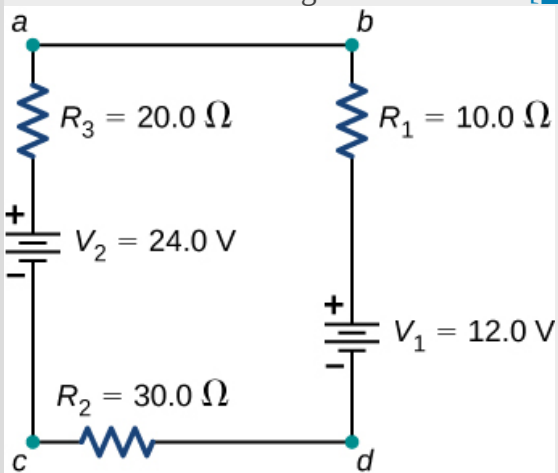
Check Your Understanding In considering the following schematic and the power supplied and consumed by a circuit, will a voltage source always provide power to the circuit, or can a voltage source consume power?

**Solution:**

The circuit can be analyzed using Kirchhoff's loop rule. The first voltage source supplies power: $P_{\text{in}} = IV_1 = 7.20\text{ mW}$. The second voltage source consumes power: $P_{\text{out}} = IV_2 + I^2R_1 + I^2R_2 = 7.2\text{ mW}$.

Example:**Calculating Current by Using Kirchhoff's Rules**

Find the current flowing in the circuit in [\[link\]](#).



This circuit consists of three resistors and two batteries connected in series.

Note that the batteries are connected with opposite polarities.

Strategy

This circuit can be analyzed using Kirchhoff's rules. There is only one loop and no nodes. Choose the direction of current flow. For this example, we will use the clockwise direction from point *a* to point *b*. Consider Loop *abcd*a and use [\[link\]](#) to write the loop equation. Note that according to [\[link\]](#), battery V_1 will be added and battery V_2 will be subtracted.

Solution

Applying the junction rule yields the following three equations. We have one unknown, so one equation is required:

Equation:

$$\text{Loop } abcd : -IR_1 - V_1 - IR_2 + V_2 - IR_3 = 0.$$

Simplify the equations by placing the unknowns on one side of the equations. Use the values given in the figure.

Equation:

$$I(R_1 + R_2 + R_3) = V_2 - V_1.$$

$$I = \frac{V_2 - V_1}{R_1 + R_2 + R_3} = \frac{24 \text{ V} - 12 \text{ V}}{10.0 \, \Omega + 30.0 \, \Omega + 10.0 \, \Omega} = 0.20 \text{ A}.$$

Significance

The power dissipated or consumed by the circuit equals the power supplied to the circuit, but notice that the current in the battery V_1 is flowing through the battery from the positive terminal to the negative terminal and consumes power.

Equation:

$$P_{R_1} = I^2 R_1 = 0.40 \text{ W}$$

$$P_{R_2} = I^2 R_2 = 1.20 \text{ W}$$

$$P_{R_3} = I^2 R_3 = 0.80 \text{ W}$$

$$P_{V_1} = IV_1 = 2.40 \text{ W}$$

$$P_{\text{dissipated}} = 4.80 \text{ W}$$

$$P_{\text{source}} = IV_2 = 4.80 \text{ W}$$

The power supplied equals the power dissipated by the resistors and consumed by the battery V_1 .

Note:

Exercise:

Problem:

Check Your Understanding When using Kirchhoff's laws, you need to decide which loops to use and the direction of current flow through each loop. In analyzing the circuit in [\[link\]](#), the direction of current flow was chosen to be clockwise, from point a to point b . How would the results change if the direction of the current was chosen to be counterclockwise, from point b to point a ?

Solution:

The current calculated would be equal to $I = -0.20 \text{ A}$ instead of $I = 0.20 \text{ A}$. The sum of the power dissipated and the power consumed would still equal the power supplied.

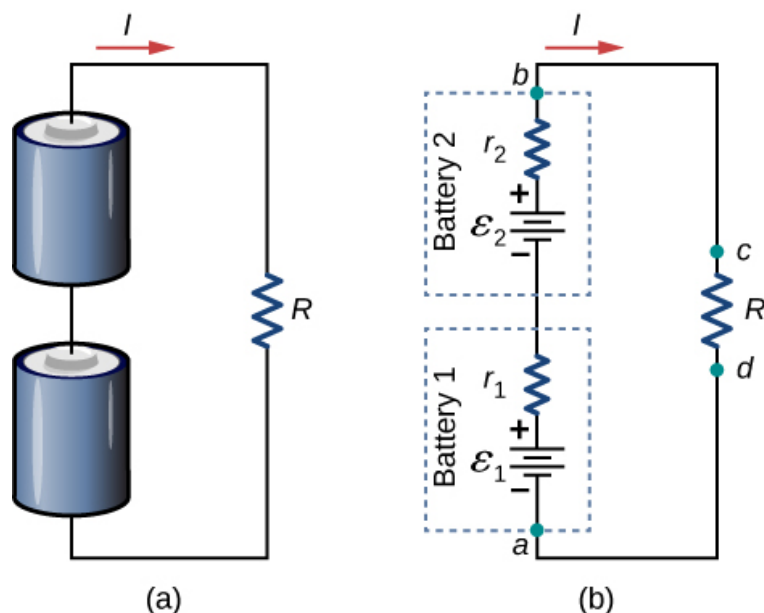
Multiple Voltage Sources

Many devices require more than one battery. Multiple voltage sources, such as batteries, can be connected in series configurations, parallel configurations, or a combination of the two.

In series, the positive terminal of one battery is connected to the negative terminal of another battery. Any number of voltage sources, including batteries, can be connected in series. Two batteries connected in series are shown in [\[link\]](#). Using Kirchhoff's loop rule for the circuit in part (b) gives the result

Equation:

$$\begin{aligned}\varepsilon_1 - Ir_1 + \varepsilon_2 - Ir_2 - IR &= 0, \\ [(\varepsilon_1 + \varepsilon_2) - I(r_1 + r_2)] - IR &= 0.\end{aligned}$$



(a) Two batteries connected in series with a load resistor. (b) The circuit diagram of the two batteries and the load resistor, with each battery modeled as an idealized emf source and an internal resistance.

When voltage sources are in series, their internal resistances can be added together and their emfs can be added together to get the total values. Series connections of voltage sources are common—for example, in flashlights, toys, and other appliances. Usually, the cells are in series in order to produce a larger total emf. In [\[link\]](#), the terminal voltage is

Equation:

$$V_{\text{terminal}} = (\varepsilon_1 - Ir_1) + (\varepsilon_2 - Ir_2) = [(\varepsilon_1 + \varepsilon_2) - I(r_1 + r_2)] = (\varepsilon_1 + \varepsilon_2) + Ir_{\text{eq}}.$$

Note that the same current I is found in each battery because they are connected in series. The disadvantage of series connections of cells is that their internal resistances are additive.

Batteries are connected in series to increase the voltage supplied to the circuit. For instance, an LED flashlight may have two AAA cell batteries, each with a terminal voltage of 1.5 V, to provide 3.0 V to the flashlight.

Any number of batteries can be connected in series. For N batteries in series, the terminal voltage is equal to

Note:

Equation:

$$V_{\text{terminal}} = (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{N-1} + \varepsilon_N) - I(r_1 + r_2 + \cdots + r_{N-1} + r_N) = \sum_{i=1}^N \varepsilon_i - I r_{\text{eq}}$$

where the equivalent resistance is $r_{\text{eq}} = \sum_{i=1}^N r_i$.

When a load is placed across voltage sources in series, as in [\[link\]](#), we can find the current:

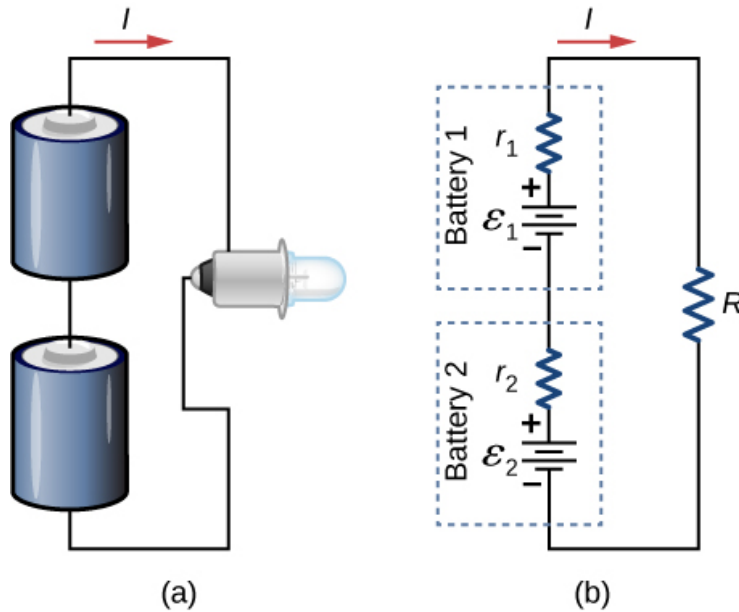
Equation:

$$(\varepsilon_1 - I r_1) + (\varepsilon_2 - I r_2) = I R,$$

$$I r_1 + I r_2 + I R = \varepsilon_1 + \varepsilon_2,$$

$$I = \frac{\varepsilon_1 + \varepsilon_2}{r_1 + r_2 + R}.$$

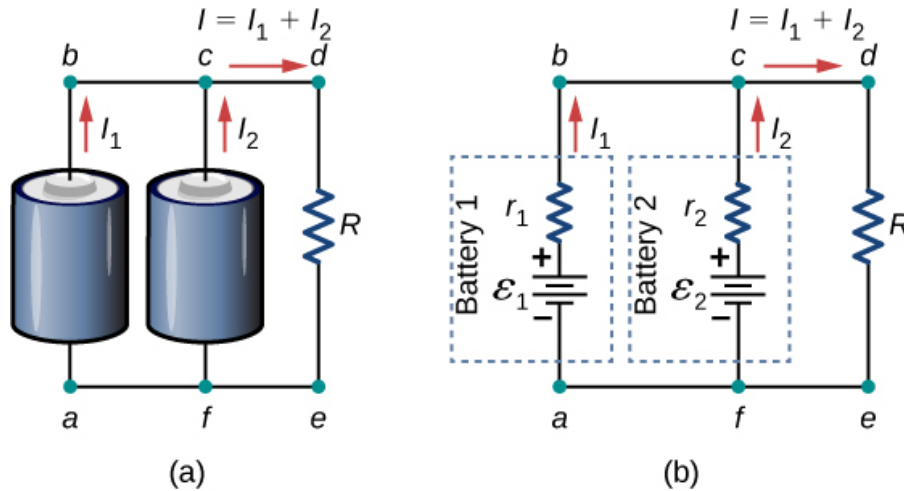
As expected, the internal resistances increase the equivalent resistance.



Two batteries connect in series to an LED bulb, as found in a flashlight.

Voltage sources, such as batteries, can also be connected in parallel. [\[link\]](#) shows two batteries with identical emfs in parallel and connected to a load resistance. When the batteries are

connect in parallel, the positive terminals are connected together and the negative terminals are connected together, and the load resistance is connected to the positive and negative terminals. Normally, voltage sources in parallel have identical emfs. In this simple case, since the voltage sources are in parallel, the total emf is the same as the individual emfs of each battery.



(a) Two batteries connect in parallel to a load resistor. (b) The circuit diagram shows the battery as an emf source and an internal resistor. The two emf sources have identical emfs (each labeled by ε) connected in parallel that produce the same emf.

Consider the Kirchhoff analysis of the circuit in [link](#)(b). There are two loops and a node at point b and $\varepsilon = \varepsilon_1 = \varepsilon_2$.

Node b : $I_1 + I_2 - I = 0$.

$$\begin{aligned} \text{Loop } abcfa: \quad \varepsilon - I_1 r_1 + I_2 r_2 - \varepsilon &= 0, \\ I_1 r_1 &= I_2 r_2. \end{aligned}$$

$$\begin{aligned} \text{Loop } fcdef: \quad \varepsilon_2 - I_2 r_2 - IR &= 0, \\ \varepsilon - I_2 r_2 - IR &= 0. \end{aligned}$$

Solving for the current through the load resistor results in $I = \frac{\varepsilon}{r_{\text{eq}} + R}$, where

$r_{\text{eq}} = \left(\frac{1}{r_1} + \frac{1}{r_2} \right)^{-1}$. The terminal voltage is equal to the potential drop across the load resistor $IR = \left(\frac{\varepsilon}{r_{\text{eq}} + R} \right)$. The parallel connection reduces the internal resistance and thus can produce a larger current.

Any number of batteries can be connected in parallel. For N batteries in parallel, the terminal voltage is equal to

Note:

Equation:

$$V_{\text{terminal}} = \varepsilon - I \left(\frac{1}{r_1} + \frac{1}{r_2} + \cdots + \frac{1}{r_{N-1}} + \frac{1}{r_N} \right)^{-1} = \varepsilon - I r_{\text{eq}}$$

where the equivalent resistance is $r_{\text{eq}} = \sum_{i=1}^N \frac{1}{r_i}^{-1}$.

As an example, some diesel trucks use two 12-V batteries in parallel; they produce a total emf of 12 V but can deliver the larger current needed to start a diesel engine.

In summary, the terminal voltage of batteries in series is equal to the sum of the individual emfs minus the sum of the internal resistances times the current. When batteries are connected in parallel, they usually have equal emfs and the terminal voltage is equal to the emf minus the equivalent internal resistance times the current, where the equivalent internal resistance is smaller than the individual internal resistances. Batteries are connected in series to increase the terminal voltage to the load. Batteries are connected in parallel to increase the current to the load.

Solar Cell Arrays

Another example dealing with multiple voltage sources is that of combinations of solar cells—wired in both series and parallel combinations to yield a desired voltage and current. Photovoltaic generation, which is the conversion of sunlight directly into electricity, is based upon the photoelectric effect. The photoelectric effect is beyond the scope of this chapter and is covered in [Photons and Matter Waves](#), but in general, photons hitting the surface of a solar cell create an electric current in the cell.

Most solar cells are made from pure silicon. Most single cells have a voltage output of about 0.5 V, while the current output is a function of the amount of sunlight falling on the cell (the incident solar radiation known as the insolation). Under bright noon sunlight, a current per unit area of about 100 mA/cm^2 of cell surface area is produced by typical single-crystal cells.

Individual solar cells are connected electrically in modules to meet electrical energy needs. They can be wired together in series or in parallel—connected like the batteries discussed earlier. A solar-cell array or module usually consists of between 36 and 72 cells, with a power output of 50 W to 140 W.

Solar cells, like batteries, provide a direct current (dc) voltage. Current from a dc voltage source is unidirectional. Most household appliances need an alternating current (ac) voltage.

Summary

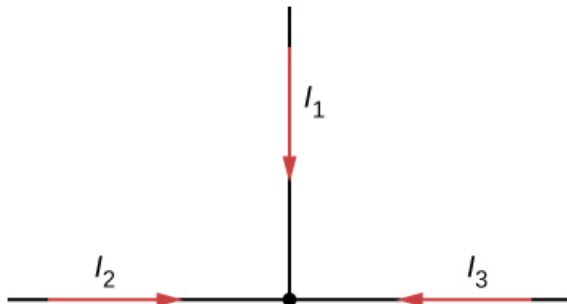
- Kirchhoff's rules can be used to analyze any circuit, simple or complex. The simpler series and parallel connection rules are special cases of Kirchhoff's rules.
- Kirchhoff's first rule, also known as the junction rule, applies to the charge to a junction. Current is the flow of charge; thus, whatever charge flows into the junction must flow out.
- Kirchhoff's second rule, also known as the loop rule, states that the voltage drop around a loop is zero.
- When calculating potential and current using Kirchhoff's rules, a set of conventions must be followed for determining the correct signs of various terms.
- When multiple voltage sources are in series, their internal resistances add together and their emfs add together to get the total values.
- When multiple voltage sources are in parallel, their internal resistances combine to an equivalent resistance that is less than the individual resistance and provides a higher current than a single cell.
- Solar cells can be wired in series or parallel to provide increased voltage or current, respectively.

Conceptual Questions

Exercise:

Problem:

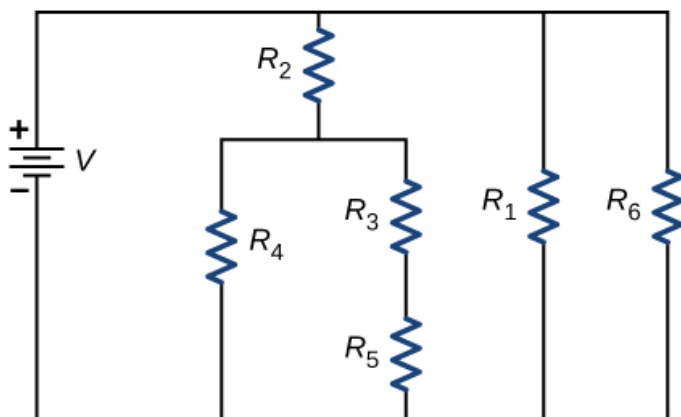
Can all of the currents going into the junction shown below be positive? Explain.



Exercise:

Problem:

Consider the circuit shown below. Does the analysis of the circuit require Kirchhoff's method, or can it be redrawn to simplify the circuit? If it is a circuit of series and parallel connections, what is the equivalent resistance?



Solution:

It can be redrawn.

$$R_{eq} = \frac{1}{R_6} + \frac{1}{R_1} + \frac{1}{R_2 + \left(\frac{1}{R_4} + \frac{1}{R_3 + R_5} \right)^{-1}} \quad -1$$

Exercise:

Problem:

Do batteries in a circuit always supply power to a circuit, or can they absorb power in a circuit? Give an example.

Exercise:

Problem:

What are the advantages and disadvantages of connecting batteries in series? In parallel?

Solution:

In series the voltages add, but so do the internal resistances, because the internal resistances are in series. In parallel, the terminal voltage is the same, but the equivalent internal resistance is smaller than the smallest individual internal resistance and a higher current can be provided.

Exercise:

Problem:

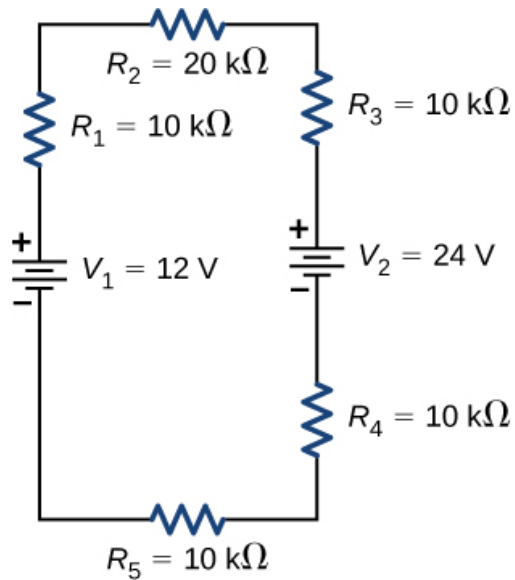
Semi-tractor trucks use four large 12-V batteries. The starter system requires 24 V, while normal operation of the truck's other electrical components utilizes 12 V. How could the four batteries be connected to produce 24 V? To produce 12 V? Why is 24 V better than 12 V for starting the truck's engine (a very heavy load)?

Problems

Exercise:

Problem:

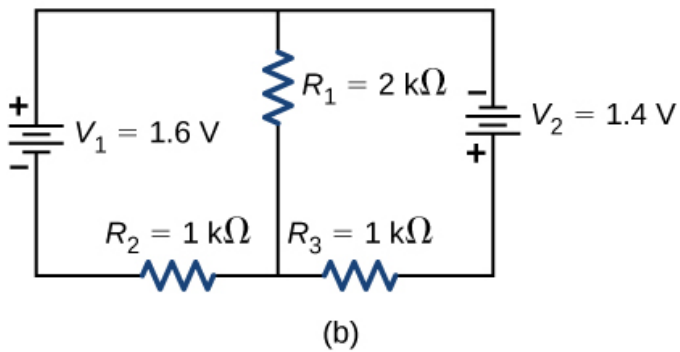
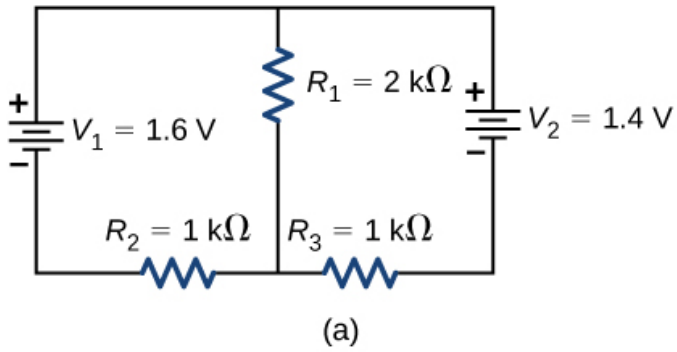
Consider the circuit shown below. (a) Find the voltage across each resistor. (b) What is the power supplied to the circuit and the power dissipated or consumed by the circuit?



Exercise:

Problem:

Consider the circuits shown below. (a) What is the current through each resistor in part (a)? (b) What is the current through each resistor in part (b)? (c) What is the power dissipated or consumed by each circuit? (d) What is the power supplied to each circuit?

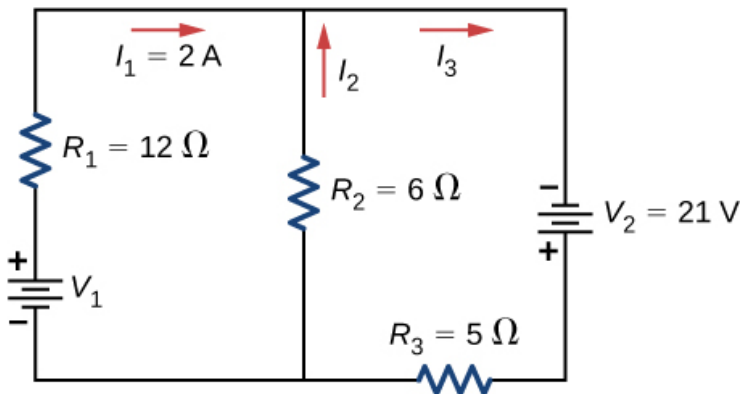


Solution:

- a. $I_1 = 0.6\text{ mA}$, $I_2 = 0.4\text{ mA}$, $I_3 = 0.2\text{ mA}$;
 b. $I_1 = 0.04\text{ mA}$, $I_2 = 1.52\text{ mA}$, $I_3 = -1.48\text{ mA}$; c.
 $P_{\text{out}} = 0.92\text{ mW}$, $P_{\text{out}} = 4.50\text{ mW}$;
 d. $P_{\text{in}} = 0.92\text{ mW}$, $P_{\text{in}} = 4.50\text{ mW}$

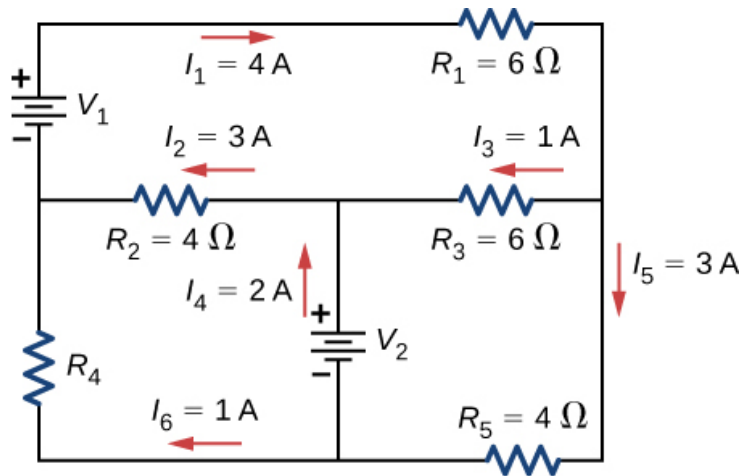
Exercise:

Problem: Consider the circuit shown below. Find V_1 , I_2 , and I_3 .



Exercise:

Problem: Consider the circuit shown below. Find V_1 , V_2 , and R_4 .

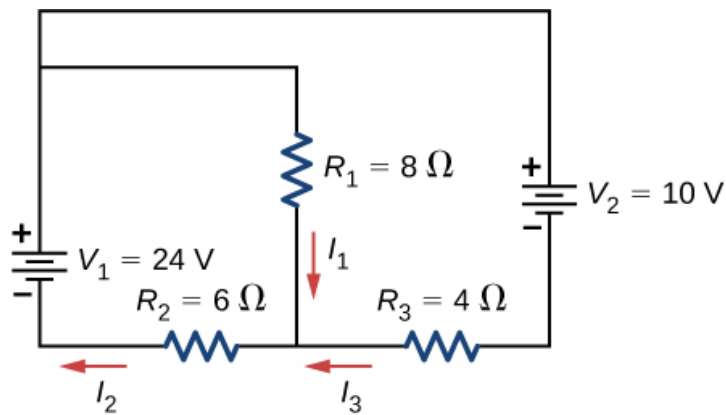


Solution:

$$V_1 = 42 \text{ V}, V_2 = 6 \text{ V}, R_4 = 18 \Omega$$

Exercise:

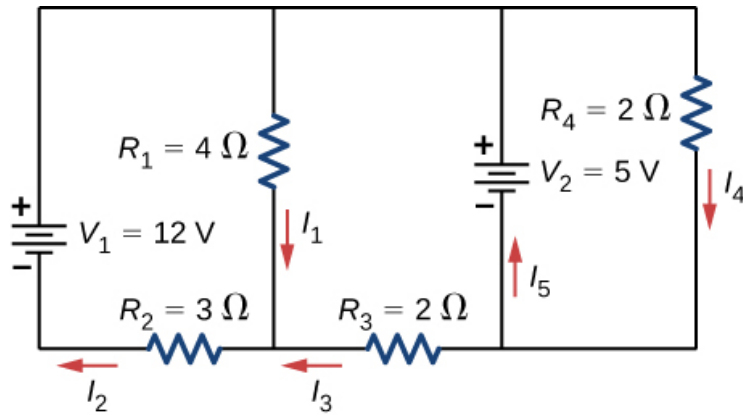
Problem: Consider the circuit shown below. Find I_1 , I_2 , and I_3 .



Exercise:

Problem:

Consider the circuit shown below. (a) Find I_1 , I_2 , I_3 , I_4 , and I_5 . (b) Find the power supplied by the voltage sources. (c) Find the power dissipated by the resistors.



Solution:

a. $I_1 = 1.5 \text{ A}$, $I_2 = 2 \text{ A}$, $I_3 = 0.5 \text{ A}$, $I_4 = 2.5 \text{ A}$, $I_5 = 2 \text{ A}$; b.

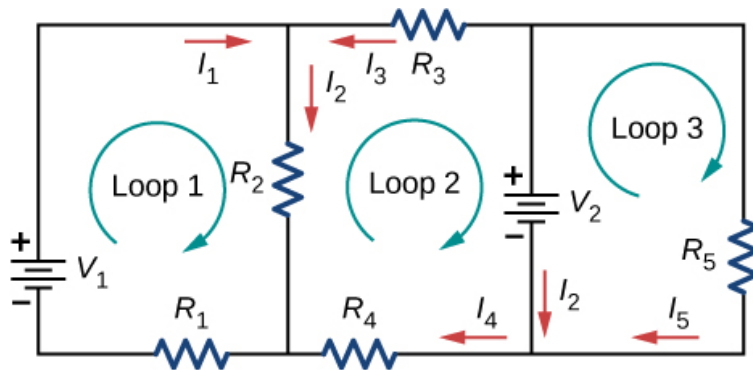
$P_{\text{in}} = I_2 V_1 + I_5 V_5 = 34 \text{ W}$;

c. $P_{\text{out}} = I_1^2 R_1 + I_2^2 R_2 + I_3^2 R_3 + I_4^2 R_4 = 34 \text{ W}$

Exercise:

Problem:

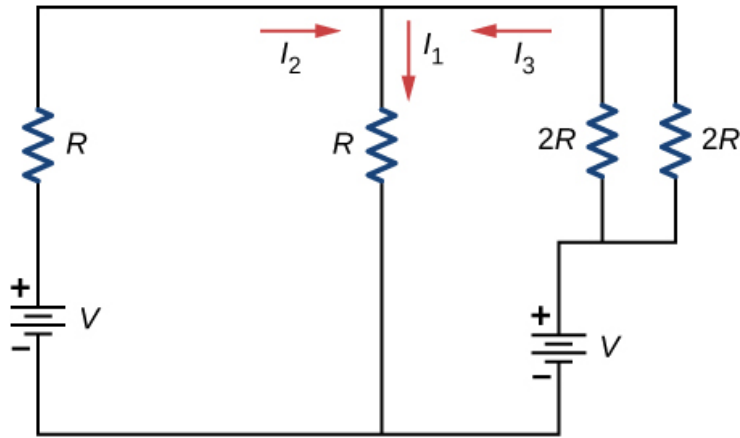
Consider the circuit shown below. Write the three loop equations for the loops shown.



Exercise:

Problem:

Consider the circuit shown below. Write equations for the three currents in terms of R and V .



Solution:

$$I_1 = \frac{2}{3} \frac{V}{R}, I_2 = \frac{V}{3R}, I_3 = \frac{V}{3R}$$

Exercise:

Problem:

Consider the circuit shown in the preceding problem. Write equations for the power supplied by the voltage sources and the power dissipated by the resistors in terms of R and V .

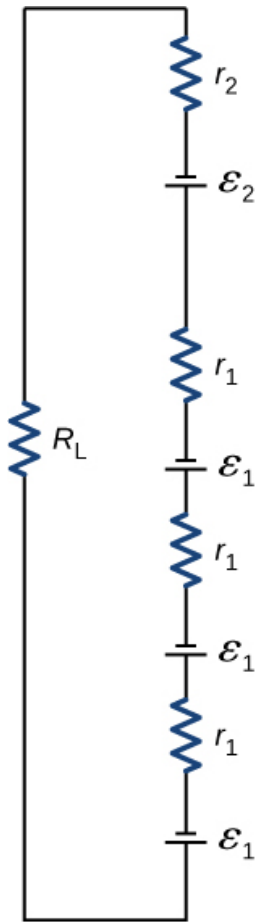
Exercise:

Problem:

A child's electronic toy is supplied by three 1.58-V alkaline cells having internal resistances of $0.0200\ \Omega$ in series with a 1.53-V carbon-zinc dry cell having a $0.100\text{-}\Omega$ internal resistance. The load resistance is $10.0\ \Omega$. (a) Draw a circuit diagram of the toy and its batteries. (b) What current flows? (c) How much power is supplied to the load? (d) What is the internal resistance of the dry cell if it goes bad, resulting in only 0.500 W being supplied to the load?

Solution:

a.

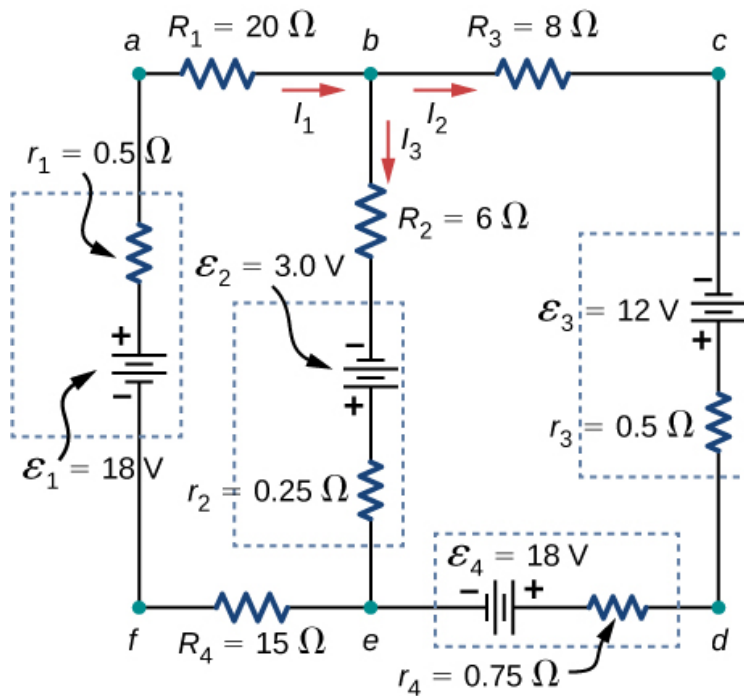


; b. 0.617 A; c. 3.81 W; d. $18.0\ \Omega$

Exercise:

Problem:

Apply the junction rule to Junction *b* shown below. Is any new information gained by applying the junction rule at *e*?



Exercise:

Problem: Apply the loop rule to Loop *afedcba* in the preceding problem.

Solution:

$$I_1 r_1 - \varepsilon_1 + I_1 R_4 + \varepsilon_4 + I_2 r_4 + I_4 r_3 - \varepsilon_3 + I_2 R_3 + I_1 R_1 = 0$$

Glossary

junction rule

sum of all currents entering a junction must equal the sum of all currents leaving the junction

Kirchhoff's rules

set of two rules governing current and changes in potential in an electric circuit

loop rule

algebraic sum of changes in potential around any closed circuit path (loop) must be zero

Electrical Measuring Instruments

By the end of the section, you will be able to:

- Describe how to connect a voltmeter in a circuit to measure voltage
- Describe how to connect an ammeter in a circuit to measure current
- Describe the use of an ohmmeter

Ohm's law and Kirchhoff's method are useful to analyze and design electrical circuits, providing you with the voltages across, the current through, and the resistance of the components that compose the circuit. To measure these parameters require instruments, and these instruments are described in this section.

DC Voltmeters and Ammeters

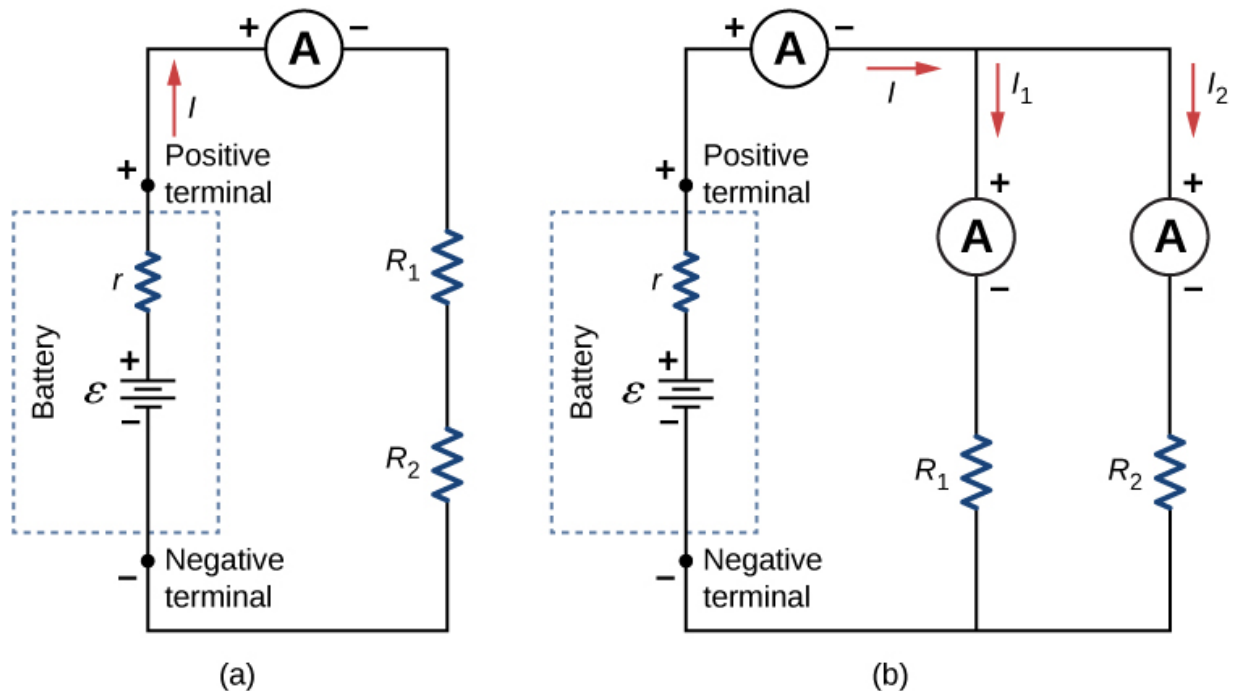
Whereas **voltmeters** measure voltage, **ammeters** measure current. Some of the meters in automobile dashboards, digital cameras, cell phones, and tuner-amplifiers are actually voltmeters or ammeters ([\[link\]](#)). The internal construction of the simplest of these meters and how they are connected to the system they monitor give further insight into applications of series and parallel connections.



The fuel and temperature gauges (far right and far left, respectively) in this 1996 Volkswagen are voltmeters that register the voltage output of “sender” units. These units are proportional to the amount of gasoline in the tank and to the engine temperature. (credit: Christian Giersing)

Measuring Current with an Ammeter

To measure the current through a device or component, the ammeter is placed in series with the device or component. A series connection is used because objects in series have the same current passing through them. (See [\[link\]](#), where the ammeter is represented by the symbol A.)

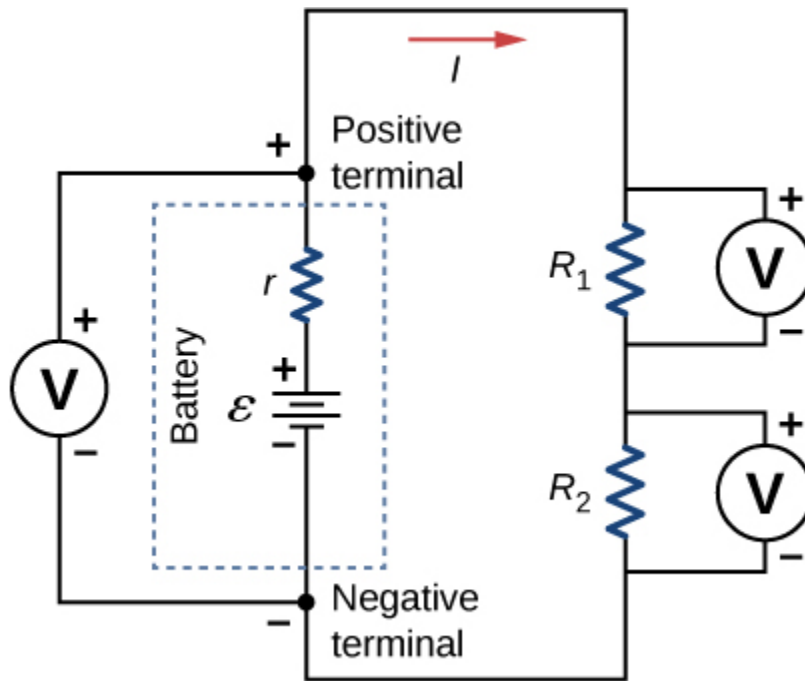


(a) When an ammeter is used to measure the current through two resistors connected in series to a battery, a single ammeter is placed in series with the two resistors because the current is the same through the two resistors in series. (b) When two resistors are connected in parallel with a battery, three meters, or three separate ammeter readings, are necessary to measure the current from the battery and through each resistor. The ammeter is connected in series with the component in question.

Ammeters need to have a very low resistance, a fraction of a milliohm. If the resistance is not negligible, placing the ammeter in the circuit would change the equivalent resistance of the circuit and modify the current that is being measured. Since the current in the circuit travels through the meter, ammeters normally contain a fuse to protect the meter from damage from currents which are too high.

Measuring Voltage with a Voltmeter

A voltmeter is connected in parallel with whatever device it is measuring. A parallel connection is used because objects in parallel experience the same potential difference. (See [\[link\]](#), where the voltmeter is represented by the symbol V.)



To measure potential differences in this series circuit, the voltmeter (V) is placed in parallel with the voltage source or either of the resistors. Note that terminal voltage is measured between the positive terminal and the negative terminal of the battery or voltage source. It is not possible to connect a voltmeter directly across the emf without including the internal resistance r of the battery.

Since voltmeters are connected in parallel, the voltmeter must have a very large resistance. Digital voltmeters convert the analog voltage into a digital

value to display on a digital readout ([link](#)). Inexpensive voltmeters have resistances on the order of $R_M = 10 \text{ M}\Omega$, whereas high-precision voltmeters have resistances on the order of $R_M = 10 \text{ G}\Omega$. The value of the resistance may vary, depending on which scale is used on the meter.



(a)



(b)

(a) An analog voltmeter uses a galvanometer to measure the voltage.

(b) Digital meters use an analog-to-digital converter to measure the voltage. (credit: modification of works by Joseph J. Trout)

Analog and Digital Meters

You may encounter two types of meters in the physics lab: analog and digital. The term 'analog' refers to signals or information represented by a continuously variable physical quantity, such as voltage or current. An analog meter uses a galvanometer, which is essentially a coil of wire with a small resistance, in a magnetic field, with a pointer attached that points to a scale. Current flows through the coil, causing the coil to rotate. To use the galvanometer as an ammeter, a small resistance is placed in parallel with the coil. For a voltmeter, a large resistance is placed in series with the coil. A

digital meter uses a component called an analog-to-digital (A to D) converter and expresses the current or voltage as a series of the digits 0 and 1, which are used to run a digital display. Most analog meters have been replaced by digital meters.

Note:

Exercise:

Problem:

Check Your Understanding Digital meters are able to detect smaller currents than analog meters employing galvanometers. How does this explain their ability to measure voltage and current more accurately than analog meters?

Solution:

Since digital meters require less current than analog meters, they alter the circuit less than analog meters. Their resistance as a voltmeter can be far greater than an analog meter, and their resistance as an ammeter can be far less than an analog meter. Consult [\[link\]](#) and [\[link\]](#) and their discussion in the text.

Note:

In this [virtual lab](#) simulation, you may construct circuits with resistors, voltage sources, ammeters and voltmeters to test your knowledge of circuit design.

Ohmmeters

An ohmmeter is an instrument used to measure the resistance of a component or device. The operation of the ohmmeter is based on Ohm's

law. Traditional ohmmeters contained an internal voltage source (such as a battery) that would be connected across the component to be tested, producing a current through the component. A galvanometer was then used to measure the current and the resistance was deduced using Ohm's law. Modern digital meters use a constant current source to pass current through the component, and the voltage difference across the component is measured. In either case, the resistance is measured using Ohm's law ($R = V/I$), where the voltage is known and the current is measured, or the current is known and the voltage is measured.

The component of interest should be isolated from the circuit; otherwise, you will be measuring the equivalent resistance of the circuit. An ohmmeter should never be connected to a "live" circuit, one with a voltage source connected to it and current running through it. Doing so can damage the meter.

Summary

- Voltmeters measure voltage, and ammeters measure current. Analog meters are based on the combination of a resistor and a galvanometer, a device that gives an analog reading of current or voltage. Digital meters are based on analog-to-digital converters and provide a discrete or digital measurement of the current or voltage.
- A voltmeter is placed in parallel with the voltage source to receive full voltage and must have a large resistance to limit its effect on the circuit.
- An ammeter is placed in series to get the full current flowing through a branch and must have a small resistance to limit its effect on the circuit.
- Standard voltmeters and ammeters alter the circuit they are connected to and are thus limited in accuracy.
- Ohmmeters are used to measure resistance. The component in which the resistance is to be measured should be isolated (removed) from the circuit.

Conceptual Questions

Exercise:**Problem:**

What would happen if you placed a voltmeter in series with a component to be tested?

Solution:

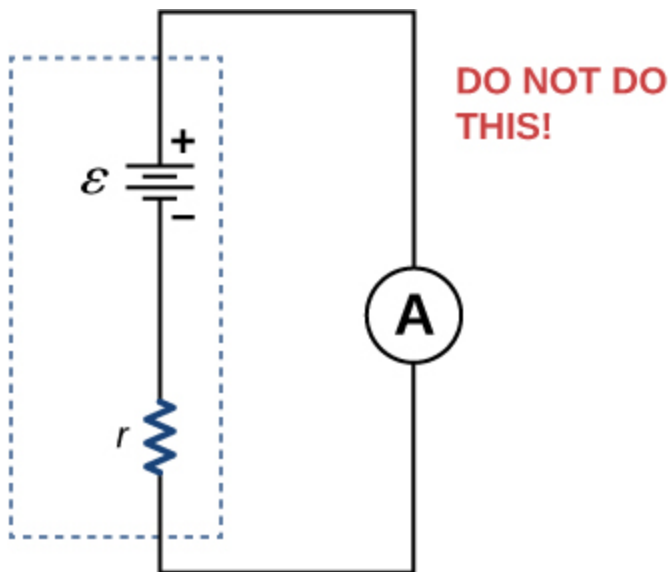
The voltmeter would put a large resistance in series with the circuit, significantly changing the circuit. It would probably give a reading, but it would be meaningless.

Exercise:**Problem:**

What is the basic operation of an ohmmeter as it measures a resistor?

Exercise:**Problem:**

Why should you not connect an ammeter directly across a voltage source as shown below?

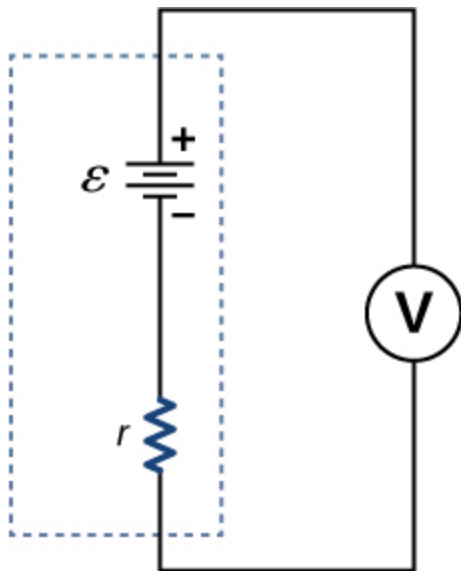


Solution:

The ammeter has a small resistance; therefore, a large current will be produced and could damage the meter and/or overheat the battery.

Problems**Exercise:****Problem:**

Suppose you measure the terminal voltage of a 1.585-V alkaline cell having an internal resistance of $0.100\ \Omega$ by placing a $1.00\text{-k}\Omega$ voltmeter across its terminals (see below). (a) What current flows? (b) Find the terminal voltage. (c) To see how close the measured terminal voltage is to the emf, calculate their ratio.

**Glossary**

ammeter

instrument that measures current

voltmeter

instrument that measures voltage

RC Circuits

By the end of the section, you will be able to:

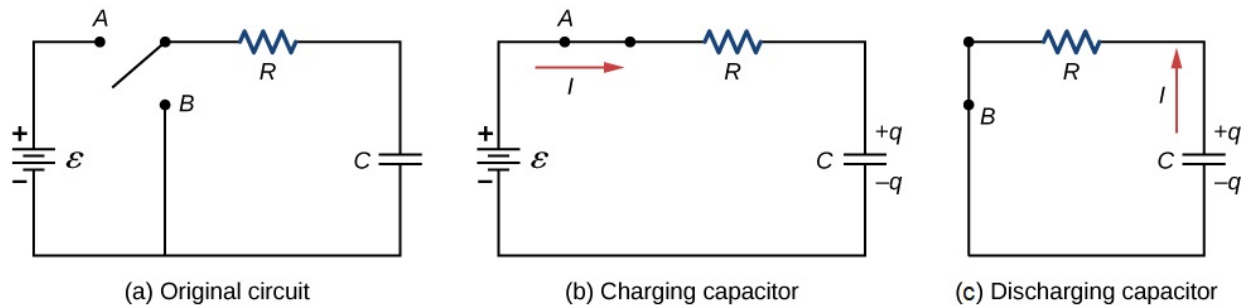
- Describe the charging process of a capacitor
- Describe the discharging process of a capacitor
- List some applications of RC circuits

When you use a flash camera, it takes a few seconds to charge the capacitor that powers the flash. The light flash discharges the capacitor in a tiny fraction of a second. Why does charging take longer than discharging? This question and several other phenomena that involve charging and discharging capacitors are discussed in this module.

Circuits with Resistance and Capacitance

An **RC circuit** is a circuit containing resistance and capacitance. As presented in [Capacitance](#), the capacitor is an electrical component that stores electric charge, storing energy in an electric field.

[\[link\]](#)(a) shows a simple RC circuit that employs a dc (direct current) voltage source \mathcal{E} , a resistor R , a capacitor C , and a two-position switch. The circuit allows the capacitor to be charged or discharged, depending on the position of the switch. When the switch is moved to position A, the capacitor charges, resulting in the circuit in part (b). When the switch is moved to position B, the capacitor discharges through the resistor.



(a) An RC circuit with a two-pole switch that can be used to charge and discharge a capacitor. (b) When the switch is moved to position A, the circuit reduces to a simple series connection of the voltage source, the resistor, the capacitor, and the switch. (c) When the switch is moved to position B, the circuit reduces to a simple series connection of the resistor, the capacitor, and the switch. The voltage source is removed from the circuit.

Charging a Capacitor

We can use Kirchhoff's loop rule to understand the charging of the capacitor. This results in the equation $\mathcal{E} - V_R - V_C = 0$. This equation can be used to model the charge as a function of time as the capacitor charges. Capacitance is defined as $C = q/V$, so the voltage across the capacitor is $V_C = \frac{q}{C}$. Using Ohm's law, the potential drop across the resistor is $V_R = IR$, and the current is defined as $I = dq/dt$.

Equation:

$$\begin{aligned}\varepsilon - V_R - V_c &= 0, \\ \varepsilon - IR - \frac{q}{C} &= 0, \\ \varepsilon - R\frac{dq}{dt} - \frac{q}{C} &= 0.\end{aligned}$$

This differential equation can be integrated to find an equation for the charge on the capacitor as a function of time.

Equation:

$$\begin{aligned}\varepsilon - R\frac{dq}{dt} - \frac{q}{C} &= 0, \\ \frac{dq}{dt} &= \frac{\varepsilon C - q}{RC}, \\ \int_0^q \frac{dq}{\varepsilon C - q} &= \frac{1}{RC} \int_0^t dt.\end{aligned}$$

Let $u = \varepsilon C - q$, then $du = -dq$. The result is

Equation:

$$\begin{aligned}-\int_0^q \frac{du}{u} &= \frac{1}{RC} \int_0^t dt, \\ \ln\left(\frac{\varepsilon C - q}{\varepsilon C}\right) &= -\frac{1}{RC}t, \\ \frac{\varepsilon C - q}{\varepsilon C} &= e^{-t/RC}.\end{aligned}$$

Simplifying results in an equation for the charge on the charging capacitor as a function of time:

Note:

Equation:

$$q(t) = C\varepsilon \left(1 - e^{-\frac{t}{RC}}\right) = Q\left(1 - e^{-\frac{t}{\tau}}\right).$$

A graph of the charge on the capacitor versus time is shown in [\[link\]](#)(a). First note that as time approaches infinity, the exponential goes to zero, so the charge approaches the maximum charge $Q = C\varepsilon$ and has units of coulombs. The units of RC are seconds, units of time. This quantity is known as the time constant:

Note:

Equation:

$$\tau = RC.$$

At time $t = \tau = RC$, the charge is equal to $1 - e^{-1} = 1 - 0.368 = 0.632$ of the maximum charge $Q = C\varepsilon$. Notice that the time rate change of the charge is the slope at a point of the charge versus time plot. The slope of the graph is large at time $t = 0.0$ s and approaches zero as time increases.

As the charge on the capacitor increases, the current through the resistor decreases, as shown in [\[link\]\(b\)](#). The current through the resistor can be found by taking the time derivative of the charge.

Equation:

$$I(t) = \frac{dq}{dt} = \frac{d}{dt} \left[C\varepsilon \left(1 - e^{-\frac{t}{RC}} \right) \right],$$

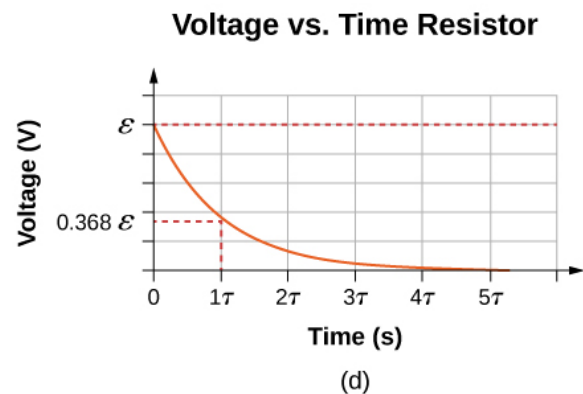
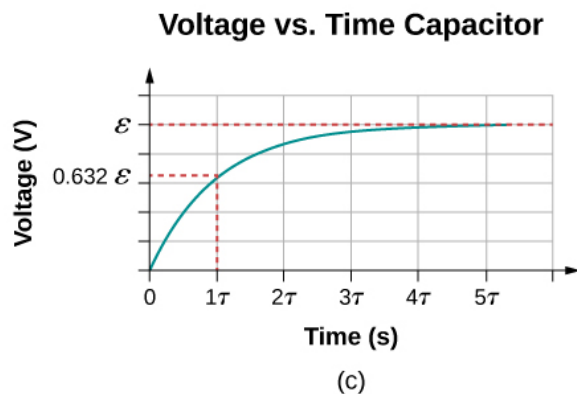
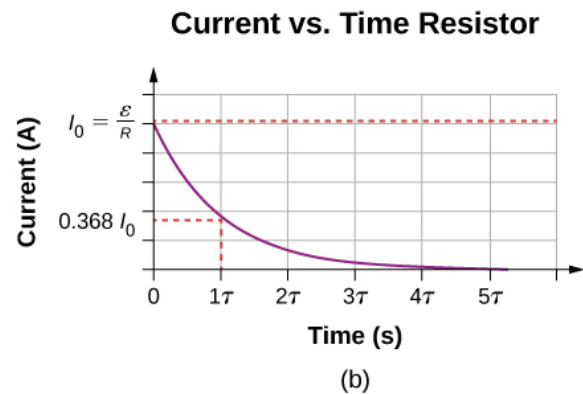
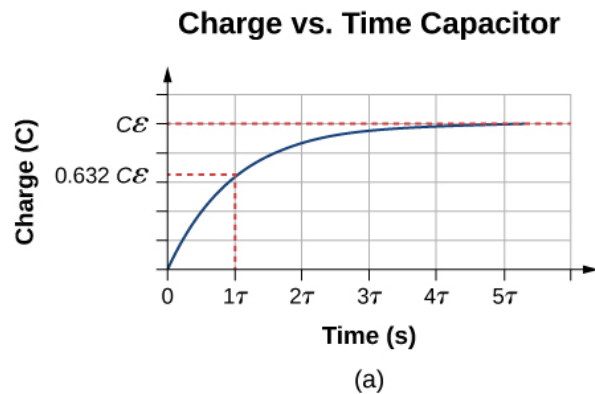
$$I(t) = C\varepsilon \left(\frac{1}{RC} \right) e^{-\frac{t}{RC}} = \frac{\varepsilon}{R} e^{-\frac{t}{RC}} = I_0 e^{-\frac{t}{RC}},$$

Note:

Equation:

$$I(t) = I_0 e^{-t/\tau}.$$

At time $t = 0.00$ s, the current through the resistor is $I_0 = \frac{\varepsilon}{R}$. As time approaches infinity, the current approaches zero. At time $t = \tau$, the current through the resistor is $I(t = \tau) = I_0 e^{-1} = 0.368 I_0$.



- (a) Charge on the capacitor versus time as the capacitor charges. (b) Current through the resistor versus time. (c) Voltage difference across the capacitor. (d) Voltage difference across the resistor.

[\[link\]](#)(c) and [\[link\]](#)(d) show the voltage differences across the capacitor and the resistor, respectively. As the charge on the capacitor increases, the current decreases, as does the voltage difference across the resistor $V_R(t) = (I_0 R)e^{-t/\tau} = \varepsilon e^{-t/\tau}$. The voltage difference across the capacitor increases as $V_C(t) = \varepsilon(1 - e^{-t/\tau})$.

Discharging a Capacitor

When the switch in [\[link\]](#)(a) is moved to position *B*, the circuit reduces to the circuit in part (c), and the charged capacitor is allowed to discharge through the resistor. A graph of the charge on the capacitor as a function of time is shown in [\[link\]](#)(a). Using Kirchhoff's loop rule to analyze the circuit as the capacitor discharges results in the equation $-V_R - V_c = 0$, which simplifies to $IR + \frac{q}{C} = 0$. Using the definition of current $\frac{dq}{dt}R = -\frac{q}{C}$ and integrating the loop equation yields an equation for the charge on the capacitor as a function of time:

Note:

Equation:

$$q(t) = Qe^{-t/\tau}.$$

Here, Q is the initial charge on the capacitor and $\tau = RC$ is the time constant of the circuit. As shown in the graph, the charge decreases exponentially from the initial charge, approaching zero as time approaches infinity.

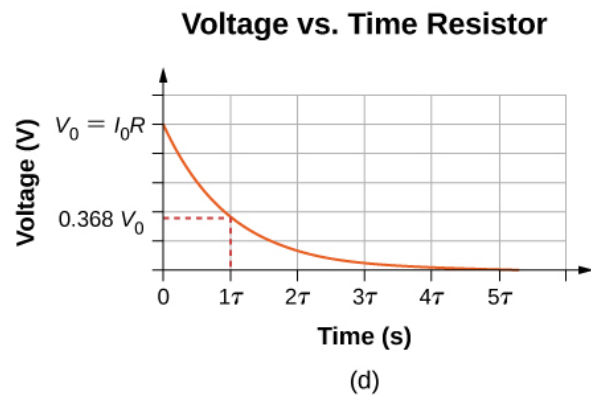
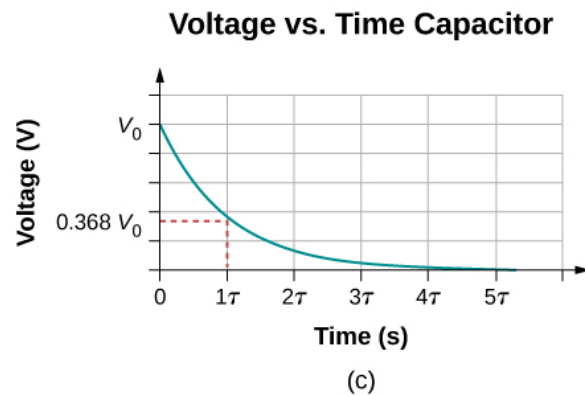
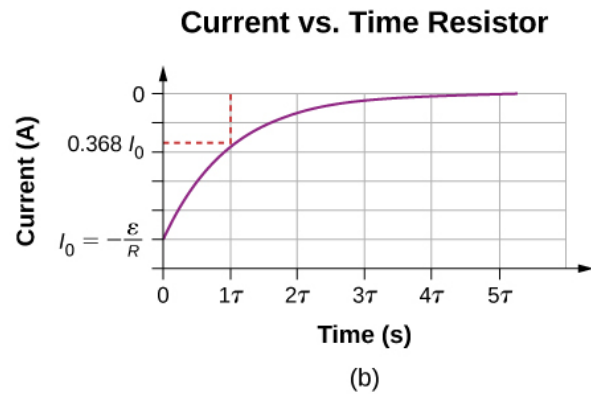
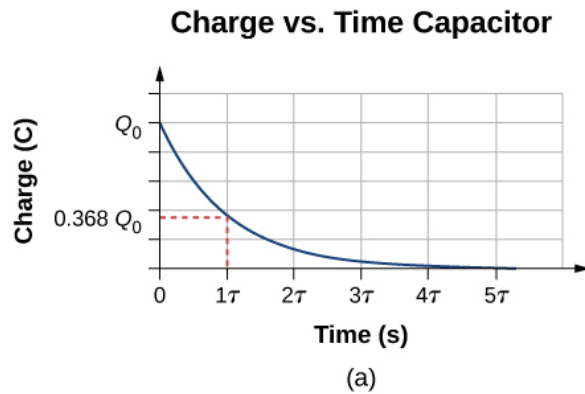
The current as a function of time can be found by taking the time derivative of the charge:

Note:

Equation:

$$I(t) = -\frac{Q}{RC}e^{-t/\tau}.$$

The negative sign shows that the current flows in the opposite direction of the current found when the capacitor is charging. [\[link\]](#)(b) shows an example of a plot of charge versus time and current versus time. A plot of the voltage difference across the capacitor and the voltage difference across the resistor as a function of time are shown in parts (c) and (d) of the figure. Note that the magnitudes of the charge, current, and voltage all decrease exponentially, approaching zero as time increases.



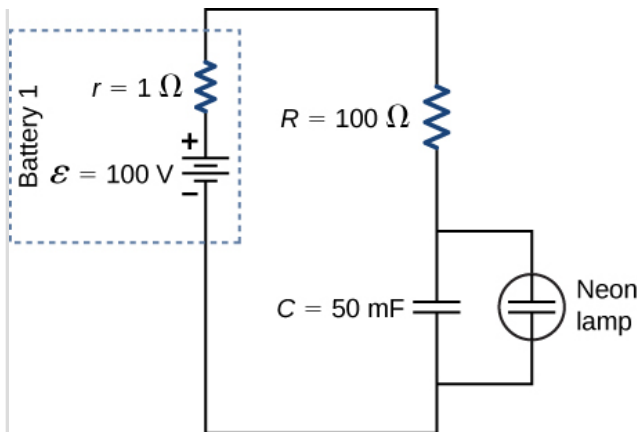
(a) Charge on the capacitor versus time as the capacitor discharges. (b) Current through the resistor versus time. (c) Voltage difference across the capacitor. (d) Voltage difference across the resistor.

Now we can explain why the flash camera mentioned at the beginning of this section takes so much longer to charge than discharge: The resistance while charging is significantly greater than while discharging. The internal resistance of the battery accounts for most of the resistance while charging. As the battery ages, the increasing internal resistance makes the charging process even slower.

Example:

The Relaxation Oscillator

One application of an RC circuit is the relaxation oscillator, as shown below. The relaxation oscillator consists of a voltage source, a resistor, a capacitor, and a neon lamp. The neon lamp acts like an open circuit (infinite resistance) until the potential difference across the neon lamp reaches a specific voltage. At that voltage, the lamp acts like a short circuit (zero resistance), and the capacitor discharges through the neon lamp and produces light. In the relaxation oscillator shown, the voltage source charges the capacitor until the voltage across the capacitor is 80 V. When this happens, the neon in the lamp breaks down and allows the capacitor to discharge through the lamp, producing a bright flash. After the capacitor fully discharges through the neon lamp, it begins to charge again, and the process repeats. Assuming that the time it takes the capacitor to discharge is negligible, what is the time interval between flashes?



Strategy

The time period can be found from considering the equation $V_C(t) = \varepsilon (1 - e^{-t/\tau})$, where $\tau = (R + r)C$.

Solution

The neon lamp flashes when the voltage across the capacitor reaches 80 V. The RC time constant is equal to $\tau = (R + r)C = (101 \Omega)(50 \times 10^{-3} \text{ F}) = 5.05 \text{ s}$. We can solve the voltage equation for the time it takes the capacitor to reach 80 V:

Equation:

$$\begin{aligned} V_C(t) &= \varepsilon (1 - e^{-t/\tau}), \\ e^{-t/\tau} &= 1 - \frac{V_C(t)}{\varepsilon}, \\ \ln(e^{-t/\tau}) &= \ln\left(1 - \frac{V_C(t)}{\varepsilon}\right), \\ t &= -\tau \ln\left(1 - \frac{V_C(t)}{\varepsilon}\right) = -5.05 \text{ s} \cdot \ln\left(1 - \frac{80 \text{ V}}{100 \text{ V}}\right) = 8.13 \text{ s}. \end{aligned}$$

Significance

One application of the relaxation oscillator is for controlling indicator lights that flash at a frequency determined by the values for R and C . In this example, the neon lamp will flash every 8.13 seconds, a frequency of $f = \frac{1}{T} = \frac{1}{8.13 \text{ s}} = 0.123 \text{ Hz}$. The relaxation oscillator has many other practical uses. It is often used in electronic circuits, where the neon lamp is replaced by a transistor or a device known as a tunnel diode. The description of the transistor and tunnel diode is beyond the scope of this chapter, but you can think of them as voltage controlled switches. They are normally open switches, but when the right voltage is applied, the switch closes and conducts. The “switch” can be used to turn on another circuit, turn on a light, or run a small motor. A relaxation oscillator can be used to make the turn signals of your car blink or your cell phone to vibrate.

RC circuits have many applications. They can be used effectively as timers for applications such as intermittent windshield wipers, pace makers, and strobe lights. Some models of intermittent windshield wipers use a variable resistor to adjust the interval between sweeps of the wiper. Increasing the resistance increases the RC time constant, which increases the time between the operation of the wipers.

Another application is the pacemaker. The heart rate is normally controlled by electrical signals, which cause the muscles of the heart to contract and pump blood. When the heart rhythm is abnormal (the heartbeat is too high or too low), pace makers can be used to correct this abnormality. Pacemakers have sensors that detect body motion and breathing to increase the heart rate during physical activities, thus meeting the increased need for blood and oxygen, and an RC timing circuit can be used to control the time between voltage signals to the heart.

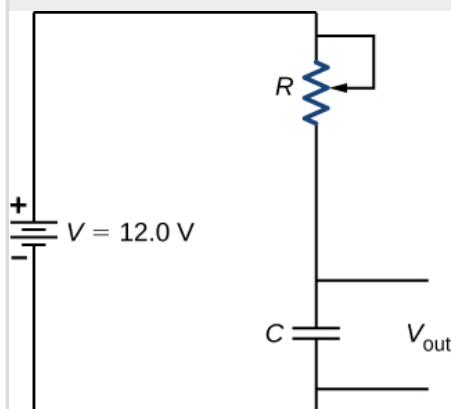
Looking ahead to the study of ac circuits ([Alternating-Current Circuits](#)), ac voltages vary as sine functions with specific frequencies. Periodic variations in voltage, or electric signals, are often recorded by scientists. These voltage signals could come from music recorded by a microphone or atmospheric data collected by radar. Occasionally, these signals can contain unwanted frequencies known as “noise.” RC filters can be used to filter out the unwanted frequencies.

In the study of electronics, a popular device known as a 555 timer provides timed voltage pulses. The time between pulses is controlled by an RC circuit. These are just a few of the countless applications of RC circuits.

Example:

Intermittent Windshield Wipers

A relaxation oscillator is used to control a pair of windshield wipers. The relaxation oscillator consists of a 10.00-mF capacitor and a 10.00-k Ω variable resistor known as a rheostat. A knob connected to the variable resistor allows the resistance to be adjusted from 0.00 Ω to 10.00 k Ω . The output of the capacitor is used to control a voltage-controlled switch. The switch is normally open, but when the output voltage reaches 10.00 V, the switch closes, energizing an electric motor and discharging the capacitor. The motor causes the windshield wipers to sweep once across the windshield and the capacitor begins to charge again. To what resistance should the rheostat be adjusted for the period of the wiper blades be 10.00 seconds?



Strategy

The resistance considers the equation $V_{\text{out}}(t) = V(1 - e^{-t/\tau})$, where $\tau = RC$. The capacitance, output voltage, and voltage of the battery are given. We need to solve this equation for the resistance.

Solution

The output voltage will be 10.00 V and the voltage of the battery is 12.00 V. The capacitance is given as 10.00 mF. Solving for the resistance yields

Equation:

$$\begin{aligned}
 V_{\text{out}}(t) &= V(1 - e^{-t/\tau}), \\
 e^{-t/RC} &= 1 - \frac{V_{\text{out}}(t)}{V}, \\
 \ln(e^{-t/RC}) &= \ln\left(1 - \frac{V_{\text{out}}(t)}{V}\right), \\
 -\frac{t}{RC} &= \ln\left(1 - \frac{V_{\text{out}}(t)}{V}\right), \\
 R &= \frac{-t}{C \ln\left(1 - \frac{V_{\text{out}}(t)}{V}\right)} = \frac{-10.00 \text{ s}}{10 \times 10^{-3} \text{ F} \ln\left(1 - \frac{10 \text{ V}}{12 \text{ V}}\right)} = 558.11 \Omega.
 \end{aligned}$$

Significance

Increasing the resistance increases the time delay between operations of the windshield wipers. When the resistance is zero, the windshield wipers run continuously. At the maximum resistance, the period of the operation of the wipers is:

Equation:

$$t = -RC \ln \left(1 - \frac{V_{\text{out}}(t)}{V} \right) = - (10 \times 10^{-3} \text{ F}) (10 \times 10^3 \Omega) \ln \left(1 - \frac{10 \text{ V}}{12 \text{ V}} \right) = 179.18 \text{ s} = 2.98 \text{ min.}$$

The RC circuit has thousands of uses and is a very important circuit to study. Not only can it be used to time circuits, it can also be used to filter out unwanted frequencies in a circuit and used in power supplies, like the one for your computer, to help turn ac voltage to dc voltage.

Summary

- An RC circuit is one that has both a resistor and a capacitor.
- The time constant τ for an RC circuit is $\tau = RC$.
- When an initially uncharged ($q = 0$ at $t = 0$) capacitor in series with a resistor is charged by a dc voltage source, the capacitor asymptotically approaches the maximum charge.
- As the charge on the capacitor increases, the current exponentially decreases from the initial current: $I_0 = \mathcal{E}/R$.
- If a capacitor with an initial charge Q is discharged through a resistor starting at $t = 0$, then its charge decreases exponentially. The current flows in the opposite direction, compared to when it charges, and the magnitude of the charge decreases with time.

Conceptual Questions

Exercise:

Problem:

A battery, switch, capacitor, and lamp are connected in series. Describe what happens to the lamp when the switch is closed.

Exercise:

Problem:

When making an ECG measurement, it is important to measure voltage variations over small time intervals. The time is limited by the RC constant of the circuit—it is not possible to measure time variations shorter than RC . How would you manipulate R and C in the circuit to allow the necessary measurements?

Solution:

The time constant can be shortened by using a smaller resistor and/or a smaller capacitor. Care should be taken when reducing the resistance because the initial current will increase as the resistance decreases.

Problems

Exercise:

Problem:

The timing device in an automobile's intermittent wiper system is based on an RC time constant and utilizes a $0.500\text{-}\mu\text{F}$ capacitor and a variable resistor. Over what range must R be made to vary to achieve time constants from 2.00 to 15.0 s?

Solution:

4.00 to 30.0 $\text{M}\Omega$

Exercise:**Problem:**

A heart pacemaker fires 72 times a minute, each time a 25.0-nF capacitor is charged (by a battery in series with a resistor) to 0.632 of its full voltage. What is the value of the resistance?

Exercise:**Problem:**

The duration of a photographic flash is related to an RC time constant, which is $0.100\mu\text{s}$ for a certain camera. (a) If the resistance of the flash lamp is $0.0400\ \Omega$ during discharge, what is the size of the capacitor supplying its energy? (b) What is the time constant for charging the capacitor, if the charging resistance is $800\ \text{k}\Omega$?

Solution:

a. $2.50\ \mu\text{F}$; b. 2.00 s

Exercise:**Problem:**

A 2.00- and a $7.50\text{-}\mu\text{F}$ capacitor can be connected in series or parallel, as can a 25.0- and a $100\text{-k}\Omega$ resistor. Calculate the four RC time constants possible from connecting the resulting capacitance and resistance in series.

Exercise:**Problem:**

A $500\text{-}\Omega$ resistor, an uncharged $1.50\text{-}\mu\text{F}$ capacitor, and a 6.16-V emf are connected in series. (a) What is the initial current? (b) What is the RC time constant? (c) What is the current after one time constant? (d) What is the voltage on the capacitor after one time constant?

Solution:

a. $12.3\ \text{mA}$; b. $7.50 \times 10^{-4}\text{s}$; c. $4.53\ \text{mA}$; d. $3.89\ \text{V}$

Exercise:**Problem:**

A heart defibrillator being used on a patient has an RC time constant of 10.0 ms due to the resistance of the patient and the capacitance of the defibrillator. (a) If the defibrillator has a capacitance of $8.00\mu\text{F}$, what is the resistance of the path through the patient? (You may neglect the capacitance of the patient and the resistance of the defibrillator.) (b) If the initial voltage is $12.0\ \text{kV}$, how long does it take to decline to $6.00 \times 10^2\ \text{V}$?

Exercise:

Problem:

An ECG monitor must have an RC time constant less than $1.00 \times 10^2 \mu\text{s}$ to be able to measure variations in voltage over small time intervals. (a) If the resistance of the circuit (due mostly to that of the patient's chest) is $1.00 \text{ k}\Omega$, what is the maximum capacitance of the circuit? (b) Would it be difficult in practice to limit the capacitance to less than the value found in (a)?

Solution:

a. $1.00 \times 10^{-7} \text{ F}$; b. No, in practice it would not be difficult to limit the capacitance to less than 100 nF , since typical capacitors range from fractions of a picofarad (pF) to milifarad (mF).

Exercise:**Problem:**

Using the exact exponential treatment, determine how much time is required to charge an initially uncharged 100-pF capacitor through a $75.0\text{-M}\Omega$ resistor to 90.0% of its final voltage.

Exercise:**Problem:**

If you wish to take a picture of a bullet traveling at 500 m/s , then a very brief flash of light produced by an RC discharge through a flash tube can limit blurring. Assuming 1.00 mm of motion during one RC constant is acceptable, and given that the flash is driven by a $600\text{-}\mu\text{F}$ capacitor, what is the resistance in the flash tube?

Solution:

$$3.33 \times 10^{-3} \Omega$$

Glossary **RC circuit**

circuit that contains both a resistor and a capacitor

Household Wiring and Electrical Safety

By the end of the section, you will be able to:

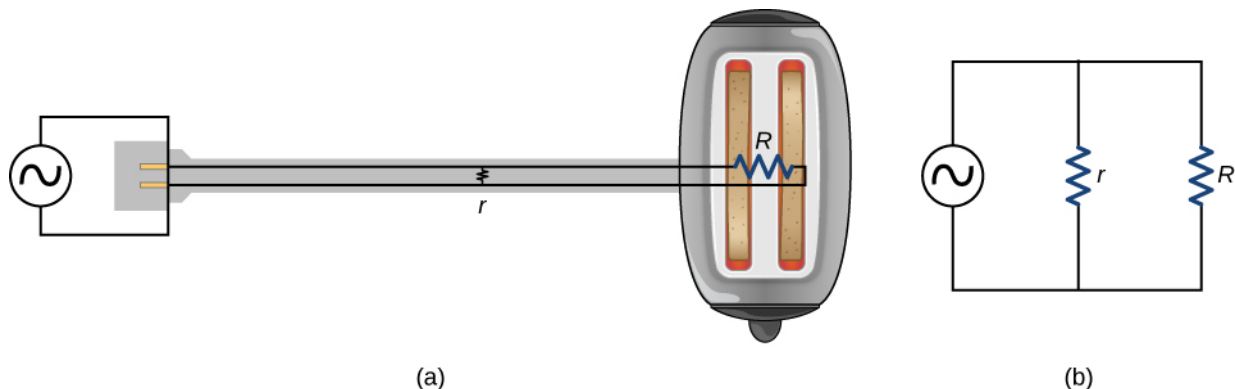
- List the basic concepts involved in house wiring
- Define the terms thermal hazard and shock hazard
- Describe the effects of electrical shock on human physiology and their relationship to the amount of current through the body
- Explain the function of fuses and circuit breakers

Electricity presents two known hazards: thermal and shock. A **thermal hazard** is one in which an excessive electric current causes undesired thermal effects, such as starting a fire in the wall of a house. A **shock hazard** occurs when an electric current passes through a person. Shocks range in severity from painful, but otherwise harmless, to heart-stopping lethality. In this section, we consider these hazards and the various factors affecting them in a quantitative manner. We also examine systems and devices for preventing electrical hazards.

Thermal Hazards

Electric power causes undesired heating effects whenever electric energy is converted into thermal energy at a rate faster than it can be safely dissipated. A classic example of this is the short circuit, a low-resistance path between terminals of a voltage source. An example of a short circuit is shown in [\[link\]](#). A toaster is plugged into a common household electrical outlet. Insulation on wires leading to an appliance has worn through, allowing the two wires to come into contact, or “short.” As a result, thermal energy can quickly raise the temperature of surrounding materials, melting the insulation and perhaps causing a fire.

The circuit diagram shows a symbol that consists of a sine wave enclosed in a circle. This symbol represents an alternating current (ac) voltage source. In an ac voltage source, the voltage oscillates between a positive and negative maximum amplitude. Up to now, we have been considering direct current (dc) voltage sources, but many of the same concepts are applicable to ac circuits.



A short circuit is an undesired low-resistance path across a voltage source. (a) Worn

insulation on the wires of a toaster allow them to come into contact with a low resistance r . Since $P = V^2/r$, thermal power is created so rapidly that the cord melts or burns. (b)
A schematic of the short circuit.

Another serious thermal hazard occurs when wires supplying power to an appliance are overloaded. Electrical wires and appliances are often rated for the maximum current they can safely handle. The term “overloaded” refers to a condition where the current exceeds the rated maximum current. As current flows through a wire, the power dissipated in the supply wires is $P = I^2 R_W$, where R_W is the resistance of the wires and I is the current flowing through the wires. If either I or R_W is too large, the wires overheat. Fuses and circuit breakers are used to limit excessive currents.

Shock Hazards

Electric shock is the physiological reaction or injury caused by an external electric current passing through the body. The effect of an electric shock can be negative or positive. When a current with a magnitude above 300 mA passes through the heart, death may occur. Most electrical shock fatalities occur because a current causes ventricular fibrillation, a massively irregular and often fatal, beating of the heart. On the other hand, a heart attack victim, whose heart is in fibrillation, can be saved by an electric shock from a defibrillator.

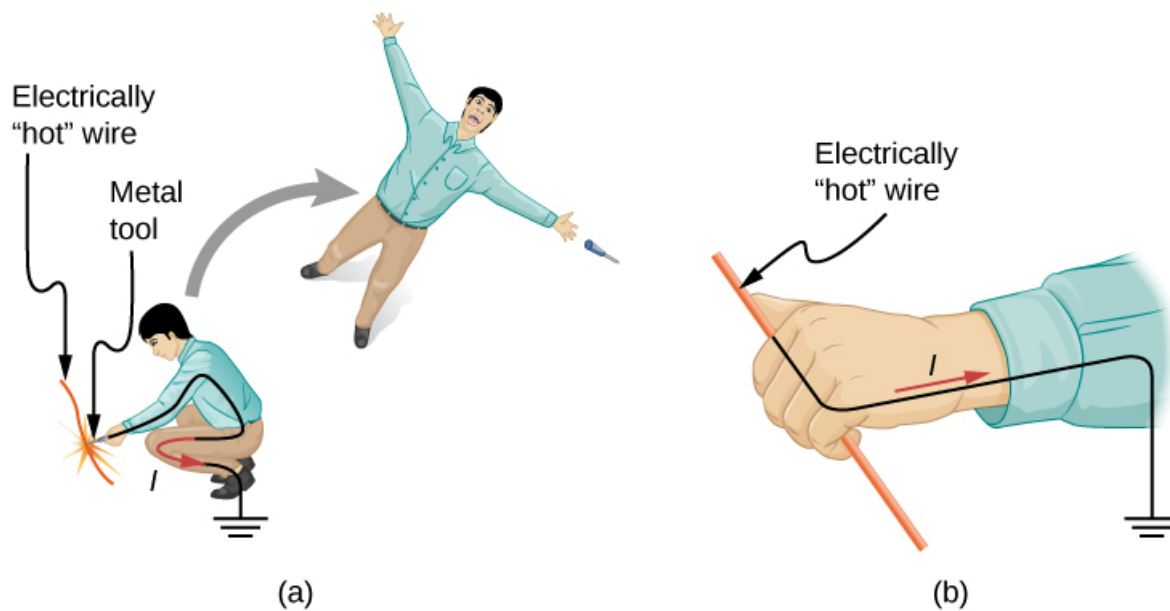
The effects of an undesirable electric shock can vary in severity: a slight sensation at the point of contact, pain, loss of voluntary muscle control, difficulty breathing, heart fibrillation, and possibly death. The loss of voluntary muscle control can cause the victim to not be able to let go of the source of the current.

The major factors upon which the severity of the effects of electrical shock depend are

1. The amount of current I
2. The path taken by the current
3. The duration of the shock
4. The frequency f of the current ($f = 0$ for dc)

Our bodies are relatively good electric conductors due to the body’s water content. A dangerous condition occurs when the body is in contact with a voltage source and “ground.” The term “ground” refers to a large sink or source of electrons, for example, the earth (thus, the name). When there is a direct path to ground, large currents will pass through the parts of the body with the lowest resistance and a direct path to ground. A safety precaution used by many professions is the wearing of insulated shoes. Insulated shoes prohibit a pathway to ground for electrons through the feet by providing a large resistance. Whenever working with high-power tools, or any electric circuit, ensure that you do not provide a pathway for current flow (especially across the heart). A common safety precaution is to work with one hand, reducing the possibility of providing a current path through the heart.

Very small currents pass harmlessly and unfelt through the body. This happens to you regularly without your knowledge. The threshold of sensation is only 1 mA and, although unpleasant, shocks are apparently harmless for currents less than 5 mA. A great number of safety rules take the 5-mA value for the maximum allowed shock. At 5–30 mA and above, the current can stimulate sustained muscular contractions, much as regular nerve impulses do ([link](#)). Very large currents (above 300 mA) cause the heart and diaphragm of the lung to contract for the duration of the shock. Both the heart and respiration stop. Both often return to normal following the shock.



An electric current can cause muscular contractions with varying effects. (a) The victim is “thrown” backward by involuntary muscle contractions that extend the legs and torso. (b) The victim can’t let go of the wire that is stimulating all the muscles in the hand. Those that close the fingers are stronger than those that open them.

Current is the major factor determining shock severity. A larger voltage is more hazardous, but since $I = V/R$, the severity of the shock depends on the combination of voltage and resistance. For example, a person with dry skin has a resistance of about 200 k Ω . If he comes into contact with 120-V ac, a current

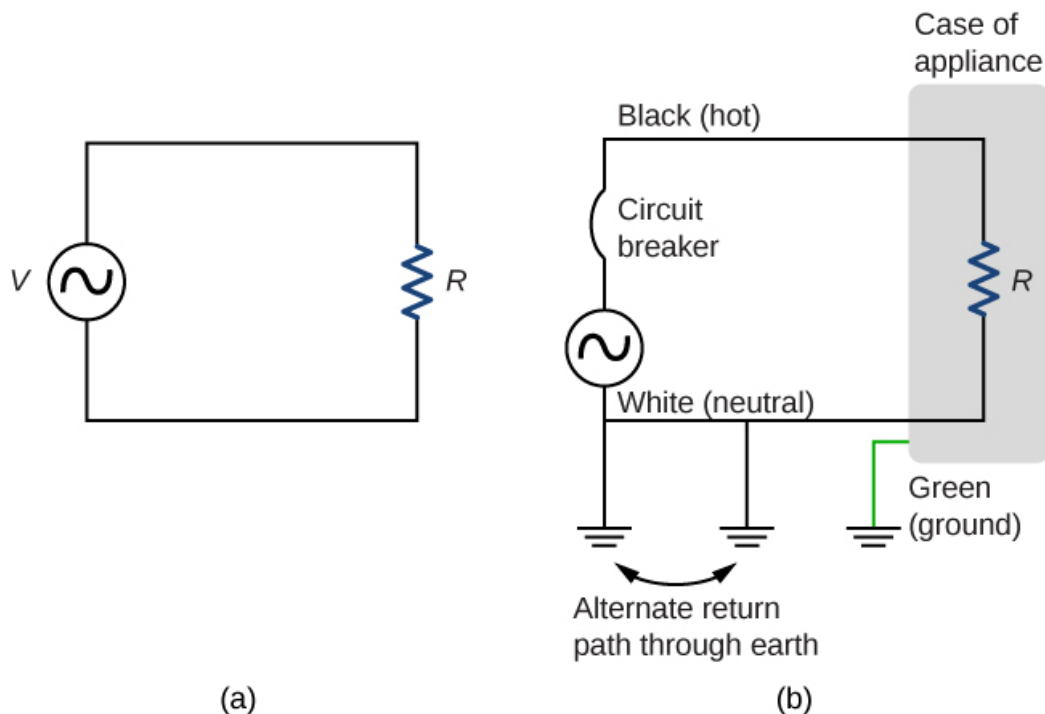
Equation:

$$I = (120 \text{ V}) / (200 \text{ k}\Omega) = 0.6 \text{ mA}$$

passes harmlessly through him. The same person soaking wet may have a resistance of 10.0 k Ω and the same 120 V will produce a current of 12 mA—above the “can’t let go” threshold and potentially dangerous.

Electrical Safety: Systems and Devices

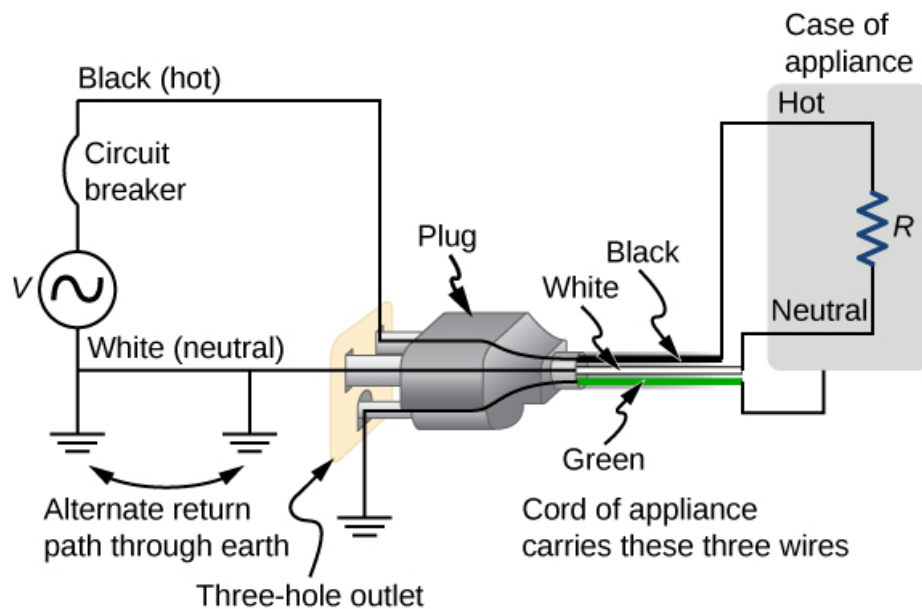
[\[link\]](#)(a) shows the schematic for a simple ac circuit with no safety features. This is not how power is distributed in practice. Modern household and industrial wiring requires the **three-wire system**, shown schematically in part (b), which has several safety features, with live, neutral, and ground wires. First is the familiar circuit breaker (or fuse) to prevent thermal overload. Second is a protective case around the appliance, such as a toaster or refrigerator. The case's safety feature is that it prevents a person from touching exposed wires and coming into electrical contact with the circuit, helping prevent shocks.



(a) Schematic of a simple ac circuit with a voltage source and a single appliance represented by the resistance R . There are no safety features in this circuit. (b) The three-wire system connects the neutral wire to ground at the voltage source and user location, forcing it to be at zero volts and supplying an alternative return path for the current through ground. Also grounded to zero volts is the case of the appliance. A circuit breaker or fuse protects against thermal overload and is in series on the active (live/hot) wire.

There are three connections to ground shown in [\[link\]](#)(b). Recall that a ground connection is a low-resistance path directly to ground. The two ground connections on the neutral wire force it to be at zero volts relative to ground, giving the wire its name. This wire is therefore safe to

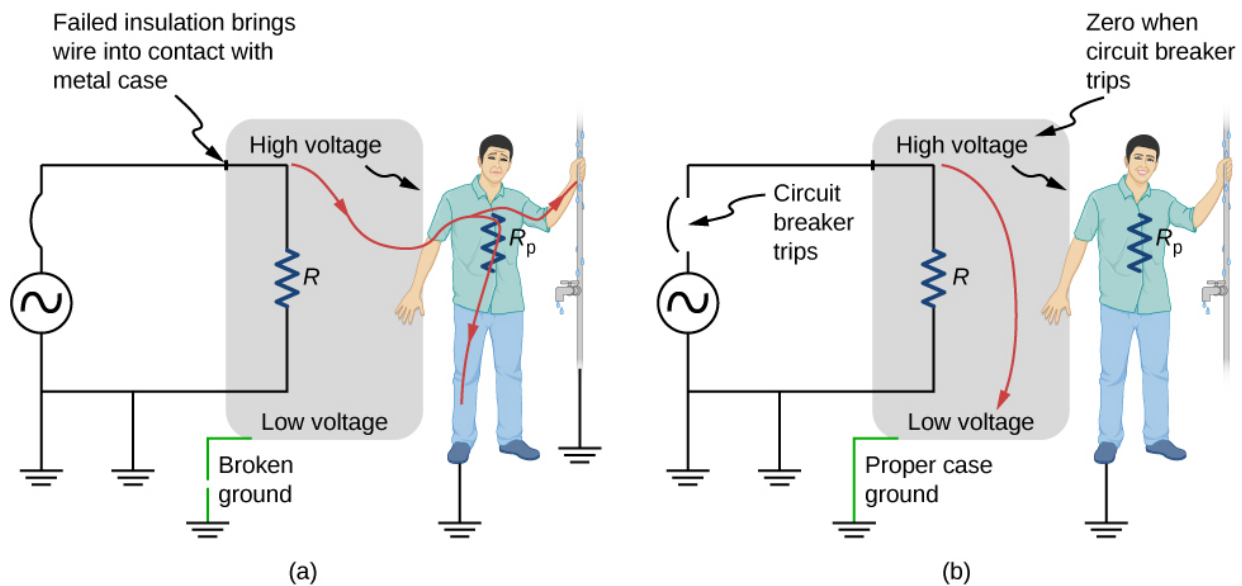
touch even if its insulation, usually white, is missing. The neutral wire is the return path for the current to follow to complete the circuit. Furthermore, the two ground connections supply an alternative path through ground (a good conductor) to complete the circuit. The ground connection closest to the power source could be at the generating plant, whereas the other is at the user's location. The third ground is to the case of the appliance, through the green ground wire, forcing the case, too, to be at zero volts. The live or hot wire (hereafter referred to as "live/hot") supplies voltage and current to operate the appliance. [\[link\]](#) shows a more pictorial version of how the three-wire system is connected through a three-prong plug to an appliance.



The standard three-prong plug can only be inserted in one way, to ensure proper function of the three-wire system.

Insulating plastic is color-coded to identify live/hot, neutral, and ground wires, but these codes vary around the world. It is essential to determine the color code in your region. Striped coatings are sometimes used for the benefit of those who are colorblind.

Grounding the case solves more than one problem. The simplest problem is worn insulation on the live/hot wire that allows it to contact the case, as shown in [\[link\]](#). Lacking a ground connection, a severe shock is possible. This is particularly dangerous in the kitchen, where a good connection to ground is available through water on the floor or a water faucet. With the ground connection intact, the circuit breaker will trip, forcing repair of the appliance.



Worn insulation allows the live/hot wire to come into direct contact with the metal case of this appliance. (a) The ground connection being broken, the person is severely shocked. The appliance may operate normally in this situation. (b) With a proper ground, the circuit breaker trips, forcing repair of the appliance.

A ground fault circuit interrupter (GFCI) is a safety device found in updated kitchen and bathroom wiring that works based on electromagnetic induction. GFCIs compare the currents in the live/hot and neutral wires. When live/hot and neutral currents are not equal, it is almost always because current in the neutral is less than in the live/hot wire. Then some of the current, called a leakage current, is returning to the voltage source by a path other than through the neutral wire. It is assumed that this path presents a hazard. GFCIs are usually set to interrupt the circuit if the leakage current is greater than 5 mA, the accepted maximum harmless shock. Even if the leakage current goes safely to ground through an intact ground wire, the GFCI will trip, forcing repair of the leakage.

Summary

- The two types of electric hazards are thermal (excessive power) and shock (current through a person). Electrical safety systems and devices are employed to prevent thermal and shock hazards.
- Shock severity is determined by current, path, duration, and ac frequency.
- Circuit breakers and fuses interrupt excessive currents to prevent thermal hazards.
- The three-wire system guards against thermal and shock hazards, utilizing live/hot, neutral, and ground wires, and grounding the neutral wire and case of the appliance.
- A ground fault circuit interrupter (GFCI) prevents shock by detecting the loss of current to unintentional paths.

Key Equations

Terminal voltage of a single voltage source	$V_{\text{terminal}} = \varepsilon - I r_{\text{eq}}$
Equivalent resistance of a series circuit	$R_{\text{eq}} = R_1 + R_2 + R_3 + \cdots + R_{N-1} + R_N = \sum_{i=1}^N R_i$
Equivalent resistance of a parallel circuit	$R_{\text{eq}} = \left(\frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_N} \right)^{-1} = \left(\sum_{i=1}^N \frac{1}{R_i} \right)^{-1}$
Junction rule	$\sum I_{\text{in}} = \sum I_{\text{out}}$
Loop rule	$\sum V = 0$
Terminal voltage of N voltage sources in series	$V_{\text{terminal}} = \sum_{i=1}^N \varepsilon_i - I \sum_{i=1}^N r_i = \sum_{i=1}^N \varepsilon_i - I r_{\text{eq}}$
Terminal voltage of N voltage sources in parallel	$V_{\text{terminal}} = \varepsilon - I \sum_{i=1}^N \left(\frac{1}{r_i} \right)^{-1} = \varepsilon - I r_{\text{eq}}$
Charge on a charging capacitor	$q(t) = C\varepsilon \left(1 - e^{-\frac{t}{RC}} \right) = Q \left(1 - e^{-\frac{t}{\tau}} \right)$
Time constant	$\tau = RC$
Current during charging of a capacitor	$I = \frac{\varepsilon}{R} e^{-\frac{t}{RC}} = I_o e^{-\frac{t}{RC}}$
Charge on a discharging capacitor	$q(t) = Q e^{-\frac{t}{\tau}}$
Current during discharging of a capacitor	$I(t) = -\frac{Q}{RC} e^{-\frac{t}{\tau}}$

Conceptual Questions

Exercise:

Problem: Why isn't a short circuit necessarily a shock hazard?

Exercise:

Problem:

We are often advised to not flick electric switches with wet hands, dry your hand first. We are also advised to never throw water on an electric fire. Why?

Solution:

Not only might water drip into the switch and cause a shock, but also the resistance of your body is lower when you are wet.

Problems

Exercise:

Problem:

(a) How much power is dissipated in a short circuit of 240-V ac through a resistance of $0.250\ \Omega$? (b) What current flows?

Exercise:

Problem:

What voltage is involved in a 1.44-kW short circuit through a $0.100\text{-}\Omega$ resistance?

Solution:

12.0 V

Exercise:

Problem:

Find the current through a person and identify the likely effect on her if she touches a 120-V ac source: (a) if she is standing on a rubber mat and offers a total resistance of $300\ \text{k}\Omega$; (b) if she is standing barefoot on wet grass and has a resistance of only $4000\ \text{k}\Omega$.

Exercise:

Problem:

While taking a bath, a person touches the metal case of a radio. The path through the person to the drainpipe and ground has a resistance of $4000\ \Omega$. What is the smallest voltage on the case of the radio that could cause ventricular fibrillation?

Solution:

400 V

Exercise:**Problem:**

A man foolishly tries to fish a burning piece of bread from a toaster with a metal butter knife and comes into contact with 120-V ac. He does not even feel it since, luckily, he is wearing rubber-soled shoes. What is the minimum resistance of the path the current follows through the person?

Exercise:**Problem:**

(a) During surgery, a current as small as $20.0 \mu\text{A}$ applied directly to the heart may cause ventricular fibrillation. If the resistance of the exposed heart is 300Ω , what is the smallest voltage that poses this danger? (b) Does your answer imply that special electrical safety precautions are needed?

Solution:

a. 6.00 mV; b. It would not be necessary to take extra precautions regarding the power coming from the wall. However, it is possible to generate voltages of approximately this value from static charge built up on gloves, for instance, so some precautions are necessary.

Exercise:**Problem:**

(a) What is the resistance of a 220-V ac short circuit that generates a peak power of 96.8 kW? (b) What would the average power be if the voltage were 120 V ac?

Exercise:**Problem:**

A heart defibrillator passes 10.0 A through a patient's torso for 5.00 ms in an attempt to restore normal beating. (a) How much charge passed? (b) What voltage was applied if 500 J of energy was dissipated? (c) What was the path's resistance? (d) Find the temperature increase caused in the 8.00 kg of affected tissue.

Solution:

a. $5.00 \times 10^{-2} \text{ C}$; b. 10.0 kV; c. $1.00 \text{ k}\Omega$; d. $1.79 \times 10^{-2} \text{ }^\circ\text{C}$

Exercise:

Problem:

A short circuit in a 120-V appliance cord has a $0.500\text{-}\Omega$ resistance. Calculate the temperature rise of the 2.00 g of surrounding materials, assuming their specific heat capacity is $0.200\text{ cal/g} \cdot ^\circ\text{C}$ and that it takes 0.0500 s for a circuit breaker to interrupt the current. Is this likely to be damaging?

Additional Problems**Exercise:****Problem:**

A circuit contains a D cell battery, a switch, a $20\text{-}\Omega$ resistor, and four 20-mF capacitors connected in series. (a) What is the equivalent capacitance of the circuit? (b) What is the RC time constant? (c) How long before the current decreases to 50 % of the initial value once the switch is closed?

Solution:

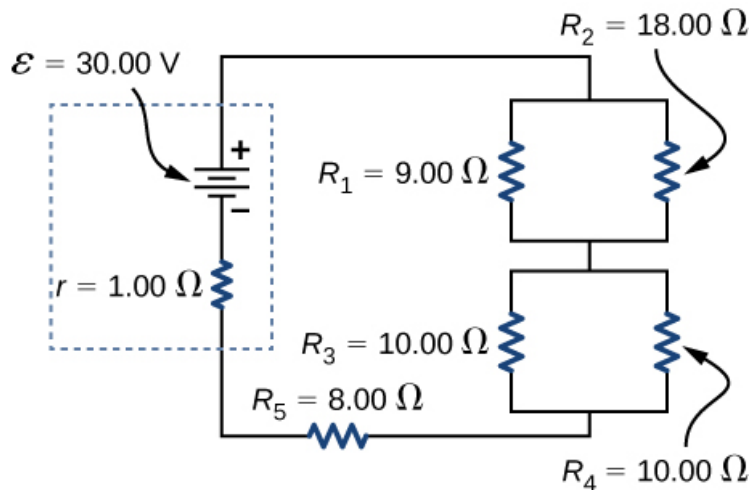
a. $C_{\text{eq}} = 5.00\text{ mF}$; b. $\tau = 0.1\text{ s}$; c. 0.069 s

Exercise:**Problem:**

A circuit contains a D-cell battery, a switch, a $20\text{-}\Omega$ resistor, and three 20-mF capacitors. The capacitors are connected in parallel, and the parallel connection of capacitors are connected in series with the switch, the resistor and the battery. (a) What is the equivalent capacitance of the circuit? (b) What is the RC time constant? (c) How long before the current decreases to 50 % of the initial value once the switch is closed?

Exercise:**Problem:**

Consider the circuit below. The battery has an emf of $\varepsilon = 30.00\text{ V}$ and an internal resistance of $r = 1.00\text{ }\Omega$. (a) Find the equivalent resistance of the circuit and the current out of the battery. (b) Find the current through each resistor. (c) Find the potential drop across each resistor. (d) Find the power dissipated by each resistor. (e) Find the total power supplied by the batteries.



Solution:

- $R_{\text{eq}} = 20.00 \, \Omega$;
- $I_r = 1.50 \, \text{A}$, $I_1 = 1.00 \, \text{A}$, $I_2 = 0.50 \, \text{A}$, $I_3 = 0.75 \, \text{A}$, $I_4 = 0.75 \, \text{A}$, $I_5 = 1.50 \, \text{A}$;
- $V_r = 1.50 \, \text{V}$, $V_1 = 9.00 \, \text{V}$, $V_2 = 9.00 \, \text{V}$, $V_3 = 7.50 \, \text{V}$, $V_4 = 7.50 \, \text{V}$, $V_5 = 12.00 \, \text{V}$;
- $P_r = 2.25 \, \text{W}$, $P_1 = 9.00 \, \text{W}$, $P_2 = 4.50 \, \text{W}$, $P_3 = 5.625 \, \text{W}$, $P_4 = 5.625 \, \text{W}$, $P_5 = 18.00 \, \text{W}$;
- $P = 45.00 \, \text{W}$

Exercise:

Problem:

A homemade capacitor is constructed of 2 sheets of aluminum foil with an area of 2.00 square meters, separated by paper, 0.05 mm thick, of the same area and a dielectric constant of 3.7. The homemade capacitor is connected in series with a 100.00- Ω resistor, a switch, and a 6.00-V voltage source. (a) What is the RC time constant of the circuit? (b) What is the initial current through the circuit, when the switch is closed? (c) How long does it take the current to reach one third of its initial value?

Exercise:

Problem:

A student makes a homemade resistor from a graphite pencil 5.00 cm long, where the graphite is 0.05 mm in diameter. The resistivity of the graphite is $\rho = 1.38 \times 10^{-5} \, \Omega/\text{m}$. The homemade resistor is placed in series with a switch, a 10.00-mF uncharged capacitor and a 0.50-V power source. (a) What is the RC time constant of the circuit? (b) What is the potential drop across the pencil 1.00 s after the switch is closed?

Solution:

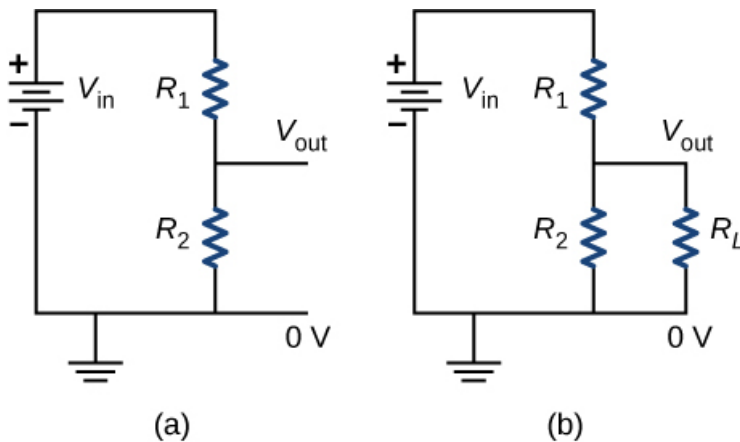
$$\text{a. } \tau = \left(1.38 \times 10^{-5} \Omega \text{m} \left(\frac{5.00 \times 10^{-2} \text{m}}{3.14 \left(\frac{0.05 \times 10^{-3}}{2} \right)^2} \right) \right) 10 \times 10^{-3} \text{F} = 3.52 \text{ s; b.}$$

$$V = 0.014 \text{ A} \left(e^{-\frac{1.00 \text{s}}{3.52 \text{s}}} \right) 351.59 \Omega = 0.376 \text{ V}$$

Exercise:

Problem:

The rather simple circuit shown below is known as a voltage divider. The symbol consisting of three horizontal lines represents “ground” and can be defined as the point where the potential is zero. The voltage divider is widely used in circuits and a single voltage source can be used to provide reduced voltage to a load resistor as shown in the second part of the figure. (a) What is the output voltage V_{out} of circuit (a) in terms of R_1 , R_2 , and V_{in} ? (b) What is the output voltage V_{out} of circuit (b) in terms of R_1 , R_2 , R_L , and V_{in} ?



Exercise:

Problem:

Three $300\text{-}\Omega$ resistors are connect in series with an AAA battery with a rating of 3 AmpHours. (a) How long can the battery supply the resistors with power? (b) If the resistors are connected in parallel, how long can the battery last?

Solution:

$$\text{a. } t = \frac{3 \text{ A}\cdot\text{h}}{\frac{1.5 \text{ V}}{900 \Omega}} = 1800 \text{ h; b. } t = \frac{3 \text{ A}\cdot\text{h}}{\frac{1.5 \text{ V}}{100 \Omega}} = 200 \text{ h}$$

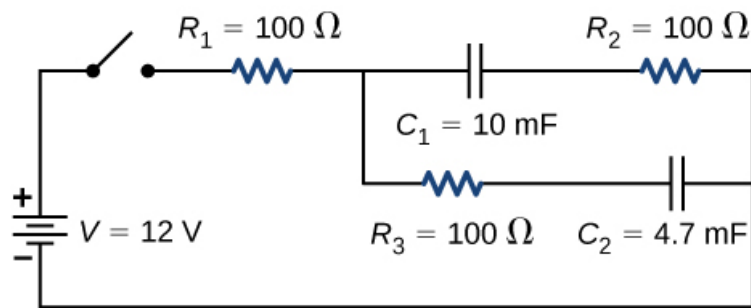
Exercise:

Problem:

Consider a circuit that consists of a real battery with an emf ε and an internal resistance of r connected to a variable resistor R . (a) In order for the terminal voltage of the battery to be equal to the emf of the battery, what should the resistance of the variable resistor be adjusted to? (b) In order to get the maximum current from the battery, what should the resistance of the variable resistor be adjusted to? (c) In order for the maximum power output of the battery to be reached, what should the resistance of the variable resistor be set to?

Exercise:**Problem:**

Consider the circuit shown below. What is the energy stored in each capacitor after the switch has been closed for a very long time?

**Solution:**

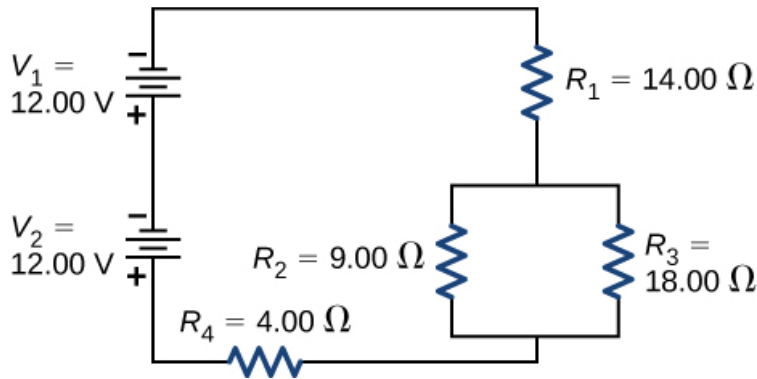
$$U_1 = C_1 V_1^2 = 0.72 \text{ J}, \quad U_2 = C_2 V_2^2 = 0.338 \text{ J}$$

Exercise:**Problem:**

Consider a circuit consisting of a battery with an emf ε and an internal resistance of r connected in series with a resistor R and a capacitor C . Show that the total energy supplied by the battery while charging the battery is equal to $\varepsilon^2 C$.

Exercise:**Problem:**

Consider the circuit shown below. The terminal voltages of the batteries are shown. (a) Find the equivalent resistance of the circuit and the current out of the battery. (b) Find the current through each resistor. (c) Find the potential drop across each resistor. (d) Find the power dissipated by each resistor. (e) Find the total power supplied by the batteries.



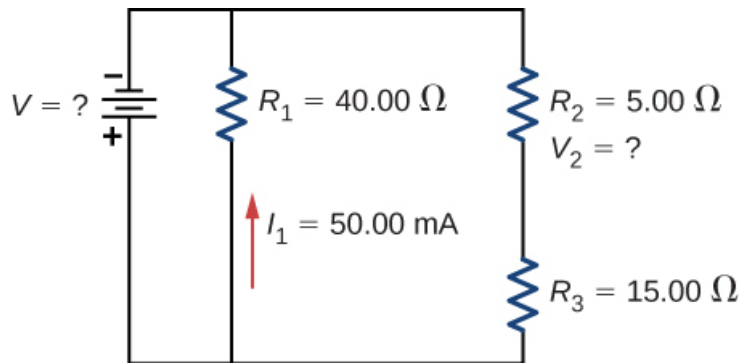
Solution:

- a. $R_{\text{eq}} = 24.00 \, \Omega$; b. $I_1 = 1.00 \, \text{A}$, $I_2 = 0.67 \, \text{A}$, $I_3 = 0.33 \, \text{A}$, $I_4 = 1.00 \, \text{A}$;
 c. $V_1 = 14.00 \, \text{V}$, $V_2 = 6.00 \, \text{V}$, $V_3 = 6.00 \, \text{V}$, $V_4 = 4.00 \, \text{V}$;
 d. $P_1 = 14.00 \, \text{W}$, $P_2 = 4.04 \, \text{W}$, $P_3 = 1.96 \, \text{W}$, $P_4 = 4.00 \, \text{W}$; e. $P = 24.00 \, \text{W}$

Exercise:

Problem:

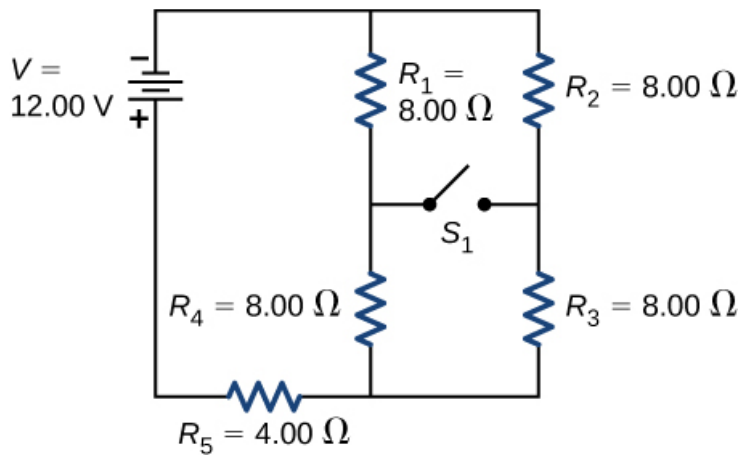
Consider the circuit shown below. (a) What is the terminal voltage of the battery? (b) What is the potential drop across resistor R_2 ?



Exercise:

Problem:

Consider the circuit shown below. (a) Determine the equivalent resistance and the current from the battery with switch S_1 open. (b) Determine the equivalent resistance and the current from the battery with switch S_1 closed.



Solution:

a. $R_{\text{eq}} = 12.00 \, \Omega$, $I = 1.00 \, \text{A}$; b. $R_{\text{eq}} = 12.00 \, \Omega$, $I = 1.00 \, \text{A}$

Exercise:

Problem:

Two resistors, one having a resistance of $145 \, \Omega$, are connected in parallel to produce a total resistance of $150 \, \Omega$. (a) What is the value of the second resistance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem:

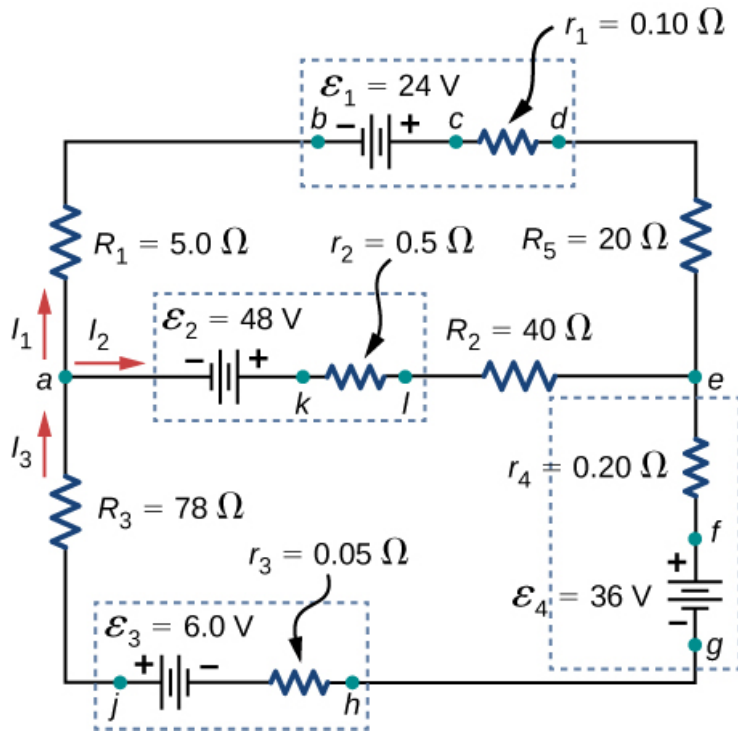
Two resistors, one having a resistance of $900 \, \text{k}\Omega$, are connected in series to produce a total resistance of $0.500 \, \text{M}\Omega$. (a) What is the value of the second resistance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. $-400 \, \text{k}\Omega$; b. You cannot have negative resistance. c. The assumption that $R_{\text{eq}} < R_1$ is unreasonable. Series resistance is always greater than any of the individual resistances.

Exercise:

Problem: Apply the junction rule at point *a* shown below.



Exercise:

Problem: Apply the loop rule to Loop $akledcba$ in the preceding problem.

Solution:

$$E_2 - I_2 r_2 - I_2 R_2 + I_1 R_5 + I_1 r_1 - E_1 + I_1 R_1 = 0$$

Exercise:

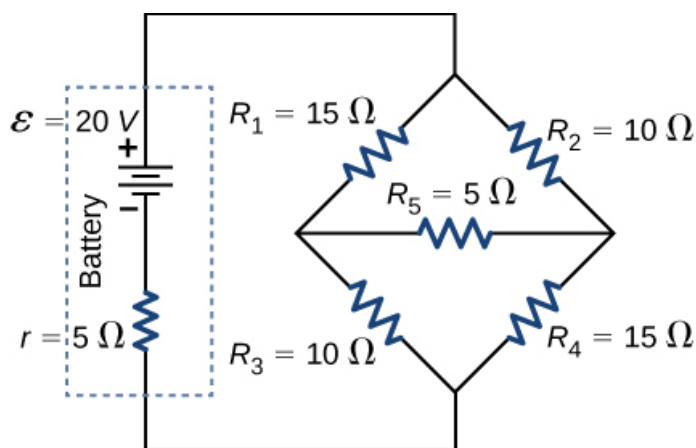
Problem:

Find the currents flowing in the circuit in the preceding problem. Explicitly show how you follow the steps in the [Problem-Solving Strategy: Series and Parallel Resistors](#).

Exercise:

Problem:

Consider the circuit shown below. (a) Find the current through each resistor. (b) Check the calculations by analyzing the power in the circuit.



Solution:

- a. $I = 1.17 \text{ A}$, $I_1 = 0.50 \text{ A}$, $I_2 = 0.67 \text{ A}$, $I_3 = 0.67 \text{ A}$, $I_4 = 0.50 \text{ A}$, $I_5 = 0.17 \text{ A}$;
 b. $P_{\text{output}} = 23.4 \text{ W}$, $P_{\text{input}} = 23.4 \text{ W}$

Exercise:

Problem:

A flashing lamp in a Christmas earring is based on an RC discharge of a capacitor through its resistance. The effective duration of the flash is 0.250 s , during which it produces an average 0.500 W from an average 3.00 V . (a) What energy does it dissipate? (b) How much charge moves through the lamp? (c) Find the capacitance. (d) What is the resistance of the lamp? (Since average values are given for some quantities, the shape of the pulse profile is not needed.)

Exercise:

Problem:

A $160\text{-}\mu\text{F}$ capacitor charged to 450 V is discharged through a $31.2\text{-k}\Omega$ resistor. (a) Find the time constant. (b) Calculate the temperature increase of the resistor, given that its mass is 2.50 g and its specific heat is $1.67 \text{ kJ/kg} \cdot ^\circ\text{C}$, noting that most of the thermal energy is retained in the short time of the discharge. (c) Calculate the new resistance, assuming it is pure carbon. (d) Does this change in resistance seem significant?

Solution:

- a. 4.99 s ; b. $3.87 ^\circ\text{C}$; c. $3.11 \times 10^4 \Omega$; d. No, this change does not seem significant. It probably would not be noticed.

Challenge Problems

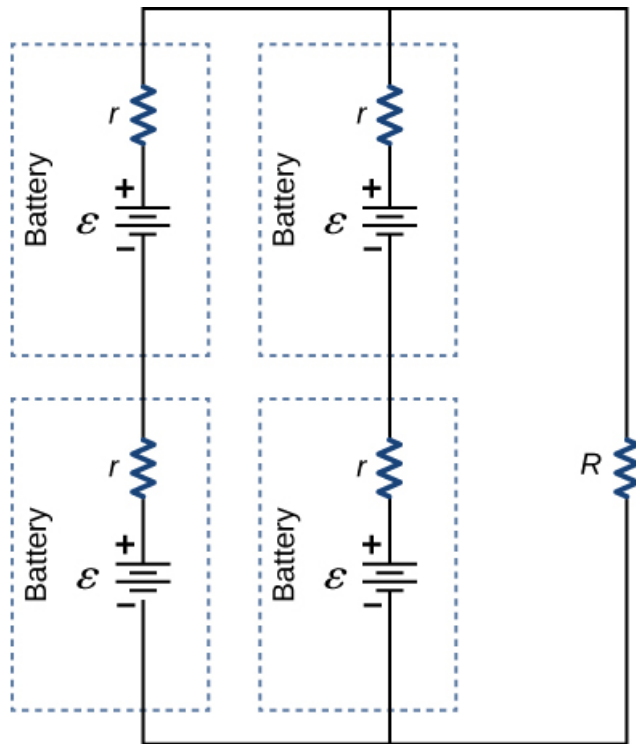
Exercise:

Problem:

Some camera flashes use flash tubes that require a high voltage. They obtain a high voltage by charging capacitors in parallel and then internally changing the connections of the capacitors to place them in series. Consider a circuit that uses four AAA batteries connected in series to charge six 10-mF capacitors through an equivalent resistance of $100\ \Omega$. The connections are then switched internally to place the capacitors in series. The capacitors discharge through a lamp with a resistance of $100\ \Omega$. (a) What is the RC time constant and the initial current out of the batteries while they are connected in parallel? (b) How long does it take for the capacitors to charge to 90 % of the terminal voltages of the batteries? (c) What is the RC time constant and the initial current of the capacitors connected in series assuming it discharges at 90 % of full charge? (d) How long does it take the current to decrease to 10 % of the initial value?

Exercise:**Problem:**

Consider the circuit shown below. Each battery has an emf of $1.50\ \text{V}$ and an internal resistance of $1.00\ \Omega$. (a) What is the current through the external resistor, which has a resistance of $10.00\ \text{ohms}$? (b) What is the terminal voltage of each battery?

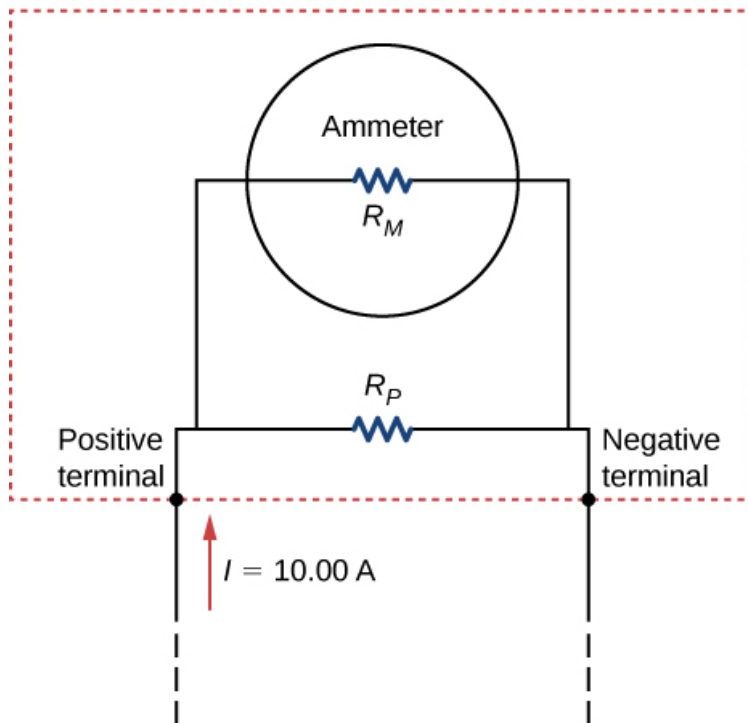
**Solution:**

a. $0.273\ \text{A}$; b. $V_T = 1.36\ \text{V}$

Exercise:

Problem:

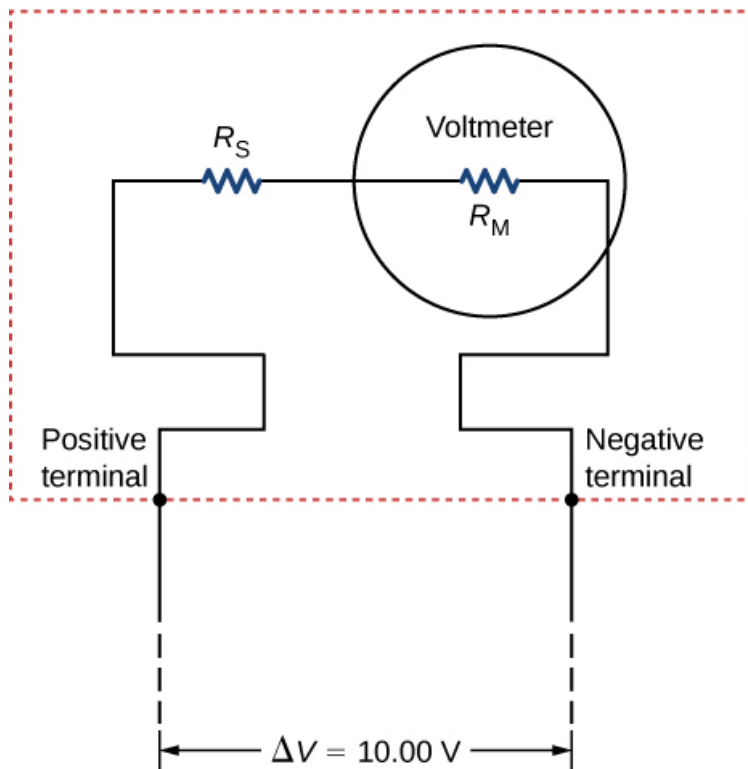
Analog meters use a galvanometer, which essentially consists of a coil of wire with a small resistance and a pointer with a scale attached. When current runs through the coil, the pointer turns; the amount the pointer turns is proportional to the amount of current running through the coil. Galvanometers can be used to make an ammeter if a resistor is placed in parallel with the galvanometer. Consider a galvanometer that has a resistance of $25.00\ \Omega$ and gives a full scale reading when a $50\text{-}\mu\text{A}$ current runs through it. The galvanometer is to be used to make an ammeter that has a full scale reading of $10.00\ \text{A}$, as shown below. Recall that an ammeter is connected in series with the circuit of interest, so all $10\ \text{A}$ must run through the meter. (a) What is the current through the parallel resistor in the meter? (b) What is the voltage across the parallel resistor? (c) What is the resistance of the series resistor?



Exercise:

Problem:

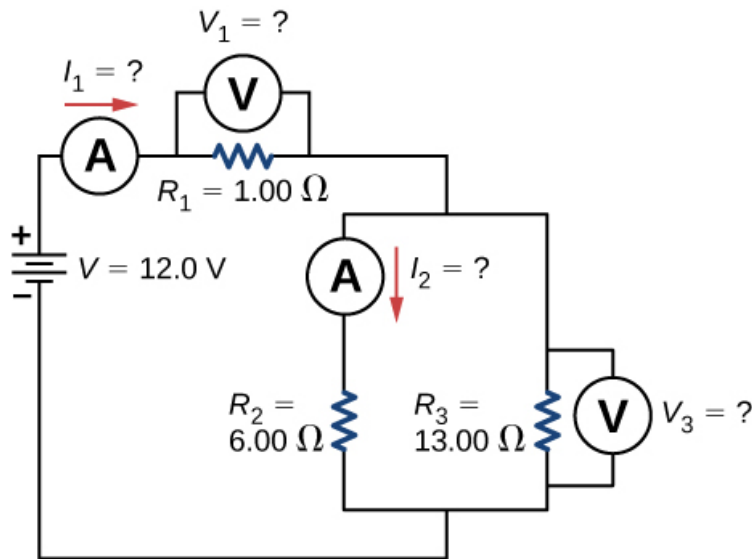
Analog meters use a galvanometer, which essentially consists of a coil of wire with a small resistance and a pointer with a scale attached. When current runs through the coil, the point turns; the amount the pointer turns is proportional to the amount of current running through the coil. Galvanometers can be used to make a voltmeter if a resistor is placed in series with the galvanometer. Consider a galvanometer that has a resistance of $25.00\ \Omega$ and gives a full scale reading when a $50\text{-}\mu\text{A}$ current runs through it. The galvanometer is to be used to make an voltmeter that has a full scale reading of $10.00\ \text{V}$, as shown below. Recall that a voltmeter is connected in parallel with the component of interest, so the meter must have a high resistance or it will change the current running through the component. (a) What is the potential drop across the series resistor in the meter? (b) What is the resistance of the parallel resistor?

**Solution:**

a. $V_s = V - I_M R_M = 9.99875\ \text{V}$; b. $R_S = \frac{V_P}{I_M} = 199.975\ \text{k}\Omega$

Exercise:

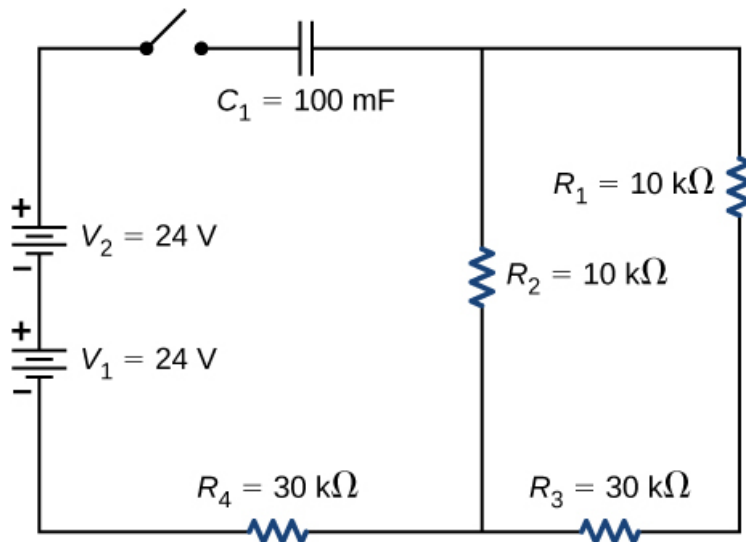
Problem: Consider the circuit shown below. Find I_1 , V_1 , I_2 , and V_3 .



Exercise:

Problem:

Consider the circuit below. (a) What is the RC time constant of the circuit? (b) What is the initial current in the circuit once the switch is closed? (c) How much time passes between the instant the switch is closed and the time the current has reached half of the initial current?



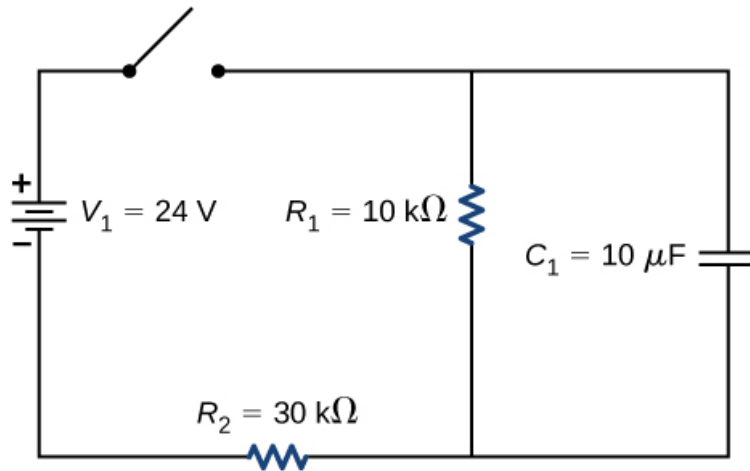
Solution:

a. $\tau = 3800 \, \text{s}$; b. $1.26 \, \text{mA}$; c. $t = 2633.96 \, \text{s}$

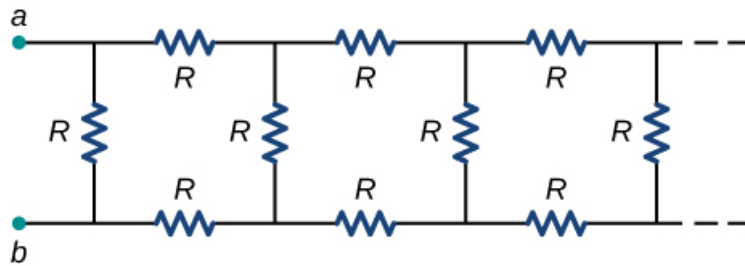
Exercise:

Problem:

Consider the circuit below. (a) What is the initial current through resistor R_2 when the switch is closed? (b) What is the current through resistor R_2 when the capacitor is fully charged, long after the switch is closed? (c) What happens if the switch is opened after it has been closed for some time? (d) If the switch has been closed for a time period long enough for the capacitor to become fully charged, and then the switch is opened, how long before the current through resistor R_1 reaches half of its initial value?

**Exercise:****Problem:**

Consider the infinitely long chain of resistors shown below. What is the resistance between terminals a and b ?

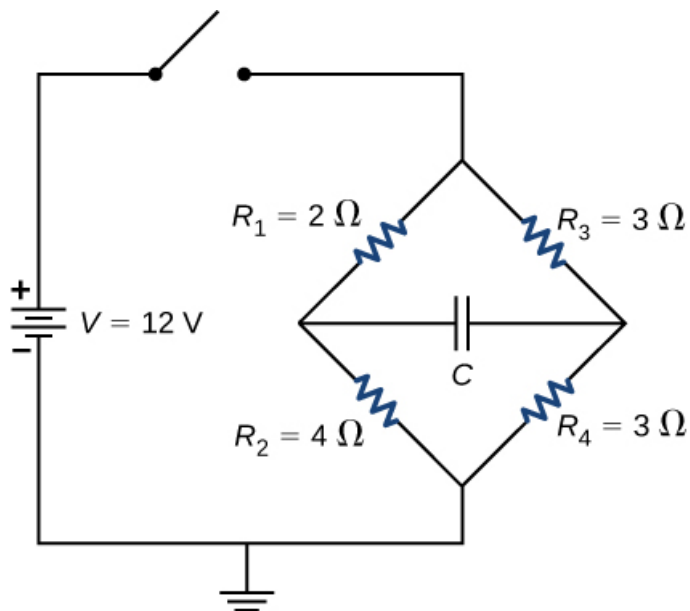
**Solution:**

$$R_{\text{eq}} = (\sqrt{3} - 1)R$$

Exercise:

Problem:

Consider the circuit below. The capacitor has a capacitance of 10 mF. The switch is closed and after a long time the capacitor is fully charged. (a) What is the current through each resistor a long time after the switch is closed? (b) What is the voltage across each resistor a long time after the switch is closed? (c) What is the voltage across the capacitor a long time after the switch is closed? (d) What is the charge on the capacitor a long time after the switch is closed? (e) The switch is then opened. The capacitor discharges through the resistors. How long from the time before the current drops to one fifth of the initial value?

**Exercise:****Problem:**

A 120-V immersion heater consists of a coil of wire that is placed in a cup to boil the water. The heater can boil one cup of $20.00\ ^\circ\text{C}$ water in 180.00 seconds. You buy one to use in your dorm room, but you are worried that you will overload the circuit and trip the 15.00-A, 120-V circuit breaker, which supplies your dorm room. In your dorm room, you have four 100.00-W incandescent lamps and a 1500.00-W space heater. (a) What is the power rating of the immersion heater? (b) Will it trip the breaker when everything is turned on? (c) If you replace the incandescent bulbs with 18.00-W LED, will the breaker trip when everything is turned on?

Solution:

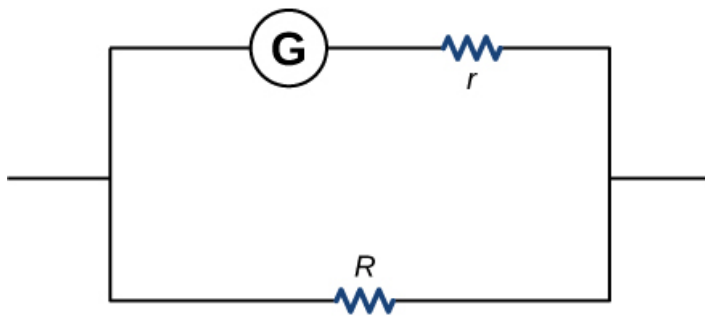
$$\begin{aligned} \text{a. } P_{\text{imheater}} &= \frac{1 \text{ cup} \left(\frac{0.000237 \text{ m}^3}{\text{cup}} \right) \left(\frac{1000 \text{ kg}}{\text{m}^3} \right) \left(4186 \frac{\text{J}}{\text{kg} \cdot ^\circ\text{C}} \right) (100\ ^\circ\text{C} - 20\ ^\circ\text{C})}{180.00 \text{ s}} \approx 441 \text{ W}; \\ \text{b. } I &= \frac{441 \text{ W}}{120 \text{ V}} + 4 \frac{100 \text{ W}}{120 \text{ V}} + \frac{1500 \text{ W}}{120 \text{ V}} = 19.51 \text{ A}; \text{ Yes, the breaker will trip.} \\ \text{c. } I &= \frac{441 \text{ W}}{120 \text{ V}} + 4 \frac{18 \text{ W}}{120 \text{ V}} + \frac{1500 \text{ W}}{120 \text{ V}} = 16.78 \text{ A}; \text{ Yes, the breaker will trip.} \end{aligned}$$

Exercise:**Problem:**

Find the resistance that must be placed in series with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as a voltmeter with a 3000-V full-scale reading. Include a circuit diagram with your solution.

Exercise:**Problem:**

Find the resistance that must be placed in parallel with a $60.0\text{-}\Omega$ galvanometer having a 1.00-mA sensitivity (the same as the one discussed in the text) to allow it to be used as an ammeter with a 25.0-A full-scale reading. Include a circuit diagram with your solution.

Solution:

$$2.40 \times 10^{-3} \Omega$$

Glossary

shock hazard

hazard in which an electric current passes through a person

thermal hazard

hazard in which an excessive electric current causes undesired thermal effects

three-wire system

wiring system used at present for safety reasons, with live, neutral, and ground wires

Introduction

class="introduction"

An industrial
electromagnet is
capable of lifting
thousands of
pounds of metallic
waste. (credit:
modification of
work by
“BedfordAl”/Flickr
)



For the past few chapters, we have been studying electrostatic forces and fields, which are caused by electric charges at rest. These electric fields can move other free charges, such as producing a current in a circuit; however, the electrostatic forces and fields themselves come from other static charges. In this chapter, we see that when an electric charge moves, it generates other forces and fields. These additional forces and fields are what we commonly call magnetism.

Before we examine the origins of magnetism, we first describe what it is and how magnetic fields behave. Once we are more familiar with magnetic effects, we can explain how they arise from the behavior of atoms and molecules, and how magnetism is related to electricity. The connection between electricity and magnetism is fascinating from a theoretical point of view, but it is also immensely practical, as shown by an industrial electromagnet that can lift thousands of pounds of metal.

Magnetism and Its Historical Discoveries

By the end of this section, you will be able to:

- Explain attraction and repulsion by magnets
- Describe the historical and contemporary applications of magnetism

Magnetism has been known since the time of the ancient Greeks, but it has always been a bit mysterious. You can see electricity in the flash of a lightning bolt, but when a compass needle points to magnetic north, you can't see any force causing it to rotate. People learned about magnetic properties gradually, over many years, before several physicists of the nineteenth century connected magnetism with electricity. In this section, we review the basic ideas of magnetism and describe how they fit into the picture of a magnetic field.

Brief History of Magnetism

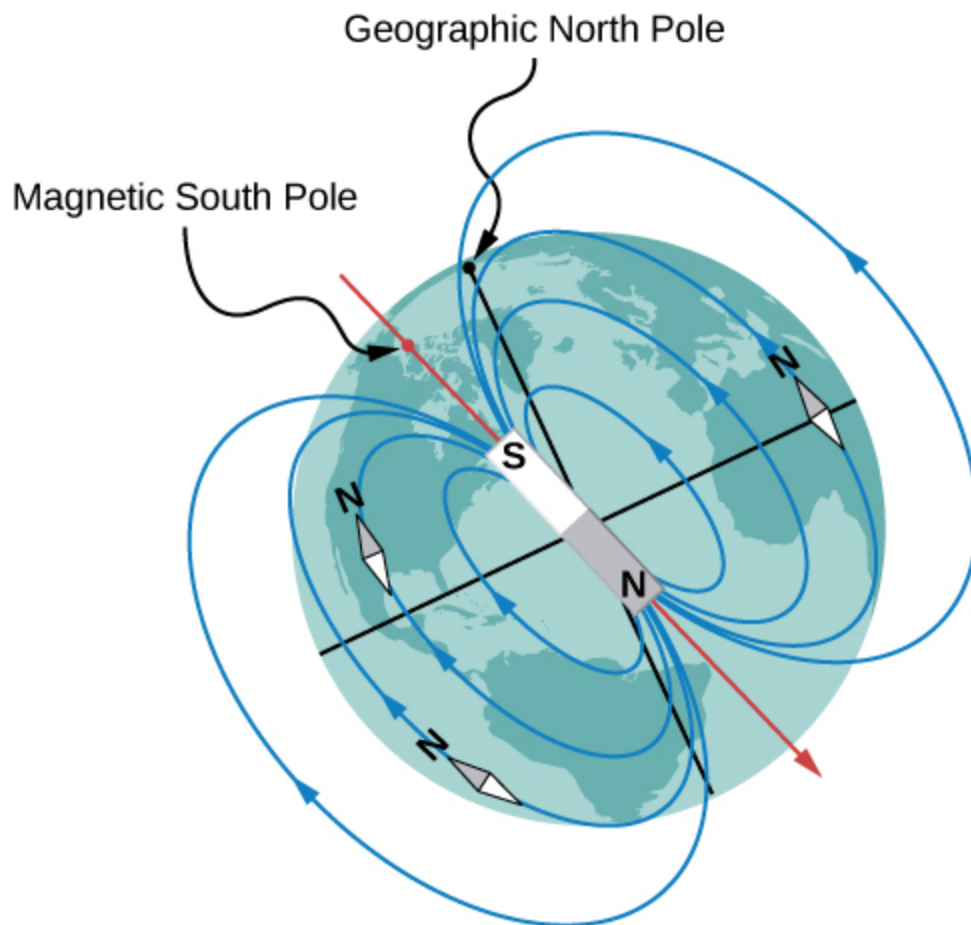
Magnets are commonly found in everyday objects, such as toys, hangers, elevators, doorbells, and computer devices. Experimentation on these magnets shows that all magnets have two poles: One is labeled north (N) and the other is labeled south (S). Magnetic poles repel if they are alike (both N or both S), they attract if they are opposite (one N and the other S), and both poles of a magnet attract unmagnetized pieces of iron. An important point to note here is that you cannot isolate an individual magnetic pole. Every piece of a magnet, no matter how small, which contains a north pole must also contain a south pole.

Note:

Visit this [website](#) for an interactive demonstration of magnetic north and south poles.

An example of a magnet is a compass needle. It is simply a thin bar magnet suspended at its center, so it is free to rotate in a horizontal plane. Earth

itself also acts like a very large bar magnet, with its south-seeking pole near the geographic North Pole ([\[link\]](#)). The north pole of a compass is attracted toward Earth's geographic North Pole because the magnetic pole that is near the geographic North Pole is actually a south magnetic pole. Confusion arises because the geographic term “North Pole” has come to be used (incorrectly) for the magnetic pole that is near the North Pole. Thus, “**north magnetic pole**” is actually a misnomer—it should be called the **south magnetic pole**. [Note that the orientation of Earth's magnetic field is not permanent but changes (“flips”) after long time intervals. Eventually, Earth's north magnetic pole may be located near its geographic North Pole.]



The north pole of a compass needle points toward the south pole of a magnet, which is how today's magnetic field is oriented from inside Earth. It also points

toward Earth's geographic North Pole because the geographic North Pole is near the magnetic south pole.

Back in 1819, the Danish physicist Hans Oersted was performing a lecture demonstration for some students and noticed that a compass needle moved whenever current flowed in a nearby wire. Further investigation of this phenomenon convinced Oersted that an electric current could somehow cause a magnetic force. He reported this finding to an 1820 meeting of the French Academy of Science.

Soon after this report, Oersted's investigations were repeated and expanded upon by other scientists. Among those whose work was especially important were Jean-Baptiste Biot and Felix Savart, who investigated the forces exerted on magnets by currents; André Marie Ampère, who studied the forces exerted by one current on another; François Arago, who found that iron could be magnetized by a current; and Humphry Davy, who discovered that a magnet exerts a force on a wire carrying an electric current. Within 10 years of Oersted's discovery, Michael Faraday found that the relative motion of a magnet and a metallic wire induced current in the wire. This finding showed not only that a current has a magnetic effect, but that a magnet can generate electric current. You will see later that the names of Biot, Savart, Ampère, and Faraday are linked to some of the fundamental laws of electromagnetism.

The evidence from these various experiments led Ampère to propose that electric current is the source of all magnetic phenomena. To explain permanent magnets, he suggested that matter contains microscopic current loops that are somehow aligned when a material is magnetized. Today, we know that permanent magnets are actually created by the alignment of spinning electrons, a situation quite similar to that proposed by Ampère. This model of permanent magnets was developed by Ampère almost a century before the atomic nature of matter was understood. (For a full quantum mechanical treatment of magnetic spins, see [Quantum Mechanics](#) and [Atomic Structure](#).)

Contemporary Applications of Magnetism

Today, magnetism plays many important roles in our lives. Physicists' understanding of magnetism has enabled the development of technologies that affect both individuals and society. The electronic tablet in your purse or backpack, for example, wouldn't have been possible without the applications of magnetism and electricity on a small scale ([\[link\]](#)). Weak changes in a magnetic field in a thin film of iron and chromium were discovered to bring about much larger changes in resistance, called giant magnetoresistance. Information can then be recorded magnetically based on the direction in which the iron layer is magnetized. As a result of the discovery of giant magnetoresistance and its applications to digital storage, the 2007 Nobel Prize in Physics was awarded to Albert Fert from France and Peter Grunberg from Germany.



Engineering technology like computer storage would not be possible without a deep understanding of magnetism. (credit: Klaus Eifert)

All electric motors—with uses as diverse as powering refrigerators, starting cars, and moving elevators—contain magnets. Generators, whether producing hydroelectric power or running bicycle lights, use magnetic fields. Recycling facilities employ magnets to separate iron from other refuse. Research into using magnetic containment of fusion as a future energy source has been continuing for several years. Magnetic resonance imaging (MRI) has become an important diagnostic tool in the field of medicine, and the use of magnetism to explore brain activity is a subject of contemporary research and development. The list of applications also includes computer hard drives, tape recording, detection of inhaled asbestos, and levitation of high-speed trains. Magnetism is involved in the structure of atomic energy levels, as well as the motion of cosmic rays and charged particles trapped in the Van Allen belts around Earth. Once again, we see that all these disparate phenomena are linked by a small number of underlying physical principles.

Summary

- Magnets have two types of magnetic poles, called the north magnetic pole and the south magnetic pole. North magnetic poles are those that are attracted toward Earth's geographic North Pole.
- Like poles repel and unlike poles attract.
- Discoveries of how magnets respond to currents by Oersted and others created a framework that led to the invention of modern electronic devices, electric motors, and magnetic imaging technology.

Glossary

north magnetic pole

currently where a compass points to north, near the geographic North Pole; this is the effective south pole of a bar magnet but has flipped between the effective north and south poles of a bar magnet multiple times over the age of Earth

south magnetic pole

currently where a compass points to the south, near the geographic South Pole; this is the effective north pole of a bar magnet but has

flipped just like the north magnetic pole

Magnetic Fields and Lines

By the end of this section, you will be able to:

- Define the magnetic field based on a moving charge experiencing a force
- Apply the right-hand rule to determine the direction of a magnetic force based on the motion of a charge in a magnetic field
- Sketch magnetic field lines to understand which way the magnetic field points and how strong it is in a region of space

We have outlined the properties of magnets, described how they behave, and listed some of the applications of magnetic properties. Even though there are no such things as isolated magnetic charges, we can still define the attraction and repulsion of magnets as based on a field. In this section, we define the magnetic field, determine its direction based on the right-hand rule, and discuss how to draw magnetic field lines.

Defining the Magnetic Field

A magnetic field is defined by the force that a charged particle experiences moving in this field, after we account for the gravitational and any additional electric forces possible on the charge. The magnitude of this force is proportional to the amount of charge q , the speed of the charged particle v , and the magnitude of the applied magnetic field. The direction of this force is perpendicular to both the direction of the moving charged particle and the direction of the applied magnetic field. Based on these observations, we define the magnetic field strength B based on the **magnetic force** \vec{F} on a charge q moving at velocity \vec{v} as the cross product of the velocity and magnetic field, that is,

Note:

Equation:

$$\vec{F} = q\vec{v} \times \vec{B}.$$

In fact, this is how we define the magnetic field \vec{B} —in terms of the force on a charged particle moving in a magnetic field. The magnitude of the force is determined from the definition of the cross product as it relates to the magnitudes of each of the vectors. In other words, the magnitude of the force satisfies

Note:

Equation:

$$F = qvB\sin\theta$$

where θ is the angle between the velocity and the magnetic field.

The SI unit for magnetic field strength B is called the **tesla** (T) after the eccentric but brilliant inventor Nikola Tesla (1856–1943), where

Equation:

$$1 \text{ T} = \frac{1 \text{ N}}{\text{A} \cdot \text{m}}.$$

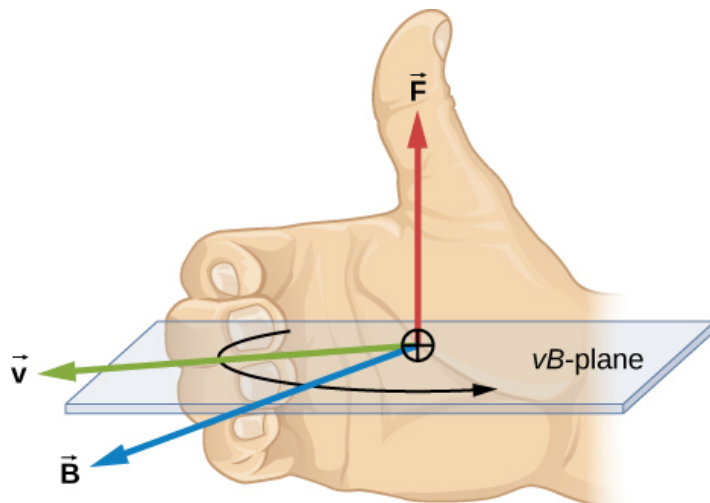
A smaller unit, called the **gauss** (G), where $1 \text{ G} = 10^{-4} \text{ T}$, is sometimes used. The strongest permanent magnets have fields near 2 T; superconducting electromagnets may attain 10 T or more. Earth's magnetic field on its surface is only about $5 \times 10^{-5} \text{ T}$, or 0.5 G.

Note:

Direction of the Magnetic Field by the Right-Hand Rule

The direction of the magnetic force \vec{F} is perpendicular to the plane formed by \vec{v} and \vec{B} , as determined by the **right-hand rule-1** (or RHR-1), which is illustrated in [\[link\]](#).

1. Orient your right hand so that your fingers curl in the plane defined by the velocity and magnetic field vectors.
2. Using your right hand, sweep from the velocity toward the magnetic field with your fingers through the smallest angle possible.
3. The magnetic force is directed where your thumb is pointing.
4. If the charge was negative, reverse the direction found by these steps.



Magnetic fields exert forces on moving charges. The direction of the magnetic force on a moving charge is perpendicular to the plane formed by \vec{v} and \vec{B} and

follows the right-hand rule-1 (RHR-1) as shown.

The magnitude of the force is proportional to q , v , B , and the sine of the angle between \vec{v} and \vec{B} .

Note:

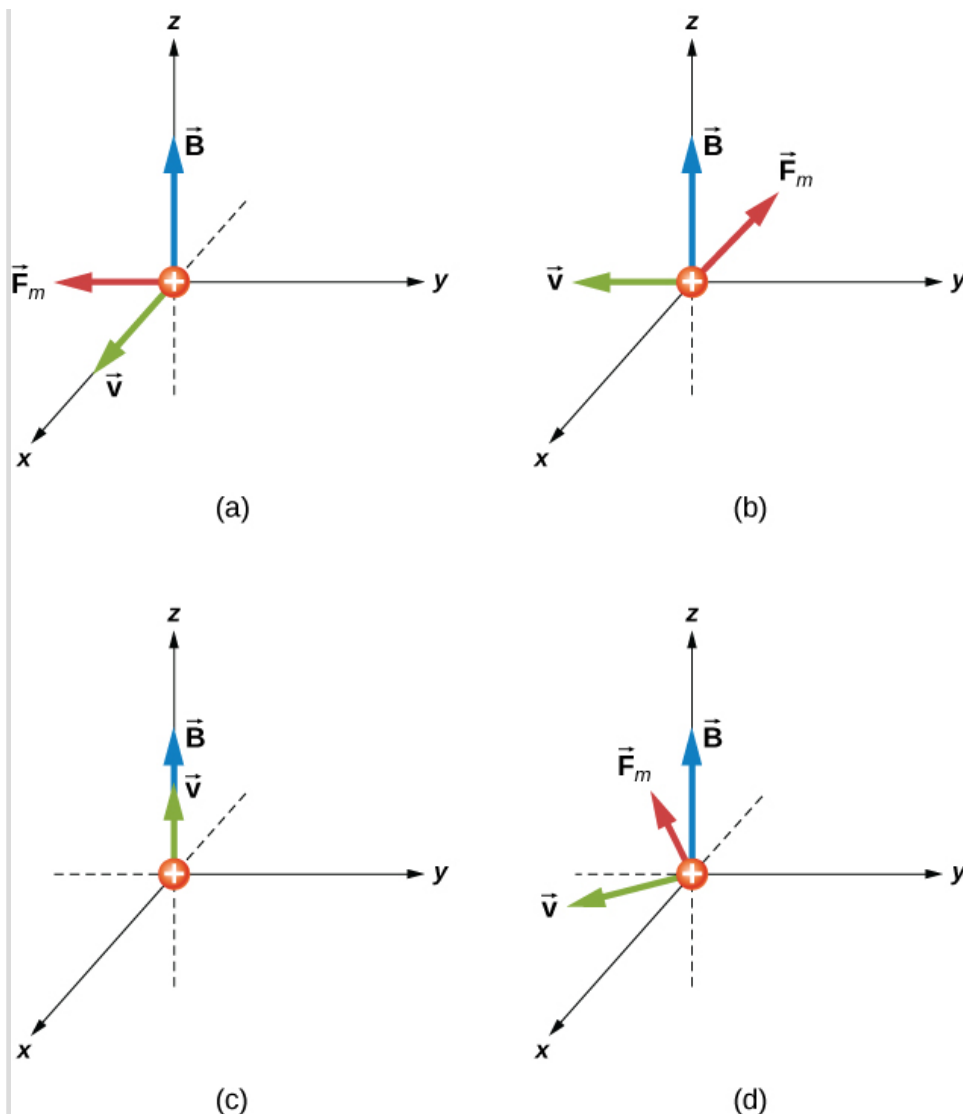
Visit this [website](#) for additional practice with the direction of magnetic fields.

There is no magnetic force on static charges. However, there is a magnetic force on charges moving at an angle to a magnetic field. When charges are stationary, their electric fields do not affect magnets. However, when charges move, they produce magnetic fields that exert forces on other magnets. When there is relative motion, a connection between electric and magnetic forces emerges—each affects the other.

Example:

An Alpha-Particle Moving in a Magnetic Field

An alpha-particle ($q = 3.2 \times 10^{-19}\text{C}$) moves through a uniform magnetic field whose magnitude is 1.5 T. The field is directly parallel to the positive z-axis of the rectangular coordinate system of [\[link\]](#). What is the magnetic force on the alpha-particle when it is moving (a) in the positive x-direction with a speed of $5.0 \times 10^4\text{m/s}$? (b) in the negative y-direction with a speed of $5.0 \times 10^4\text{m/s}$? (c) in the positive z-direction with a speed of $5.0 \times 10^4\text{m/s}$? (d) with a velocity $\vec{v} = (2.0\hat{i} - 3.0\hat{j} + 1.0\hat{k}) \times 10^4\text{m/s}$?



The magnetic forces on an alpha-particle moving in a uniform magnetic field. The field is the same in each drawing, but the velocity is different.

Strategy

We are given the charge, its velocity, and the magnetic field strength and direction. We can thus use the equation $\vec{F} = q\vec{v} \times \vec{B}$ or $F = qvB\sin\theta$ to calculate the force. The direction of the force is determined by RHR-1.

Solution

- First, to determine the direction, start with your fingers pointing in the positive x -direction. Sweep your fingers upward in the direction of magnetic field. Your thumb should point in the negative y -direction. This should match the mathematical answer. To calculate the force, we use the given charge, velocity, and magnetic field and the definition of the magnetic force in cross-product form to calculate:

Equation:

$$\vec{\mathbf{F}} = q\vec{\mathbf{v}} \times \vec{\mathbf{B}} = (3.2 \times 10^{-19}\text{C}) (5.0 \times 10^4\text{m/s} \hat{\mathbf{i}}) \times (1.5\text{ T} \hat{\mathbf{k}}) = -2.4 \times 10^{-14}\text{N} \hat{\mathbf{j}}.$$

- b. First, to determine the directionality, start with your fingers pointing in the negative y -direction. Sweep your fingers upward in the direction of magnetic field as in the previous problem. Your thumb should be open in the negative x -direction. This should match the mathematical answer. To calculate the force, we use the given charge, velocity, and magnetic field and the definition of the magnetic force in cross-product form to calculate:

Equation:

$$\vec{\mathbf{F}} = q\vec{\mathbf{v}} \times \vec{\mathbf{B}} = (3.2 \times 10^{-19}\text{C}) (-5.0 \times 10^4\text{m/s} \hat{\mathbf{j}}) \times (1.5\text{ T} \hat{\mathbf{k}}) = -2.4 \times 10^{-14}\text{N} \hat{\mathbf{i}}.$$

An alternative approach is to use [\[link\]](#) to find the magnitude of the force. This applies for both parts (a) and (b). Since the velocity is perpendicular to the magnetic field, the angle between them is 90 degrees. Therefore, the magnitude of the force is:

Equation:

$$F = qvB\sin\theta = (3.2 \times 10^{-19}\text{C})(5.0 \times 10^4\text{m/s})(1.5\text{ T})\sin(90^\circ) = 2.4 \times 10^{-14}\text{N}.$$

- c. Since the velocity and magnetic field are parallel to each other, there is no orientation of your hand that will result in a force direction. Therefore, the force on this moving charge is zero. This is confirmed by the cross product. When you cross two vectors pointing in the same direction, the result is equal to zero.
- d. First, to determine the direction, your fingers could point in any orientation; however, you must sweep your fingers upward in the direction of the magnetic field. As you rotate your hand, notice that the thumb can point in any x - or y -direction possible, but not in the z -direction. This should match the mathematical answer. To calculate the force, we use the given charge, velocity, and magnetic field and the definition of the magnetic force in cross-product form to calculate:

Equation:

$$\begin{aligned}\vec{\mathbf{F}} &= q\vec{\mathbf{v}} \times \vec{\mathbf{B}} = (3.2 \times 10^{-19}\text{C}) \left((2.0\hat{\mathbf{i}} - 3.0\hat{\mathbf{j}} + 1.0\hat{\mathbf{k}}) \times 10^4\text{m/s} \right) \times (1.5\text{ T} \hat{\mathbf{k}}) \\ &= (-14.4\hat{\mathbf{i}} - 9.6\hat{\mathbf{j}}) \times 10^{-15}\text{N}.\end{aligned}$$

This solution can be rewritten in terms of a magnitude and angle in the xy -plane:

Equation:

$$\begin{aligned}|\vec{\mathbf{F}}| &= \sqrt{F_x^2 + F_y^2} = \sqrt{(-14.4)^2 + (-9.6)^2} \times 10^{-15}\text{N} = 1.7 \times 10^{-14}\text{N} \\ \theta &= \tan^{-1} \left(\frac{F_y}{F_x} \right) = \tan^{-1} \left(\frac{-9.6 \times 10^{-15}\text{N}}{-14.4 \times 10^{-15}\text{N}} \right) = 34^\circ.\end{aligned}$$

The magnitude of the force can also be calculated using [\[link\]](#). The velocity in this question, however, has three components. The z -component of the velocity can be neglected, because it is parallel to the magnetic field and therefore generates no force. The magnitude of the velocity is calculated from the x - and y -components. The angle between the velocity in the xy -plane and the magnetic field in the z -plane is 90 degrees. Therefore, the force is calculated to be:

Equation:

$$|\vec{v}| = \sqrt{(2)^2 + (-3)^2} \times 10^4 \frac{\text{m}}{\text{s}} = 3.6 \times 10^4 \frac{\text{m}}{\text{s}}$$

$$F = qvB\sin\theta = (3.2 \times 10^{-19}\text{C})(3.6 \times 10^4\text{m/s})(1.5\text{ T})\sin(90^\circ) = 1.7 \times 10^{-14}\text{N}.$$

This is the same magnitude of force calculated by unit vectors.

Significance

The cross product in this formula results in a third vector that must be perpendicular to the other two. Other physical quantities, such as angular momentum, also have three vectors that are related by the cross product. Note that typical force values in magnetic force problems are much larger than the gravitational force. Therefore, for an isolated charge, the magnetic force is the dominant force governing the charge's motion.

Note:

Exercise:

Problem:

Check Your Understanding Repeat the previous problem with the magnetic field in the x-direction rather than in the z-direction. Check your answers with RHR-1.

Solution:

a. 0 N; b. $2.4 \times 10^{-14}\hat{\mathbf{k}}\text{N}$; c. $2.4 \times 10^{-14}\hat{\mathbf{j}}\text{N}$; d. $(7.2\hat{\mathbf{j}} + 2.2\hat{\mathbf{k}}) \times 10^{-15}\text{N}$

Representing Magnetic Fields

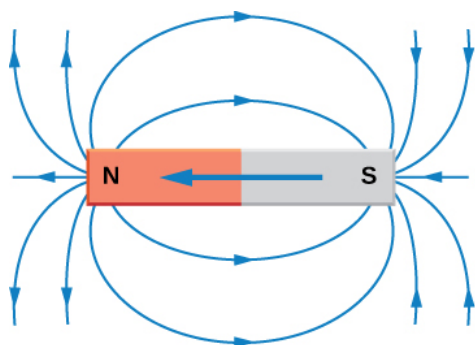
The representation of magnetic fields by **magnetic field lines** is very useful in visualizing the strength and direction of the magnetic field. As shown in [\[link\]](#), each of these lines forms a closed loop, even if not shown by the constraints of the space available for the figure. The field lines emerge from the north pole (N), loop around to the south pole (S), and continue through the bar magnet back to the north pole.

Magnetic field lines have several hard-and-fast rules:

1. The direction of the magnetic field is tangent to the field line at any point in space. A small compass will point in the direction of the field line.
2. The strength of the field is proportional to the closeness of the lines. It is exactly proportional to the number of lines per unit area perpendicular to the lines (called the areal density).
3. Magnetic field lines can never cross, meaning that the field is unique at any point in space.
4. Magnetic field lines are continuous, forming closed loops without a beginning or end. They are directed from the north pole to the south pole.

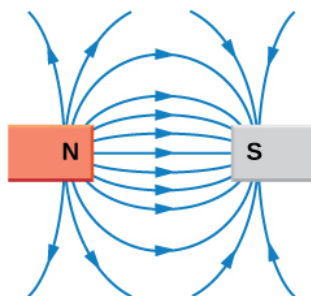
The last property is related to the fact that the north and south poles cannot be separated. It is a distinct difference from electric field lines, which generally begin on positive charges and end on

negative charges or at infinity. If isolated magnetic charges (referred to as magnetic monopoles) existed, then magnetic field lines would begin and end on them.



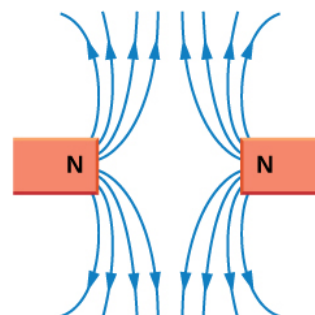
Magnetic field lines of a bar magnet

(a)



Magnetic field lines
between unlike poles

(b)



Magnetic field lines
between like poles

(c)

Magnetic field lines are defined to have the direction in which a small compass points when placed at a location in the field. The strength of the field is proportional to the closeness (or density) of the lines. If the interior of the magnet could be probed, the field lines would be found to form continuous, closed loops. To fit in a reasonable space, some of these drawings may not show the closing of the loops; however, if enough space were provided, the loops would be closed.

Summary

- Charges moving across a magnetic field experience a force determined by $\vec{F} = q\vec{v} \times \vec{B}$. The force is perpendicular to the plane formed by \vec{v} and \vec{B} .
- The direction of the force on a moving charge is given by the right hand rule 1 (RHR-1): Sweep your fingers in a velocity, magnetic field plane. Start by pointing them in the direction of velocity and sweep towards the magnetic field. Your thumb points in the direction of the magnetic force for positive charges.
- Magnetic fields can be pictorially represented by magnetic field lines, which have the following properties:
 - The field is tangent to the magnetic field line.
 - Field strength is proportional to the line density.
 - Field lines cannot cross.
 - Field lines form continuous, closed loops.
- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.

Conceptual Questions

Exercise:**Problem:**

Discuss the similarities and differences between the electrical force on a charge and the magnetic force on a charge.

Solution:

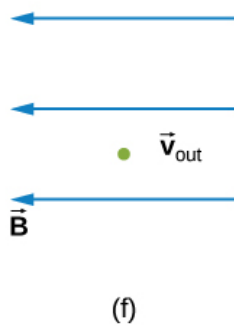
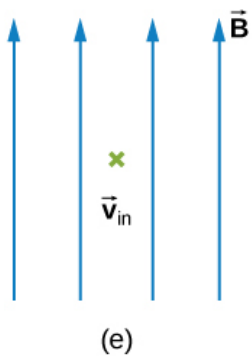
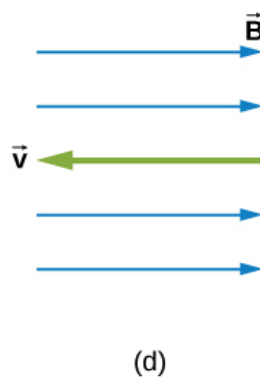
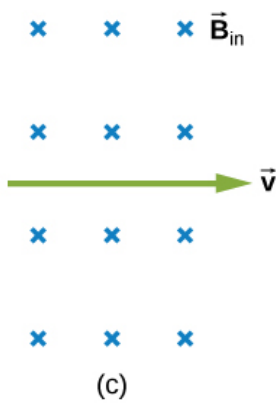
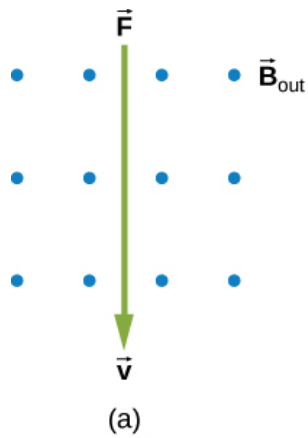
Both are field dependent. Electrical force is dependent on charge, whereas magnetic force is dependent on current or rate of charge flow.

Exercise:**Problem:**

(a) Is it possible for the magnetic force on a charge moving in a magnetic field to be zero? (b) Is it possible for the electric force on a charge moving in an electric field to be zero? (c) Is it possible for the resultant of the electric and magnetic forces on a charge moving simultaneously through both fields to be zero?

Problems**Exercise:****Problem:**

What is the direction of the magnetic force on a positive charge that moves as shown in each of the six cases?



Solution:

a. left; b. into the page; c. up the page; d. no force; e. right; f. down

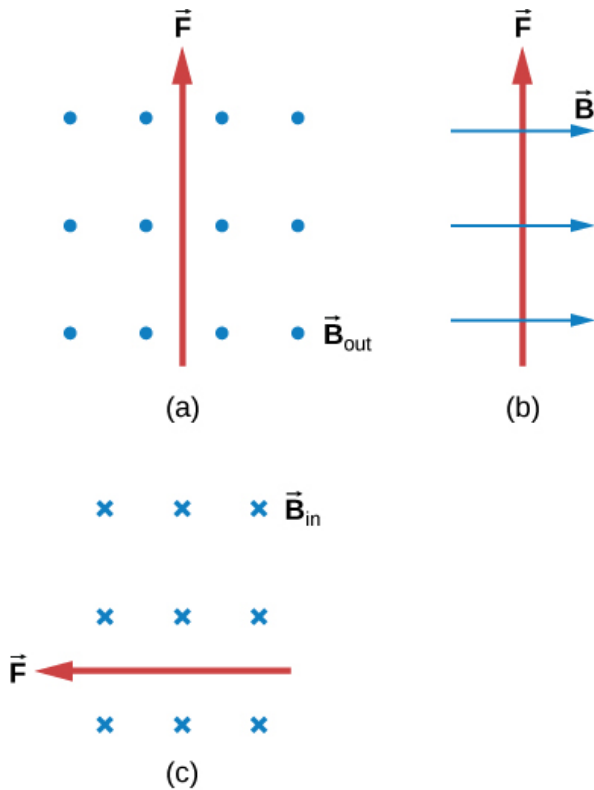
Exercise:

Problem: Repeat previous exercise for a negative charge.

Exercise:

Problem:

What is the direction of the velocity of a negative charge that experiences the magnetic force shown in each of the three cases, assuming it moves perpendicular to B ?



Solution:

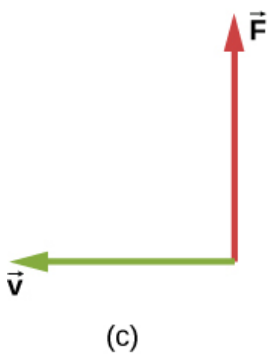
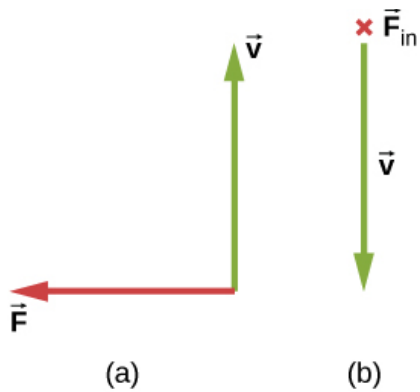
a. right; b. into the page; c. down

Exercise:

Problem: Repeat previous exercise for a positive charge.

Exercise:**Problem:**

What is the direction of the magnetic field that produces the magnetic force on a positive charge as shown in each of the three cases, assuming \vec{B} is perpendicular to \vec{v} ?



Solution:

a. into the page; b. left; c. out of the page

Exercise:

Problem: Repeat previous exercise for a negative charge.

Exercise:

Problem:

(a) Aircraft sometimes acquire small static charges. Suppose a supersonic jet has a $0.500\text{-}\mu\text{C}$ charge and flies due west at a speed of 660 m/s over Earth's south magnetic pole, where the $8.00 \times 10^{-5}\text{ T}$ magnetic field points straight down into the ground. What are the direction and the magnitude of the magnetic force on the plane? (b) Discuss whether the value obtained in part (a) implies this is a significant or negligible effect.

Solution:

a. $2.64 \times 10^{-8}\text{ N}$; north b. The force is very small, so this implies that the effect of static charges on airplanes is negligible.

Exercise:

Problem:

(a) A cosmic ray proton moving toward Earth at $5.00 \times 10^7 \text{ m/s}$ experiences a magnetic force of $1.70 \times 10^{-16} \text{ N}$. What is the strength of the magnetic field if there is a 45° angle between it and the proton's velocity? (b) Is the value obtained in part a. consistent with the known strength of Earth's magnetic field on its surface? Discuss.

Exercise:**Problem:**

An electron moving at $4.00 \times 10^3 \text{ m/s}$ in a 1.25-T magnetic field experiences a magnetic force of $1.40 \times 10^{-16} \text{ N}$. What angle does the velocity of the electron make with the magnetic field? There are two answers.

Solution:

10.1° ; 169.9°

Exercise:**Problem:**

(a) A physicist performing a sensitive measurement wants to limit the magnetic force on a moving charge in her equipment to less than $1.00 \times 10^{-12} \text{ N}$. What is the greatest the charge can be if it moves at a maximum speed of 30.0 m/s in Earth's field? (b) Discuss whether it would be difficult to limit the charge to less than the value found in (a) by comparing it with typical static electricity and noting that static is often absent.

Glossary

gauss

G, unit of the magnetic field strength; $1 \text{ G} = 10^{-4} \text{ T}$

magnetic field lines

continuous curves that show the direction of a magnetic field; these lines point in the same direction as a compass points, toward the magnetic south pole of a bar magnet

magnetic force

force applied to a charged particle moving through a magnetic field

right-hand rule-1

using your right hand to determine the direction of either the magnetic force, velocity of a charged particle, or magnetic field

tesla

SI unit for magnetic field: $1 \text{ T} = 1 \text{ N/A}\cdot\text{m}$

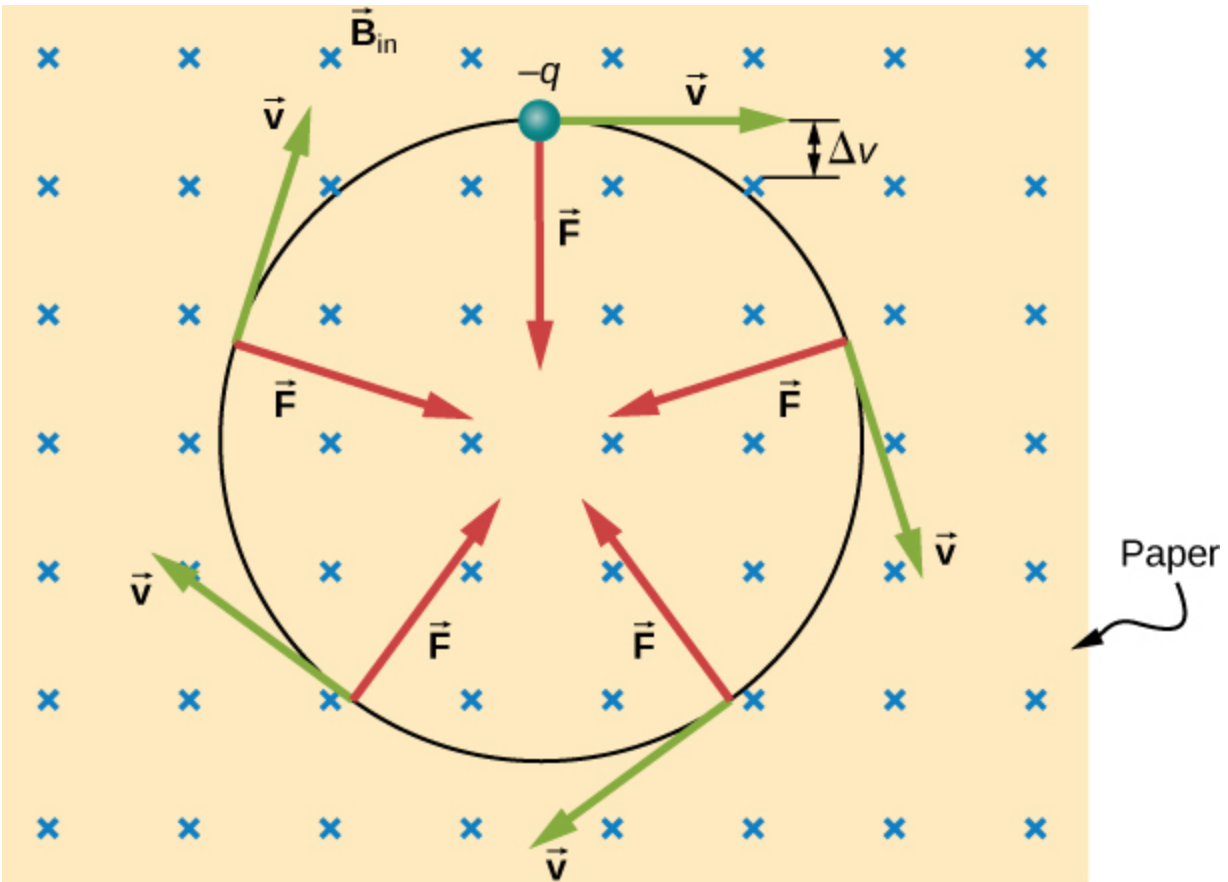
Motion of a Charged Particle in a Magnetic Field

By the end of this section, you will be able to:

- Explain how a charged particle in an external magnetic field undergoes circular motion
- Describe how to determine the radius of the circular motion of a charged particle in a magnetic field

A charged particle experiences a force when moving through a magnetic field. What happens if this field is uniform over the motion of the charged particle? What path does the particle follow? In this section, we discuss the circular motion of the charged particle as well as other motion that results from a charged particle entering a magnetic field.

The simplest case occurs when a charged particle moves perpendicular to a uniform B -field ([\[link\]](#)). If the field is in a vacuum, the magnetic field is the dominant factor determining the motion. Since the magnetic force is perpendicular to the direction of travel, a charged particle follows a curved path in a magnetic field. The particle continues to follow this curved path until it forms a complete circle. Another way to look at this is that the magnetic force is always perpendicular to velocity, so that it does no work on the charged particle. The particle's kinetic energy and speed thus remain constant. The direction of motion is affected but not the speed.



A negatively charged particle moves in the plane of the paper in a region where the magnetic field is perpendicular to the paper (represented by the small \times 's—like the tails of arrows). The magnetic force is perpendicular to the velocity, so velocity changes in direction but not magnitude. The result is uniform circular motion. (Note that because the charge is negative, the force is opposite in direction to the prediction of the right-hand rule.)

In this situation, the magnetic force supplies the centripetal force $F_c = \frac{mv^2}{r}$. Noting that the velocity is perpendicular to the magnetic field, the magnitude of the magnetic force is reduced to $F = qvB$. Because the magnetic force F supplies the centripetal force F_c , we have

Equation:

$$qvB = \frac{mv^2}{r}.$$

Solving for r yields

Note:

Equation:

$$r = \frac{mv}{qB}.$$

Here, r is the radius of curvature of the path of a charged particle with mass m and charge q , moving at a speed v that is perpendicular to a magnetic field of strength B . The time for the charged particle to go around the circular path is defined as the period, which is the same as the distance traveled (the circumference) divided by the speed. Based on this and [\[link\]](#), we can derive the period of motion as

Note:

Equation:

$$T = \frac{2\pi r}{v} = \frac{2\pi}{v} \frac{mv}{qB} = \frac{2\pi m}{qB}.$$

If the velocity is not perpendicular to the magnetic field, then we can compare each component of the velocity separately with the magnetic field. The component of the velocity perpendicular to the magnetic field produces a magnetic force perpendicular to both this velocity and the field:

Equation:

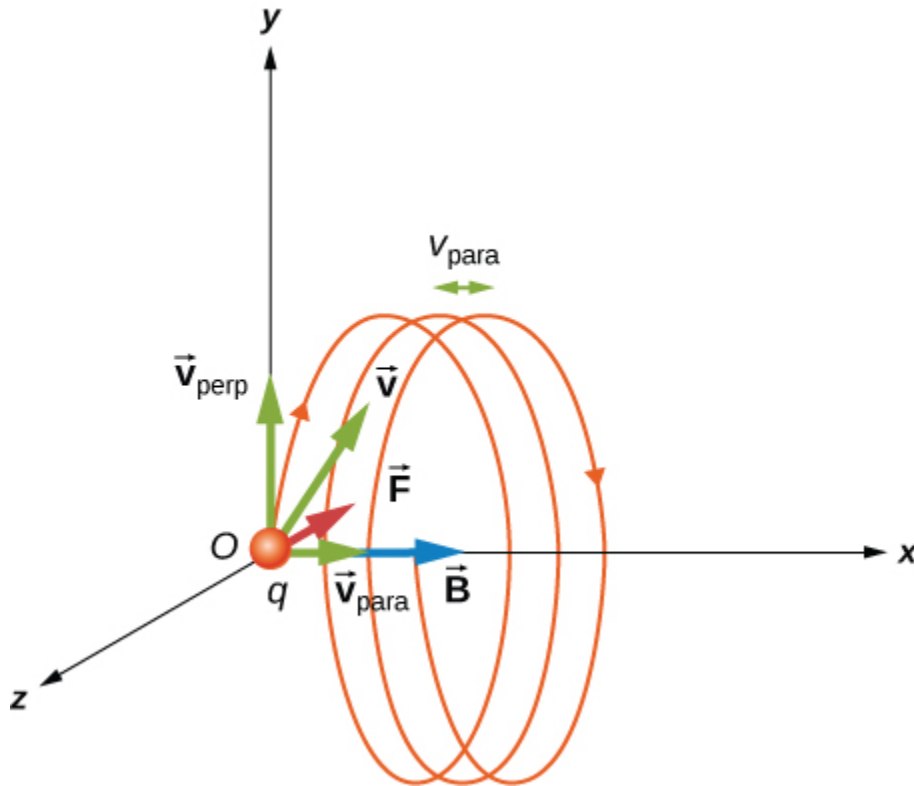
$$v_{\text{perp}} = v \sin \theta, \quad v_{\text{para}} = v \cos \theta.$$

where θ is the angle between v and B . The component parallel to the magnetic field creates constant motion along the same direction as the magnetic field, also shown in [\[link\]](#). The parallel motion determines the *pitch* p of the helix, which is the distance between adjacent turns. This distance equals the parallel component of the velocity times the period:

Equation:

$$p = v_{\text{para}} T.$$

The result is a **helical motion**, as shown in the following figure.

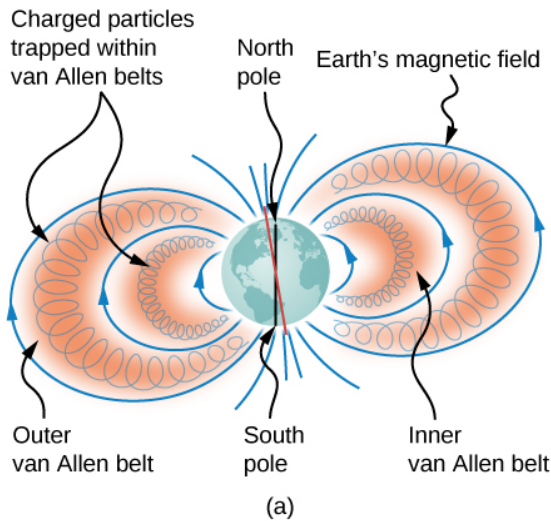


A charged particle moving with a velocity not in the same direction as the magnetic field. The velocity component perpendicular to the magnetic field creates circular motion, whereas the

component of the velocity parallel to the field moves the particle along a straight line. The pitch is the horizontal distance between two consecutive circles. The resulting motion is helical.

While the charged particle travels in a helical path, it may enter a region where the magnetic field is not uniform. In particular, suppose a particle travels from a region of strong magnetic field to a region of weaker field, then back to a region of stronger field. The particle may reflect back before entering the stronger magnetic field region. This is similar to a wave on a string traveling from a very light, thin string to a hard wall and reflecting backward. If the reflection happens at both ends, the particle is trapped in a so-called magnetic bottle.

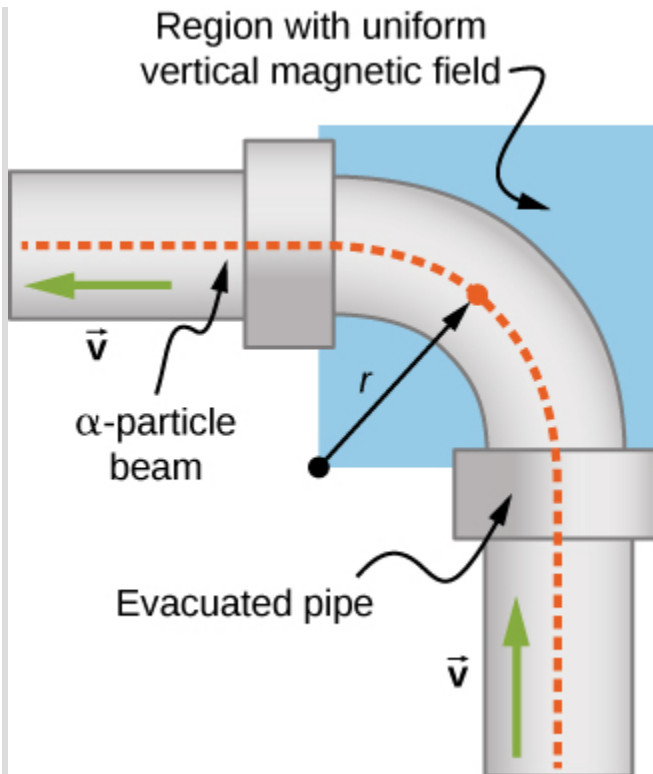
Trapped particles in magnetic fields are found in the Van Allen radiation belts around Earth, which are part of Earth's magnetic field. These belts were discovered by James Van Allen while trying to measure the flux of **cosmic rays** on Earth (high-energy particles that come from outside the solar system) to see whether this was similar to the flux measured on Earth. Van Allen found that due to the contribution of particles trapped in Earth's magnetic field, the flux was much higher on Earth than in outer space. Aurorae, like the famous aurora borealis (northern lights) in the Northern Hemisphere ([\[link\]](#)), are beautiful displays of light emitted as ions recombine with electrons entering the atmosphere as they spiral along magnetic field lines. (The ions are primarily oxygen and nitrogen atoms that are initially ionized by collisions with energetic particles in Earth's atmosphere.) Aurorae have also been observed on other planets, such as Jupiter and Saturn.



(a) The Van Allen radiation belts around Earth trap ions produced by cosmic rays striking Earth's atmosphere. (b) The magnificent spectacle of the aurora borealis, or northern lights, glows in the northern sky above Bear Lake near Eielson Air Force Base, Alaska. Shaped by Earth's magnetic field, this light is produced by glowing molecules and ions of oxygen and nitrogen. (credit b: modification of work by USAF Senior Airman Joshua Strang)

Example: Beam Deflector

A research group is investigating short-lived radioactive isotopes. They need to design a way to transport alpha-particles (helium nuclei) from where they are made to a place where they will collide with another material to form an isotope. The beam of alpha-particles ($m = 6.64 \times 10^{-27}\text{kg}$, $q = 3.2 \times 10^{-19}\text{C}$) bends through a 90-degree region with a uniform magnetic field of 0.050 T ([link](#)). (a) In what direction should the magnetic field be applied? (b) How much time does it take the alpha-particles to traverse the uniform magnetic field region?



Top view of the beam deflector setup.

Strategy

- The direction of the magnetic field is shown by the RHR-1. Your fingers point in the direction of \vec{v} , and your thumb needs to point in the direction of the force, to the left. Therefore, since the alpha-particles are positively charged, the magnetic field must point down.
- The period of the alpha-particle going around the circle is

Note:

Equation:

$$T = \frac{2\pi m}{qB}.$$

Because the particle is only going around a quarter of a circle, we can take 0.25 times the period to find the time it takes to go around this path.

Solution

- a. Let's start by focusing on the alpha-particle entering the field near the bottom of the picture. First, point your thumb up the page. In order for your palm to open to the left where the centripetal force (and hence the magnetic force) points, your fingers need to change orientation until they point into the page. This is the direction of the applied magnetic field.
- b. The period of the charged particle going around a circle is calculated by using the given mass, charge, and magnetic field in the problem. This works out to be

Equation:

$$T = \frac{2\pi m}{qB} = \frac{2\pi (6.64 \times 10^{-27} \text{kg})}{(3.2 \times 10^{-19} \text{C}) (0.050 \text{ T})} = 2.6 \times 10^{-6} \text{s}.$$

However, for the given problem, the alpha-particle goes around a quarter of the circle, so the time it takes would be

Equation:

$$t = 0.25 \times 2.61 \times 10^{-6} \text{s} = 6.5 \times 10^{-7} \text{s}.$$

Significance

This time may be quick enough to get to the material we would like to bombard, depending on how short-lived the radioactive isotope is and continues to emit alpha-particles. If we could increase the magnetic field applied in the region, this would shorten the time even more. The path the particles need to take could be shortened, but this may not be economical given the experimental setup.

Note:**Exercise:****Problem:**

Check Your Understanding A uniform magnetic field of magnitude 1.5 T is directed horizontally from west to east. (a) What is the magnetic force on a proton at the instant when it is moving vertically downward in the field with a speed of 4×10^7 m/s? (b) Compare this force with the weight w of a proton.

Solution:

a. 9.6×10^{-12} N toward the south; b. $\frac{w}{F_m} = 1.7 \times 10^{-15}$

Example:**Helical Motion in a Magnetic Field**

A proton enters a uniform magnetic field of 1.0×10^{-4} T with a speed of 5×10^5 m/s. At what angle must the magnetic field be from the velocity so that the pitch of the resulting helical motion is equal to the radius of the helix?

Strategy

The pitch of the motion relates to the parallel velocity times the period of the circular motion, whereas the radius relates to the perpendicular velocity component. After setting the radius and the pitch equal to each other, solve for the angle between the magnetic field and velocity or θ .

Solution

The pitch is given by [\[link\]](#), the period is given by [\[link\]](#), and the radius of circular motion is given by [\[link\]](#). Note that the velocity in the radius equation is related to only the perpendicular velocity, which is where the circular motion occurs. Therefore, we substitute the sine component of the overall velocity into the radius equation to equate the pitch and radius:

Equation:

$$\begin{aligned}
 p &= r \\
 v_{\parallel} T &= \frac{mv_{\perp}}{qB} \\
 v \cos \theta \frac{2\pi m}{qB} &= \frac{mv \sin \theta}{qB} \\
 2\pi &= \tan \theta \\
 \theta &= 81.0^{\circ}.
 \end{aligned}$$

Significance

If this angle were 0° , only parallel velocity would occur and the helix would not form, because there would be no circular motion in the perpendicular plane. If this angle were 90° , only circular motion would occur and there would be no movement of the circles perpendicular to the motion. That is what creates the helical motion.

Summary

- A magnetic force can supply centripetal force and cause a charged particle to move in a circular path of radius $r = \frac{mv}{qB}$.
- The period of circular motion for a charged particle moving in a magnetic field perpendicular to the plane of motion is $T = \frac{2\pi m}{qB}$.
- Helical motion results if the velocity of the charged particle has a component parallel to the magnetic field as well as a component perpendicular to the magnetic field.

Conceptual Questions

Exercise:

Problem:

At a given instant, an electron and a proton are moving with the same velocity in a constant magnetic field. Compare the magnetic forces on these particles. Compare their accelerations.

Solution:

The magnitude of the proton and electron magnetic forces are the same since they have the same amount of charge. The direction of these forces however are opposite of each other. The accelerations are opposite in direction and the electron has a larger acceleration than the proton due to its smaller mass.

Exercise:

Problem:

Does increasing the magnitude of a uniform magnetic field through which a charge is traveling necessarily mean increasing the magnetic force on the charge? Does changing the direction of the field necessarily mean a change in the force on the charge?

Exercise:

Problem:

An electron passes through a magnetic field without being deflected. What do you conclude about the magnetic field?

Solution:

The magnetic field must point parallel or anti-parallel to the velocity.

Exercise:

Problem:

If a charged particle moves in a straight line, can you conclude that there is no magnetic field present?

Exercise:

Problem:

How could you determine which pole of an electromagnet is north and which pole is south?

Solution:

A compass points toward the north pole of an electromagnet.

Problems

Exercise:

Problem:

A cosmic-ray electron moves at $7.5 \times 10^6 \text{ m/s}$ perpendicular to Earth's magnetic field at an altitude where the field strength is $1.0 \times 10^{-5} \text{ T}$. What is the radius of the circular path the electron follows?

Solution:

4.27 m

Exercise:

Problem:

(a) Viewers of Star Trek have heard of an antimatter drive on the Starship *Enterprise*. One possibility for such a futuristic energy source is to store antimatter charged particles in a vacuum chamber, circulating in a magnetic field, and then extract them as needed. Antimatter annihilates normal matter, producing pure energy. What strength magnetic field is needed to hold antiprotons, moving at $5.0 \times 10^7 \text{ m/s}$ in a circular path 2.00 m in radius? Antiprotons have the same mass as protons but the opposite (negative) charge. (b) Is this field strength obtainable with today's technology or is it a futuristic possibility?

Exercise:

Problem:

(a) An oxygen-16 ion with a mass of 2.66×10^{-26} kg travels at 5.0×10^6 m/s perpendicular to a 1.20-T magnetic field, which makes it move in a circular arc with a 0.231-m radius. What positive charge is on the ion? (b) What is the ratio of this charge to the charge of an electron? (c) Discuss why the ratio found in (b) should be an integer.

Solution:

a. 4.80×10^{-19} C; b. 3; c. This ratio must be an integer because charges must be integer numbers of the basic charge of an electron. There are no free charges with values less than this basic charge, and all charges are integer multiples of this basic charge.

Exercise:**Problem:**

An electron in a TV CRT moves with a speed of 6.0×10^6 m/s, in a direction perpendicular to Earth's field, which has a strength of 5.0×10^{-5} T. (a) What strength electric field must be applied perpendicular to the Earth's field to make the electron moves in a straight line? (b) If this is done between plates separated by 1.00 cm, what is the voltage applied? (Note that TVs are usually surrounded by a ferromagnetic material to shield against external magnetic fields and avoid the need for such a correction.)

Exercise:**Problem:**

(a) At what speed will a proton move in a circular path of the same radius as the electron in the previous exercise? (b) What would the radius of the path be if the proton had the same speed as the electron? (c) What would the radius be if the proton had the same kinetic energy as the electron? (d) The same momentum?

Solution:

(a) 3.27×10^4 m/s (b) 12,525 m (c) 292 m (d) 6.83 m.

Exercise:

Problem:

(a) What voltage will accelerate electrons to a speed of 6.00×10^{-7} m/s? (b) Find the radius of curvature of the path of a proton accelerated through this potential in a 0.500-T field and compare this with the radius of curvature of an electron accelerated through the same potential.

Exercise:

Problem:

An alpha-particle ($m = 6.64 \times 10^{-27}$ kg, $q = 3.2 \times 10^{-19}$ C) travels in a circular path of radius 25 cm in a uniform magnetic field of magnitude 1.5 T. (a) What is the speed of the particle? (b) What is the kinetic energy in electron-volts? (c) Through what potential difference must the particle be accelerated in order to give it this kinetic energy?

Solution:

a. 1.8×10^7 m/s; b. 6.8×10^6 eV; c. 3.4×10^6 V

Exercise:

Problem:

A particle of charge q and mass m is accelerated from rest through a potential difference V , after which it encounters a uniform magnetic field B . If the particle moves in a plane perpendicular to B , what is the radius of its circular orbit?

Glossary

cosmic rays

comprised of particles that originate mainly from outside the solar system and reach Earth

helical motion

superposition of circular motion with a straight-line motion that is followed by a charged particle moving in a region of magnetic field at an angle to the field

Magnetic Force on a Current-Carrying Conductor

By the end of this section, you will be able to:

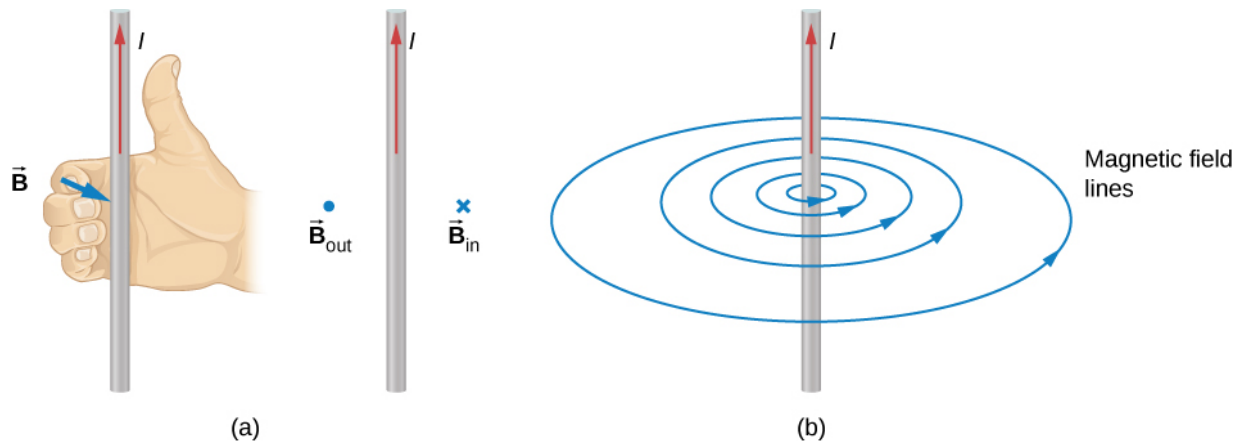
- Determine the direction in which a current-carrying wire experiences a force in an external magnetic field
- Calculate the force on a current-carrying wire in an external magnetic field

Moving charges experience a force in a magnetic field. If these moving charges are in a wire—that is, if the wire is carrying a current—the wire should also experience a force. However, before we discuss the force exerted on a current by a magnetic field, we first examine the magnetic field generated by an electric current. We are studying two separate effects here that interact closely: A current-carrying wire generates a magnetic field and the magnetic field exerts a force on the current-carrying wire.

Magnetic Fields Produced by Electrical Currents

When discussing historical discoveries in magnetism, we mentioned Oersted's finding that a wire carrying an electrical current caused a nearby compass to deflect. A connection was established that electrical currents produce magnetic fields. (This connection between electricity and magnetism is discussed in more detail in [Sources of Magnetic Fields](#).)

The compass needle near the wire experiences a force that aligns the needle tangent to a circle around the wire. Therefore, a current-carrying wire produces circular loops of magnetic field. To determine the direction of the magnetic field generated from a wire, we use a second right-hand rule. In RHR-2, your thumb points in the direction of the current while your fingers wrap around the wire, pointing in the direction of the magnetic field produced ([link](#)). If the magnetic field were coming at you or out of the page, we represent this with a dot. If the magnetic field were going into the page, we represent this with an \times . These symbols come from considering a vector arrow: An arrow pointed toward you, from your perspective, would look like a dot or the tip of an arrow. An arrow pointed away from you, from your perspective, would look like a cross or an \times . A composite sketch of the magnetic circles is shown in [link](#), where the field strength is shown to decrease as you get farther from the wire by loops that are farther separated.



(a) When the wire is in the plane of the paper, the field is perpendicular to the paper. Note the symbols used for the field pointing inward (like the tail of an arrow) and the field pointing outward (like the tip of an arrow). (b) A long and straight wire creates a field with magnetic field lines forming circular loops.

Calculating the Magnetic Force

Electric current is an ordered movement of charge. A current-carrying wire in a magnetic field must therefore experience a force due to the field. To investigate this force, let's consider the infinitesimal section of wire as shown in [\[link\]](#). The length and cross-sectional area of the section are dl and A , respectively, so its volume is $V = A \cdot dl$. The wire is formed from material that contains n charge carriers per unit volume, so the number of charge carriers in the section is $nA \cdot dl$. If the charge carriers move with drift velocity \vec{v}_d , the current I in the wire is (from [Current and Resistance](#))

Equation:

$$I = neAv_d.$$

The magnetic force on any single charge carrier is $e\vec{v}_d \times \vec{B}$, so the total magnetic force $d\vec{F}$ on the $nA \cdot dl$ charge carriers in the section of wire is

Equation:

$$d\vec{\mathbf{F}} = (nA \cdot dl)e\vec{\mathbf{v}}_d \times \vec{\mathbf{B}}.$$

We can define $d\mathbf{l}$ to be a vector of length dl pointing along $\vec{\mathbf{v}}_d$, which allows us to rewrite this equation as

Equation:

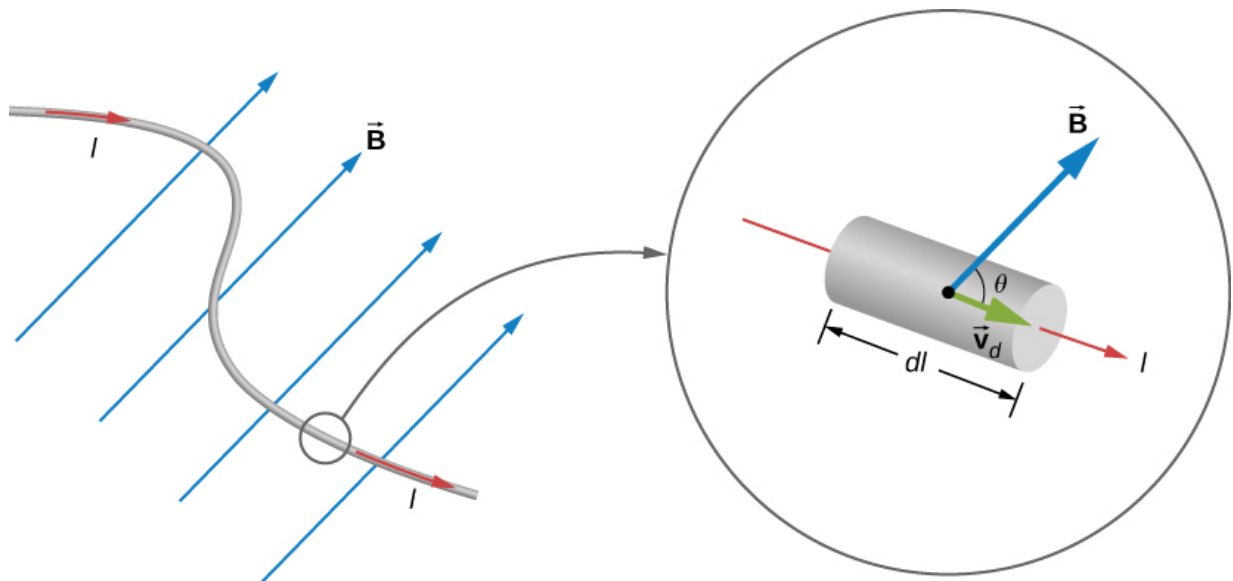
$$d\vec{\mathbf{F}} = neAv_d d\mathbf{l} \times \vec{\mathbf{B}},$$

or

Equation:

$$d\vec{\mathbf{F}} = I d\mathbf{l} \times \vec{\mathbf{B}}.$$

This is the magnetic force on the section of wire. Note that it is actually the net force exerted by the field on the charge carriers themselves. The direction of this force is given by RHR-1, where you point your fingers in the direction of the current and curl them toward the field. Your thumb then points in the direction of the force.



An infinitesimal section of current-carrying wire in a magnetic field.

To determine the magnetic force \vec{F} on a wire of arbitrary length and shape, we must integrate [\[link\]](#) over the entire wire. If the wire section happens to be straight and B is uniform, the equation differentials become absolute quantities, giving us

Note:

Equation:

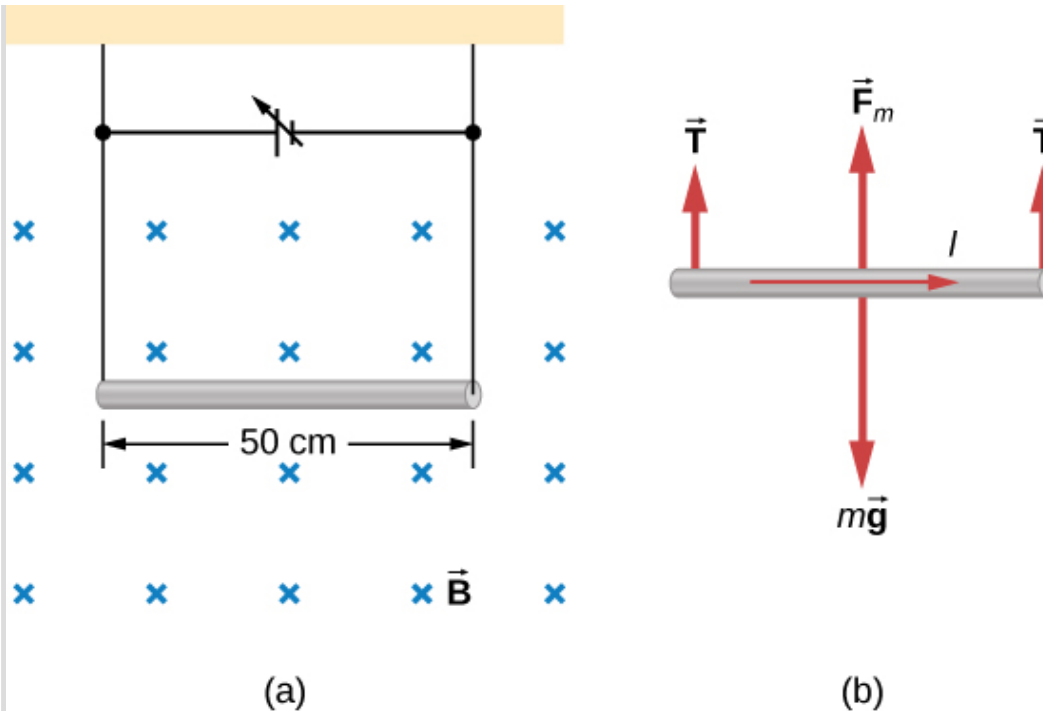
$$\vec{F} = I\vec{l} \times \vec{B}.$$

This is the force on a straight, current-carrying wire in a uniform magnetic field.

Example:

Balancing the Gravitational and Magnetic Forces on a Current-Carrying Wire

A wire of length 50 cm and mass 10 g is suspended in a horizontal plane by a pair of flexible leads ([\[link\]](#)). The wire is then subjected to a constant magnetic field of magnitude 0.50 T, which is directed as shown. What are the magnitude and direction of the current in the wire needed to remove the tension in the supporting leads?



(a) A wire suspended in a magnetic field. (b) The free-body diagram for the wire.

Strategy

From the free-body diagram in the figure, the tensions in the supporting leads go to zero when the gravitational and magnetic forces balance each other. Using the RHR-1, we find that the magnetic force points up. We can then determine the current I by equating the two forces.

Solution

Equate the two forces of weight and magnetic force on the wire:

Equation:

$$mg = IlB.$$

Thus,

Equation:

$$I = \frac{mg}{lB} = \frac{(0.010 \text{ kg})(9.8 \text{ m/s}^2)}{(0.50 \text{ m})(0.50 \text{ T})} = 0.39 \text{ A}.$$

Significance

This large magnetic field creates a significant force on a length of wire to counteract the weight of the wire.

Example:

Calculating Magnetic Force on a Current-Carrying Wire

A long, rigid wire lying along the y -axis carries a 5.0-A current flowing in the positive y -direction. (a) If a constant magnetic field of magnitude 0.30 T is directed along the positive x -axis, what is the magnetic force per unit length on the wire? (b) If a constant magnetic field of 0.30 T is directed 30 degrees from the $+x$ -axis towards the $+y$ -axis, what is the magnetic force per unit length on the wire?

Strategy

The magnetic force on a current-carrying wire in a magnetic field is given by

$\vec{F} = I\vec{l} \times \vec{B}$. For part a, since the current and magnetic field are perpendicular in this problem, we can simplify the formula to give us the magnitude and find the direction through the RHR-1. The angle θ is 90 degrees, which means $\sin\theta = 1$. Also, the length can be divided over to the left-hand side to find the force per unit length. For part b, the current times length is written in unit vector notation, as well as the magnetic field. After the cross product is taken, the directionality is evident by the resulting unit vector.

Solution

- a. We start with the general formula for the magnetic force on a wire. We are looking for the force per unit length, so we divide by the length to bring it to the left-hand side. We also set $\sin\theta = 1$. The solution therefore is

Equation:

$$\begin{aligned} F &= IlB \sin\theta \\ \frac{F}{l} &= (5.0 \text{ A})(0.30 \text{ T}) \\ \frac{F}{l} &= 1.5 \text{ N/m.} \end{aligned}$$

Directionality: Point your fingers in the positive y -direction and curl your fingers in the positive x -direction. Your thumb will point in the $-\vec{k}$ direction. Therefore, with directionality, the solution is

Equation:

$$\frac{\vec{\mathbf{F}}}{l} = -1.5\vec{\mathbf{k}} \text{ N/m.}$$

- b. The current times length and the magnetic field are written in unit vector notation. Then, we take the cross product to find the force:

Equation:

$$\begin{aligned}\vec{\mathbf{F}} &= I\vec{\mathbf{l}} \times \vec{\mathbf{B}} = (5.0A)l\hat{\mathbf{j}} \times \left(0.30T\cos(30^\circ)\hat{\mathbf{i}} + 0.30T\sin(30^\circ)\hat{\mathbf{j}}\right) \\ \vec{\mathbf{F}}/l &= -1.30\hat{\mathbf{k}} \text{ N/m.}\end{aligned}$$

Significance

This large magnetic field creates a significant force on a small length of wire. As the angle of the magnetic field becomes more closely aligned to the current in the wire, there is less of a force on it, as seen from comparing parts a and b.

Note:

Exercise:

Problem:

Check Your Understanding A straight, flexible length of copper wire is immersed in a magnetic field that is directed into the page. (a) If the wire's current runs in the $+x$ -direction, which way will the wire bend? (b) Which way will the wire bend if the current runs in the $-x$ -direction?

Solution:

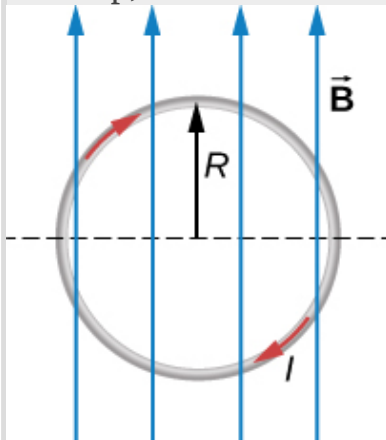
a. bends upward; b. bends downward

Example:

Force on a Circular Wire

A circular current loop of radius R carrying a current I is placed in the xy -plane. A constant uniform magnetic field cuts through the loop parallel to the y -axis

([link](#)). Find the magnetic force on the upper half of the loop, the lower half of the loop, and the total force on the loop.



A loop of wire carrying a current in a magnetic field.

Strategy

The magnetic force on the upper loop should be written in terms of the differential force acting on each segment of the loop. If we integrate over each differential piece, we solve for the overall force on that section of the loop. The force on the lower loop is found in a similar manner, and the total force is the addition of these two forces.

Solution

A differential force on an arbitrary piece of wire located on the upper ring is:

Equation:

$$dF = IB \sin \theta dl.$$

where θ is the angle between the magnetic field direction (+y) and the segment of wire. A differential segment is located at the same radius, so using an arc-length formula, we have:

Equation:

$$\begin{aligned} dl &= R d\theta \\ dF &= IB R \sin \theta d\theta. \end{aligned}$$

In order to find the force on a segment, we integrate over the upper half of the circle, from 0 to π . This results in:

Equation:

$$F = IBR \int_0^{\pi} \sin \theta d\theta = IBR(-\cos \pi + \cos 0) = 2IBR.$$

The lower half of the loop is integrated from π to zero, giving us:

Equation:

$$F = IBR \int_{\pi}^0 \sin \theta d\theta = IBR(-\cos 0 + \cos \pi) = -2IBR.$$

The net force is the sum of these forces, which is zero.

Significance

The total force on any closed loop in a uniform magnetic field is zero. Even though each piece of the loop has a force acting on it, the net force on the system is zero. (Note that there is a net torque on the loop, which we consider in the next section.)

Summary

- An electrical current produces a magnetic field around the wire.
- The directionality of the magnetic field produced is determined by the right hand rule-2, where your thumb points in the direction of the current and your fingers wrap around the wire in the direction of the magnetic field.
- The magnetic force on current-carrying conductors is given by $\vec{F} = I\vec{l} \times \vec{B}$ where I is the current and l is the length of a wire in a uniform magnetic field B .

Conceptual Questions

Exercise:

Problem:

Describe the error that results from accidentally using your left rather than your right hand when determining the direction of a magnetic force.

Exercise:**Problem:**

Considering the magnetic force law, are the velocity and magnetic field always perpendicular? Are the force and velocity always perpendicular? What about the force and magnetic field?

Solution:

Velocity and magnetic field can be set together in any direction. If there is a force, the velocity is perpendicular to it. The magnetic field is also perpendicular to the force if it exists.

Exercise:**Problem:**

Why can a nearby magnet distort a cathode ray tube television picture?

Exercise:**Problem:**

A magnetic field exerts a force on the moving electrons in a current carrying wire. What exerts the force on a wire?

Solution:

A force on a wire is exerted by an external magnetic field created by a wire or another magnet.

Exercise:**Problem:**

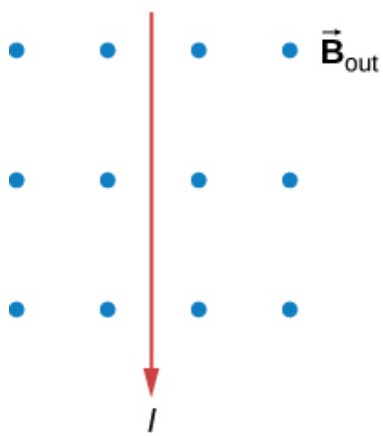
There are regions where the magnetic field of earth is almost perpendicular to the surface of Earth. What difficulty does this cause in the use of a compass?

Problems

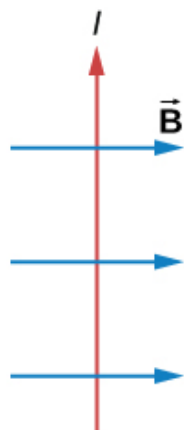
Exercise:

Problem:

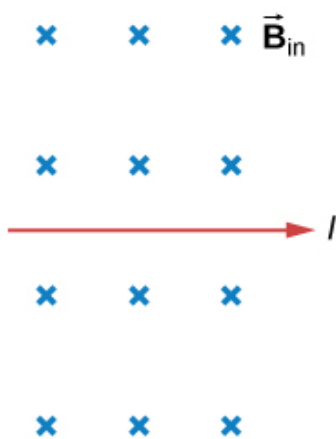
What is the direction of the magnetic force on the current in each of the six cases?



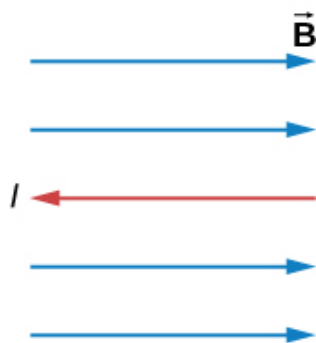
(a)



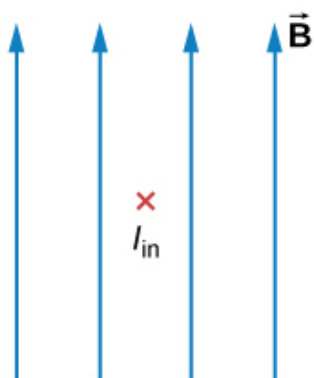
(b)



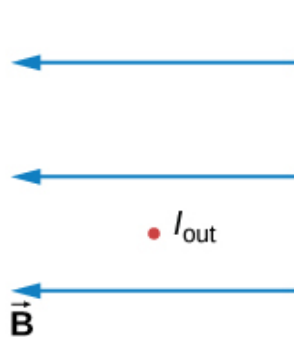
(c)



(d)



(e)



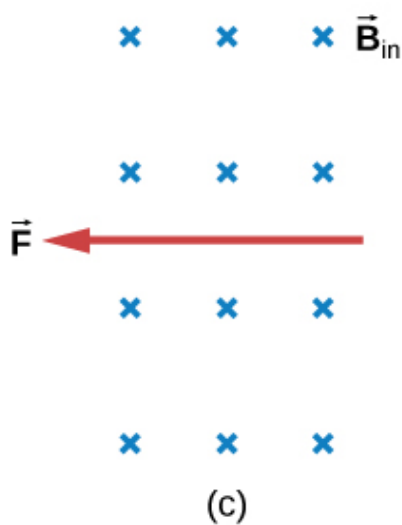
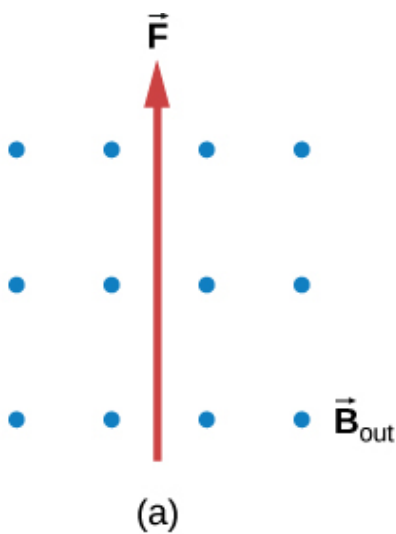
(f)

Solution:

a. left; b. into the page; c. up; d. no force; e. right; f. down

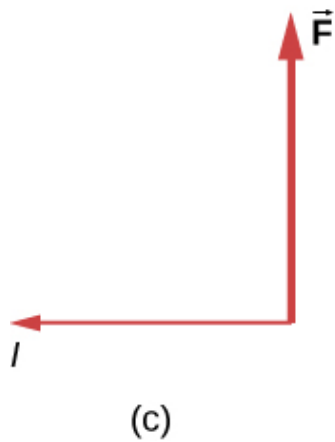
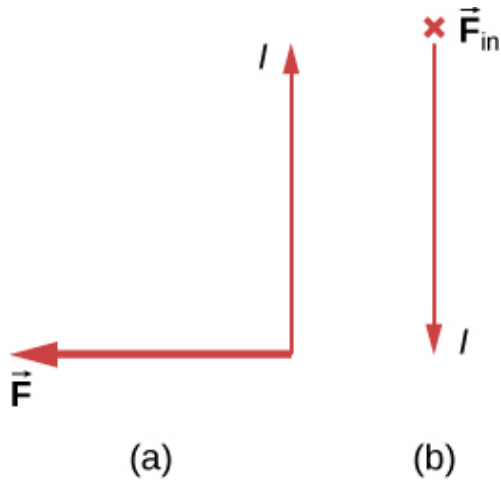
Exercise:**Problem:**

What is the direction of a current that experiences the magnetic force shown in each of the three cases, assuming the current runs perpendicular to \vec{B} ?

**Exercise:**

Problem:

What is the direction of the magnetic field that produces the magnetic force shown on the currents in each of the three cases, assuming $\vec{\mathbf{B}}$ is perpendicular to I ?



Solution:

a. into the page; b. left; c. out of the page

Exercise:

Problem:

(a) What is the force per meter on a lightning bolt at the equator that carries 20,000 A perpendicular to Earth's $3.0 \times 10^{-5} \text{ T}$ field? (b) What is the direction of the force if the current is straight up and Earth's field direction is due north, parallel to the ground?

Exercise:**Problem:**

(a) A dc power line for a light-rail system carries 1000 A at an angle of 30.0° to Earth's $5.0 \times 10^{-5} \text{ T}$ field. What is the force on a 100-m section of this line? (b) Discuss practical concerns this presents, if any.

Solution:

a. 2.50 N; b. This means that the light-rail power lines must be attached in order not to be moved by the force caused by Earth's magnetic field.

Exercise:**Problem:**

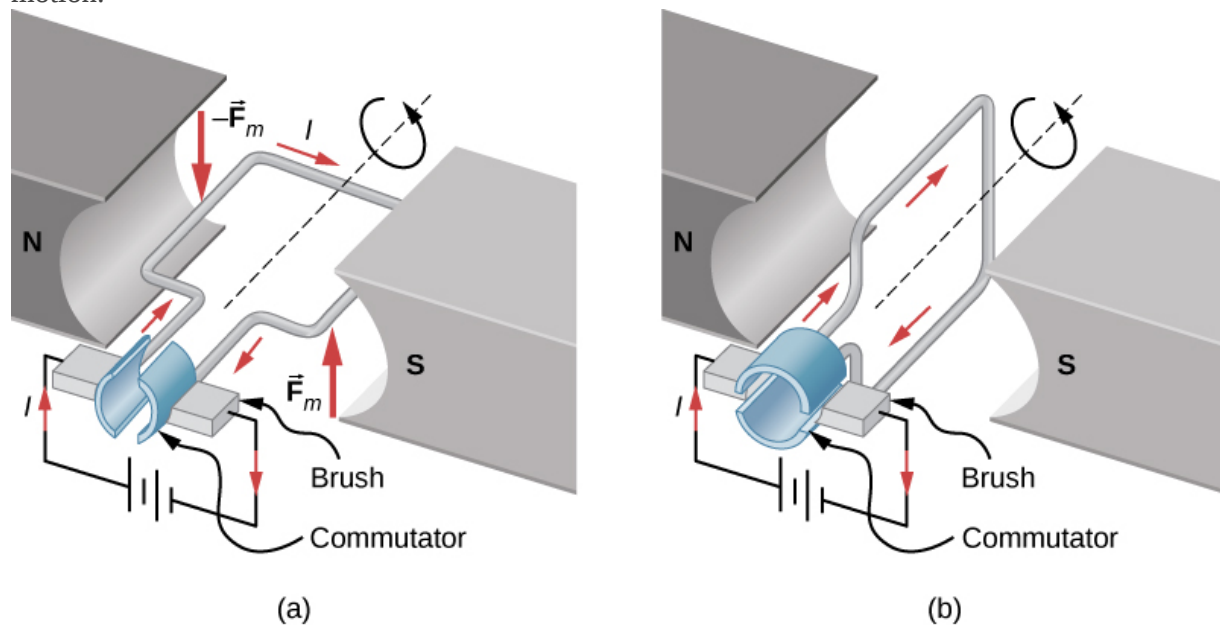
A wire carrying a 30.0-A current passes between the poles of a strong magnet that is perpendicular to its field and experiences a 2.16-N force on the 4.00 cm of wire in the field. What is the average field strength?

Force and Torque on a Current Loop

By the end of this section, you will be able to:

- Evaluate the net force on a current loop in an external magnetic field
- Evaluate the net torque on a current loop in an external magnetic field
- Define the magnetic dipole moment of a current loop

Motors are the most common application of magnetic force on current-carrying wires. Motors contain loops of wire in a magnetic field. When current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft. Electrical energy is converted into mechanical work in the process. Once the loop's surface area is aligned with the magnetic field, the direction of current is reversed, so there is a continual torque on the loop ([link](#)). This reversal of the current is done with commutators and brushes. The commutator is set to reverse the current flow at set points to keep continual motion in the motor. A basic commutator has three contact areas to avoid dead spots where the loop would have zero instantaneous torque at that point. The brushes press against the commutator, creating electrical contact between parts of the commutator during the spinning motion.



A simplified version of a dc electric motor. (a) The rectangular wire loop is placed in a magnetic field. The forces on the wires closest to the magnetic poles (N and S) are opposite in direction as determined by the right-hand rule-1. Therefore, the loop has a net torque and rotates to the position shown in (b). (b) The brushes now touch the commutator segments so that no current flows through the loop. No torque acts on the loop, but the loop continues to spin from the initial velocity given to it in part (a). By the time the loop flips over, current flows through the wires again but now in the opposite direction, and the process repeats as in part (a). This causes continual rotation of the loop.

In a uniform magnetic field, a current-carrying loop of wire, such as a loop in a motor, experiences both forces and torques on the loop. [link](#) shows a rectangular loop of wire that carries a current I

and has sides of lengths a and b . The loop is in a uniform magnetic field: $\vec{B} = B\hat{j}$. The magnetic force on a straight current-carrying wire of length l is given by $\vec{l} \times \vec{B}$. To find the net force on the loop, we have to apply this equation to each of the four sides. The force on side 1 is

Equation:

$$\vec{F}_1 = IaB\sin(90^\circ - \theta)\hat{i} = IaB\cos\theta\hat{i}$$

where the direction has been determined with the RHR-1. The current in side 3 flows in the opposite direction to that of side 1, so

Equation:

$$\vec{F}_3 = -IaB\sin(90^\circ + \theta)\hat{i} = -IaB\cos\theta\hat{i}.$$

The currents in sides 2 and 4 are perpendicular to \vec{B} and the forces on these sides are

Equation:

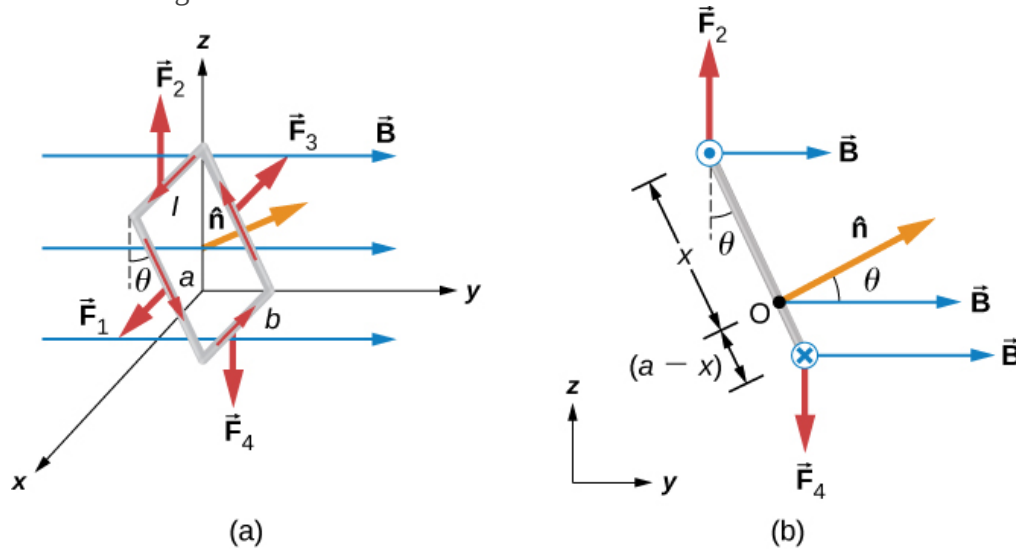
$$\vec{F}_2 = IbB\hat{k}, \quad \vec{F}_4 = -IbB\hat{k}.$$

We can now find the net force on the loop:

Equation:

$$\sum \vec{F}_{\text{net}} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \vec{F}_4 = 0.$$

Although this result ($\sum F = 0$) has been obtained for a rectangular loop, it is far more general and holds for current-carrying loops of arbitrary shapes; that is, there is no net force on a current loop in a uniform magnetic field.



(a) A rectangular current loop in a uniform magnetic field is subjected to a net torque but not a net force. (b) A side view of the coil.

To find the net torque on the current loop shown in [\[link\]](#), we first consider F_1 and F_3 . Since they have the same line of action and are equal and opposite, the sum of their torques about any axis is zero (see [Fixed-Axis Rotation](#)). Thus, if there is any torque on the loop, it must be furnished by F_2 and F_4 . Let's calculate the torques around the axis that passes through point O of [\[link\]](#) (a side view of the coil) and is perpendicular to the plane of the page. The point O is a distance x from side 2 and a distance $(a - x)$ from side 4 of the loop. The moment arms of F_2 and F_4 are $x \sin \theta$ and $(a - x) \sin \theta$, respectively, so the net torque on the loop is

Equation:

$$\begin{aligned}\sum \vec{\tau} &= \vec{\tau}_1 + \vec{\tau}_2 + \vec{\tau}_3 + \vec{\tau}_4 = F_2 x \sin \theta \hat{\mathbf{i}} - F_4 (a - x) \sin(\theta) \hat{\mathbf{i}} \\ &= -IbBx \sin \theta \hat{\mathbf{i}} - IbB(a - x) \sin \theta \hat{\mathbf{i}}.\end{aligned}$$

This simplifies to

Equation:

$$\vec{\tau} = -IAB \sin \theta \hat{\mathbf{i}}$$

where $A = ab$ is the area of the loop.

Notice that this torque is independent of x ; it is therefore independent of where point O is located in the plane of the current loop. Consequently, the loop experiences the same torque from the magnetic field about any axis in the plane of the loop and parallel to the x -axis.

A closed-current loop is commonly referred to as a **magnetic dipole** and the term IA is known as its **magnetic dipole moment** μ . Actually, the magnetic dipole moment is a vector that is defined as

Equation:

$$\vec{\mu} = IA \hat{\mathbf{n}}$$

where $\hat{\mathbf{n}}$ is a unit vector directed perpendicular to the plane of the loop (see [\[link\]](#)). The direction of $\hat{\mathbf{n}}$ is obtained with the RHR-2—if you curl the fingers of your right hand in the direction of current flow in the loop, then your thumb points along $\hat{\mathbf{n}}$. If the loop contains N turns of wire, then its magnetic dipole moment is given by

Note:

Equation:

$$\vec{\mu} = NIA \hat{\mathbf{n}}.$$

In terms of the magnetic dipole moment, the torque on a current loop due to a uniform magnetic field can be written simply as

Note:

Equation:

$$\vec{\tau} = \vec{\mu} \times \vec{B}.$$

This equation holds for a current loop in a two-dimensional plane of arbitrary shape.

Using a calculation analogous to that found in [Capacitance](#) for an electric dipole, the potential energy of a magnetic dipole is

Note:

Equation:

$$U = -\vec{\mu} \cdot \vec{B}.$$

Example:

Forces and Torques on Current-Carrying Loops

A circular current loop of radius 2.0 cm carries a current of 2.0 mA. (a) What is the magnitude of its magnetic dipole moment? (b) If the dipole is oriented at 30 degrees to a uniform magnetic field of magnitude 0.50 T, what is the magnitude of the torque it experiences and what is its potential energy?

Strategy

The dipole moment is defined by the current times the area of the loop. The area of the loop can be calculated from the area of the circle. The torque on the loop and potential energy are calculated from identifying the magnetic moment, magnetic field, and angle oriented in the field.

Solution

- a. The magnetic moment μ is calculated by the current times the area of the loop or πr^2 .

Equation:

$$\mu = IA = (2.0 \times 10^{-3} \text{ A})(\pi(0.02 \text{ m})^2) = 2.5 \times 10^{-6} \text{ A} \cdot \text{m}^2$$

- b. The torque and potential energy are calculated by identifying the magnetic moment, magnetic field, and the angle between these two vectors. The calculations of these quantities are:

Equation:

$$\begin{aligned}\tau &= \vec{\mu} \times \vec{B} = \mu B \sin \theta = (2.5 \times 10^{-6} \text{ A} \cdot \text{m}^2) (0.50 \text{ T}) \sin(30^\circ) = 6.3 \times 10^{-7} \text{ N} \cdot \text{m} \\ U &= -\vec{\mu} \cdot \vec{B} = -\mu B \cos \theta = -(2.5 \times 10^{-6} \text{ A} \cdot \text{m}^2) (0.50 \text{ T}) \cos(30^\circ) = -1.1 \times 10^{-6} \text{ J}.\end{aligned}$$

Significance

The concept of magnetic moment at the atomic level is discussed in the next chapter. The concept of aligning the magnetic moment with the magnetic field is the functionality of devices like magnetic motors, whereby switching the external magnetic field results in a constant spinning of the loop as it tries to align with the field to minimize its potential energy.

Note:

Exercise:

Problem: Check Your Understanding

In what orientation would a magnetic dipole have to be to produce (a) a maximum torque in a magnetic field? (b) A maximum energy of the dipole?

Solution:

a. aligned or anti-aligned; b. perpendicular

Summary

- The net force on a current-carrying loop of any plane shape in a uniform magnetic field is zero.
- The net torque τ on a current-carrying loop of any shape in a uniform magnetic field is calculated using $\tau = \vec{\mu} \times \vec{B}$ where $\vec{\mu}$ is the magnetic dipole moment and \vec{B} is the magnetic field strength.
- The magnetic dipole moment μ is the product of the number of turns of wire N , the current in the loop I , and the area of the loop A or $\vec{\mu} = NIA\hat{n}$.

Problems

Exercise:

Problem:

(a) By how many percent is the torque of a motor decreased if its permanent magnets lose 5.0% of their strength? (b) How many percent would the current need to be increased to return the torque to original values?

Solution:

a. $\tau = NIAB$, so τ decreases by 5.00% if B decreases by 5.00%; b. 5.26% increase

Exercise:**Problem:**

(a) What is the maximum torque on a 150-turn square loop of wire 18.0 cm on a side that carries a 50.0-A current in a 1.60-T field? (b) What is the torque when θ is 10.9° ?

Exercise:**Problem:**

Find the current through a loop needed to create a maximum torque of $9.0 \text{ N} \cdot \text{m}$. The loop has 50 square turns that are 15.0 cm on a side and is in a uniform 0.800-T magnetic field.

Solution:

10.0 A

Exercise:**Problem:**

Calculate the magnetic field strength needed on a 200-turn square loop 20.0 cm on a side to create a maximum torque of $300 \text{ N} \cdot \text{m}$ if the loop is carrying 25.0 A.

Exercise:**Problem:**

Since the equation for torque on a current-carrying loop is $\tau = NIAB \sin \theta$, the units of $\text{N} \cdot \text{m}$ must equal units of $\text{A} \cdot \text{m}^2 \cdot \text{T}$. Verify this.

Solution:

$$\text{A} \cdot \text{m}^2 \cdot \text{T} = \text{A} \cdot \text{m}^2 \cdot \frac{\text{N}}{\text{A} \cdot \text{m}} = \text{N} \cdot \text{m}$$

Exercise:**Problem:**

(a) At what angle θ is the torque on a current loop 90.0% of maximum? (b) 50.0% of maximum? (c) 10.0% of maximum?

Exercise:**Problem:**

A proton has a magnetic field due to its spin. The field is similar to that created by a circular current loop $0.65 \times 10^{-15} \text{ m}$ in radius with a current of $1.05 \times 10^4 \text{ A}$. Find the maximum torque on a proton in a 2.50-T field. (This is a significant torque on a small particle.)

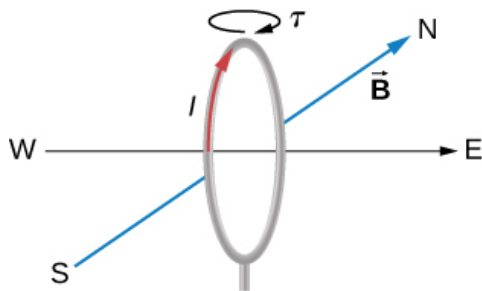
Solution:

$$3.48 \times 10^{-26} \text{ N} \cdot \text{m}$$

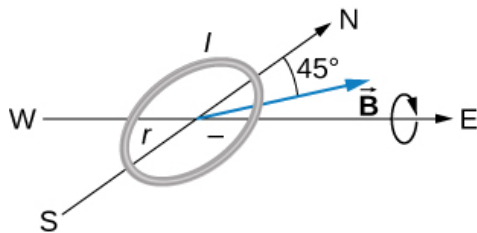
Exercise:

Problem:

(a) A 200-turn circular loop of radius 50.0 cm is vertical, with its axis on an east-west line. A current of 100 A circulates clockwise in the loop when viewed from the east. Earth's field here is due north, parallel to the ground, with a strength of $3.0 \times 10^{-5} \text{ T}$. What are the direction and magnitude of the torque on the loop? (b) Does this device have any practical applications as a motor?

**Exercise:****Problem:**

Repeat the previous problem, but with the loop lying flat on the ground with its current circulating counterclockwise (when viewed from above) in a location where Earth's field is north, but at an angle 45.0° below the horizontal and with a strength of $6.0 \times 10^{-5} \text{ T}$.

**Solution:**

$$0.666 \text{ N} \cdot \text{m}$$

Glossary

magnetic dipole
closed-current loop

magnetic dipole moment
term IA of the magnetic dipole, also called μ

motor (dc)
loop of wire in a magnetic field; when current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft; electrical energy is converted into mechanical

work in the process

The Hall Effect

By the end of this section, you will be able to:

- Explain a scenario where the magnetic and electric fields are crossed and their forces balance each other as a charged particle moves through a velocity selector
- Compare how charge carriers move in a conductive material and explain how this relates to the Hall effect

In 1879, E.H. Hall devised an experiment that can be used to identify the sign of the predominant charge carriers in a conducting material. From a historical perspective, this experiment was the first to demonstrate that the charge carriers in most metals are negative.

Note:

Visit this [website](#) to find more information about the Hall effect.

We investigate the **Hall effect** by studying the motion of the free electrons along a metallic strip of width l in a constant magnetic field ([\[link\]](#)). The electrons are moving from left to right, so the magnetic force they experience pushes them to the bottom edge of the strip. This leaves an excess of positive charge at the top edge of the strip, resulting in an electric field E directed from top to bottom. The charge concentration at both edges builds up until the electric force on the electrons in one direction is balanced by the magnetic force on them in the opposite direction. Equilibrium is reached when:

Equation:

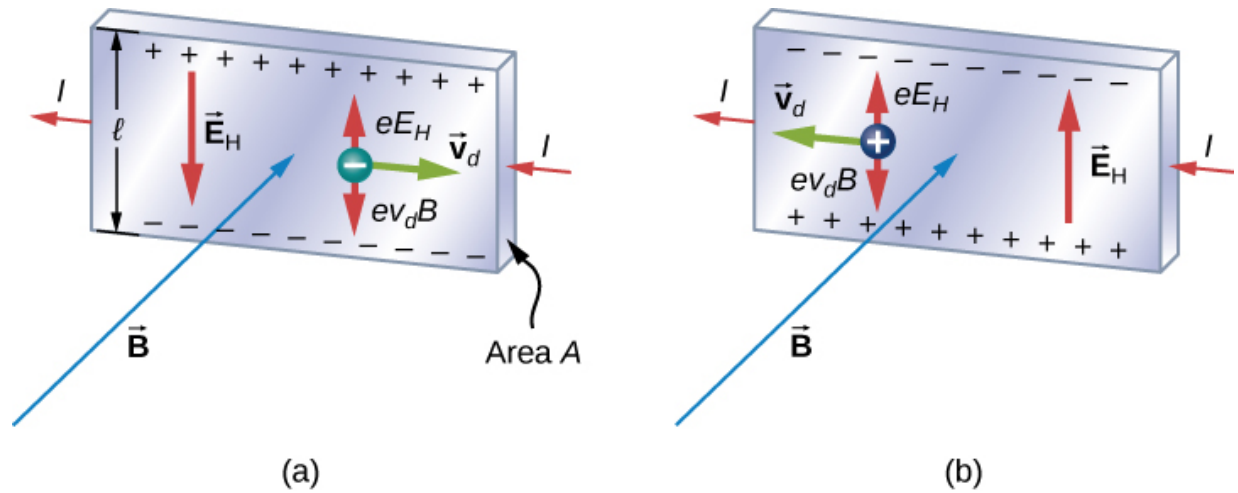
$$eE = ev_d B$$

where e is the magnitude of the electron charge, v_d is the drift speed of the electrons, and E is the magnitude of the electric field created by the separated charge. Solving this for the drift speed results in

Note:

Equation:

$$v_d = \frac{E}{B}.$$



In the Hall effect, a potential difference between the top and bottom edges of the metal strip is produced when moving charge carriers are deflected by the magnetic field. (a) Hall effect for negative charge carriers; (b) Hall effect for positive charge carriers.

A scenario where the electric and magnetic fields are perpendicular to one another is called a crossed-field situation. If these fields produce equal and opposite forces on a charged particle with the velocity that equates the forces, these particles are able to pass through an apparatus, called a **velocity selector**, undeflected. This velocity is represented in [\[link\]](#). Any other velocity of a charged particle sent into the same fields would be deflected by the magnetic force or electric force.

Going back to the Hall effect, if the current in the strip is I , then from [Current and Resistance](#), we know that

Equation:

$$I = nev_d A$$

where n is the number of charge carriers per volume and A is the cross-sectional area of the strip. Combining the equations for v_d and I results in

Equation:

$$I = ne \left(\frac{E}{B} \right) A.$$

The field E is related to the potential difference V between the edges of the strip by

Equation:

$$E = \frac{V}{l}.$$

The quantity V is called the Hall potential and can be measured with a voltmeter. Finally, combining the equations for I and E gives us

Note:
Equation:

$$V = \frac{IBl}{neA}$$

where the upper edge of the strip in [\[link\]](#) is positive with respect to the lower edge.

We can also combine [\[link\]](#) and [\[link\]](#) to get an expression for the Hall voltage in terms of the magnetic field:

Note:
Equation:

$$V = Blv_d.$$

What if the charge carriers are positive, as in [\[link\]](#)? For the same current I , the magnitude of V is still given by [\[link\]](#). However, the upper edge is now negative with respect to the lower edge. Therefore, by simply measuring the sign of V , we can determine the sign of the majority charge carriers in a metal.

Hall potential measurements show that electrons are the dominant charge carriers in most metals. However, Hall potentials indicate that for a few metals, such as tungsten, beryllium, and many semiconductors, the majority of charge carriers are positive. It turns out that conduction by positive charge is caused by the migration of missing electron sites (called holes) on ions. Conduction by holes is studied later in [Condensed Matter Physics](#).

The Hall effect can be used to measure magnetic fields. If a material with a known density of charge carriers n is placed in a magnetic field and V is measured, then the field can be determined from [\[link\]](#). In research laboratories where the fields of electromagnets used

for precise measurements have to be extremely steady, a “Hall probe” is commonly used as part of an electronic circuit that regulates the field.

Example:**Velocity Selector**

An electron beam enters a crossed-field velocity selector with magnetic and electric fields of 2.0 mT and $6.0 \times 10^3 \text{ N/C}$, respectively. (a) What must the velocity of the electron beam be to traverse the crossed fields undeflected? If the electric field is turned off, (b) what is the acceleration of the electron beam and (c) what is the radius of the circular motion that results?

Strategy

The electron beam is not deflected by either of the magnetic or electric fields if these forces are balanced. Based on these balanced forces, we calculate the velocity of the beam. Without the electric field, only the magnetic force is used in Newton’s second law to find the acceleration. Lastly, the radius of the path is based on the resulting circular motion from the magnetic force.

Solution

- a. The velocity of the unperturbed beam of electrons with crossed fields is calculated by [\[link\]](#):

Equation:

$$v_d = \frac{E}{B} = \frac{6 \times 10^3 \text{ N/C}}{2 \times 10^{-3} \text{ T}} = 3 \times 10^6 \text{ m/s}.$$

- b. The acceleration is calculated from the net force from the magnetic field, equal to mass times acceleration. The magnitude of the acceleration is:

Equation:

$$\begin{aligned} ma &= qvB \\ a &= \frac{qvB}{m} = \frac{(1.6 \times 10^{-19} \text{ C})(3 \times 10^6 \text{ m/s})(2 \times 10^{-3} \text{ T})}{9.1 \times 10^{-31} \text{ kg}} = 1.1 \times 10^{15} \text{ m/s}^2. \end{aligned}$$

- c. The radius of the path comes from a balance of the circular and magnetic forces, or [\[link\]](#):

Equation:

$$r = \frac{mv}{qB} = \frac{(9.1 \times 10^{-31} \text{ kg})(3 \times 10^6 \text{ m/s})}{(1.6 \times 10^{-19} \text{ C})(2 \times 10^{-3} \text{ T})} = 8.5 \times 10^{-3} \text{ m}.$$

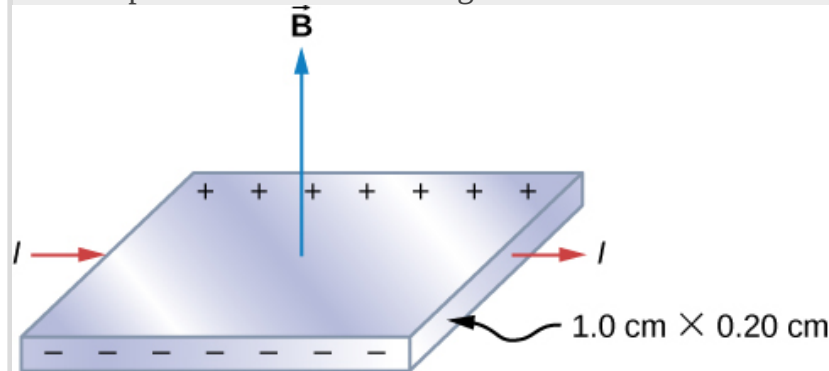
Significance

If electrons in the beam had velocities above or below the answer in part (a), those electrons would have a stronger net force exerted by either the magnetic or electric field. Therefore, only those electrons at this specific velocity would make it through.

Example:

The Hall Potential in a Silver Ribbon

[\[link\]](#) shows a silver ribbon whose cross section is 1.0 cm by 0.20 cm. The ribbon carries a current of 100 A from left to right, and it lies in a uniform magnetic field of magnitude 1.5 T. Using a density value of $n = 5.9 \times 10^{28}$ electrons per cubic meter for silver, find the Hall potential between the edges of the ribbon.



Finding the Hall potential in a silver ribbon in a magnetic field is shown.

Strategy

Since the majority of charge carriers are electrons, the polarity of the Hall voltage is that indicated in the figure. The value of the Hall voltage is calculated using [\[link\]](#):

Equation:

$$V = \frac{IBl}{neA}.$$

Solution

When calculating the Hall voltage, we need to know the current through the material, the magnetic field, the length, the number of charge carriers, and the area. Since all of these are given, the Hall voltage is calculated as:

Equation:

$$V = \frac{IBl}{neA} = \frac{(100 \text{ A})(1.5 \text{ T})}{5.9 \times 10^{28}/\text{m}^3} \frac{1.0 \times 10^{-2} \text{ m}}{1.6 \times 10^{-19} \text{ C}} \frac{1}{2.0 \times 10^{-5} \text{ m}^2} = 7.9 \times 10^{-6} \text{ V}.$$

Significance

As in this example, the Hall potential is generally very small, and careful experimentation with sensitive equipment is required for its measurement.

Note:

Exercise:

Problem:

Check Your Understanding A Hall probe consists of a copper strip, $n = 8.5 \times 10^{28}$ electrons per cubic meter, which is 2.0 cm wide and 0.10 cm thick. What is the magnetic field when $I = 50$ A and the Hall potential is (a) $4.0\mu\text{V}$ and (b) $6.0\mu\text{V}$?

Solution:

a. 1.1 T; b. 1.6 T

Summary

- Perpendicular electric and magnetic fields exert equal and opposite forces for a specific velocity of entering particles, thereby acting as a velocity selector. The velocity that passes through undeflected is calculated by $v = \frac{E}{B}$.
- The Hall effect can be used to measure the sign of the majority of charge carriers for metals. It can also be used to measure a magnetic field.

Conceptual Questions

Exercise:

Problem:

Hall potentials are much larger for poor conductors than for good conductors. Why?

Solution:

Poor conductors have a lower charge carrier density, n , which, based on the Hall effect formula, relates to a higher Hall potential. Good conductors have a higher charge carrier density, thereby a lower Hall potential.

Problems

Exercise:

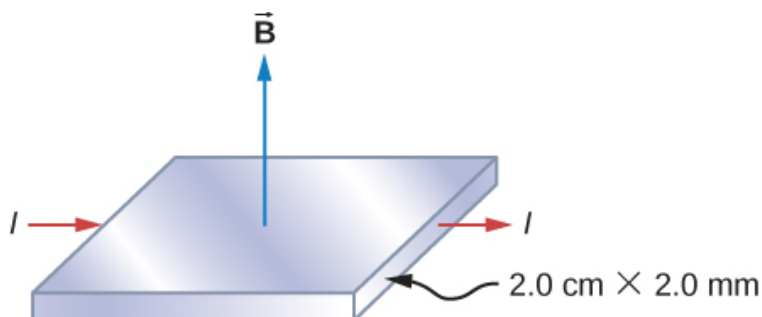
Problem:

A strip of copper is placed in a uniform magnetic field of magnitude 2.5 T. The Hall electric field is measured to be $1.5 \times 10^{-3} \text{ V/m}$. (a) What is the drift speed of the conduction electrons? (b) Assuming that $n = 8.0 \times 10^{28}$ electrons per cubic meter and that the cross-sectional area of the strip is $5.0 \times 10^{-6} \text{ m}^2$, calculate the current in the strip. (c) What is the Hall coefficient $1/nq$?

Exercise:

Problem:

The cross-sectional dimensions of the copper strip shown are 2.0 cm by 2.0 mm. The strip carries a current of 100 A, and it is placed in a magnetic field of magnitude $B = 1.5 \text{ T}$. What are the value and polarity of the Hall potential in the copper strip?



Solution:

$$5.8 \times 10^{-6} \text{ V}$$

Exercise:

Problem:

The magnitudes of the electric and magnetic fields in a velocity selector are $1.8 \times 10^5 \text{ V/m}$ and 0.080 T, respectively. (a) What speed must a proton have to pass through the selector? (b) Also calculate the speeds required for an alpha-particle and a singly ionized ${}_8\text{O}^{16}$ atom to pass through the selector.

Exercise:

Problem:

A charged particle moves through a velocity selector at constant velocity. In the selector, $E = 1.0 \times 10^4 \text{ N/C}$ and $B = 0.250 \text{ T}$. When the electric field is turned off, the charged particle travels in a circular path of radius 3.33 mm. Determine the charge-to-mass ratio of the particle.

Solution:

$$4.8 \times 10^7 \text{ C/kg}$$

Exercise:**Problem:**

A Hall probe gives a reading of $1.5 \mu\text{V}$ for a current of 2 A when it is placed in a magnetic field of 1 T. What is the magnetic field in a region where the reading is $2 \mu\text{V}$ for 1.7 A of current?

Glossary**Hall effect**

creation of voltage across a current-carrying conductor by a magnetic field

velocity selector

apparatus where the crossed electric and magnetic fields produce equal and opposite forces on a charged particle moving with a specific velocity; this particle moves through the velocity selector not affected by either field while particles moving with different velocities are deflected by the apparatus

Applications of Magnetic Forces and Fields

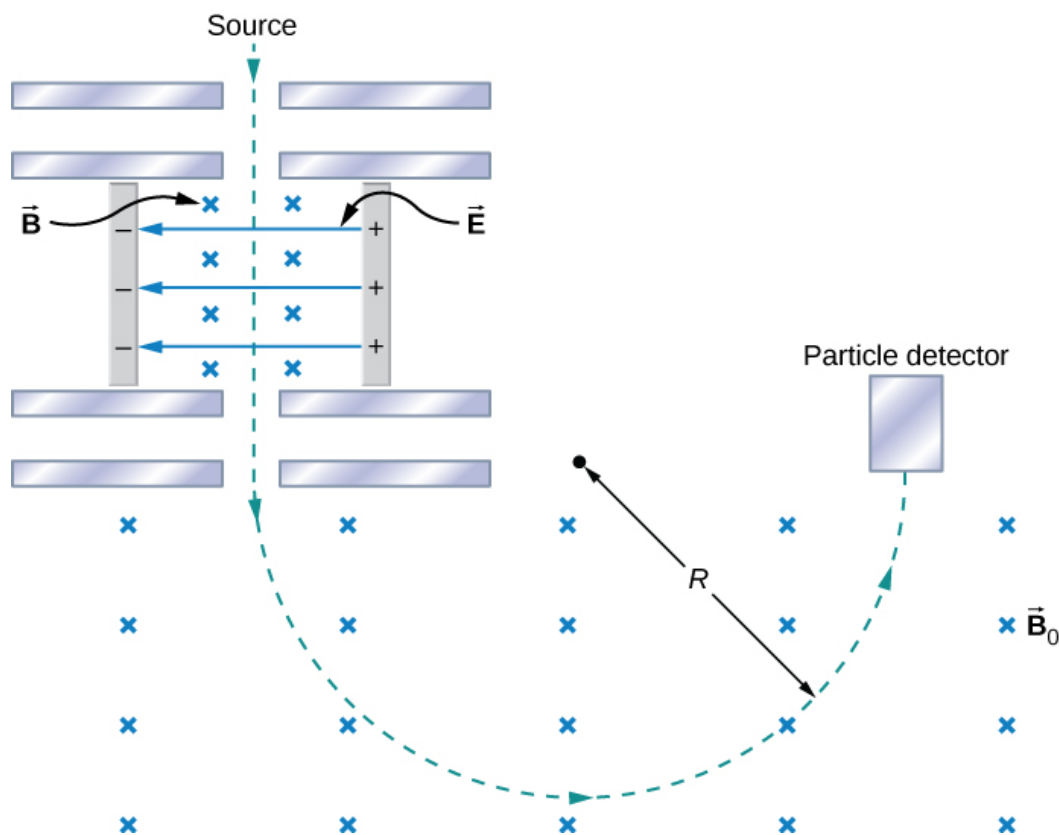
By the end of this section, you will be able to:

- Explain how a mass spectrometer works to separate charges
- Explain how a cyclotron works

Being able to manipulate and sort charged particles allows deeper experimentation to understand what matter is made of. We first look at a mass spectrometer to see how we can separate ions by their charge-to-mass ratio. Then we discuss cyclotrons as a method to accelerate charges to very high energies.

Mass Spectrometer

The **mass spectrometer** is a device that separates ions according to their charge-to-mass ratios. One particular version, the Bainbridge mass spectrometer, is illustrated in [\[link\]](#). Ions produced at a source are first sent through a velocity selector, where the magnetic force is equally balanced with the electric force. These ions all emerge with the same speed $v = E/B$ since any ion with a different velocity is deflected preferentially by either the electric or magnetic force, and ultimately blocked from the next stage. They then enter a uniform magnetic field B_0 where they travel in a circular path whose radius R is given by [\[link\]](#). The radius is measured by a particle detector located as shown in the figure.



A schematic of the Bainbridge mass spectrometer, showing charged particles leaving a source, followed by a velocity selector where the electric and magnetic forces are balanced, followed by a region of uniform magnetic field where the particle is ultimately detected.

The relationship between the charge-to-mass ratio q/m and the radius R is determined by combining [\[link\]](#) and [\[link\]](#):

Note:

Equation:

$$\frac{q}{m} = \frac{E}{BB_0R}.$$

Since most ions are singly charged ($q = 1.6 \times 10^{-19} \text{ C}$), measured values of R can be used with this equation to determine the mass of ions. With modern instruments, masses can be determined to one part in 10^8 .

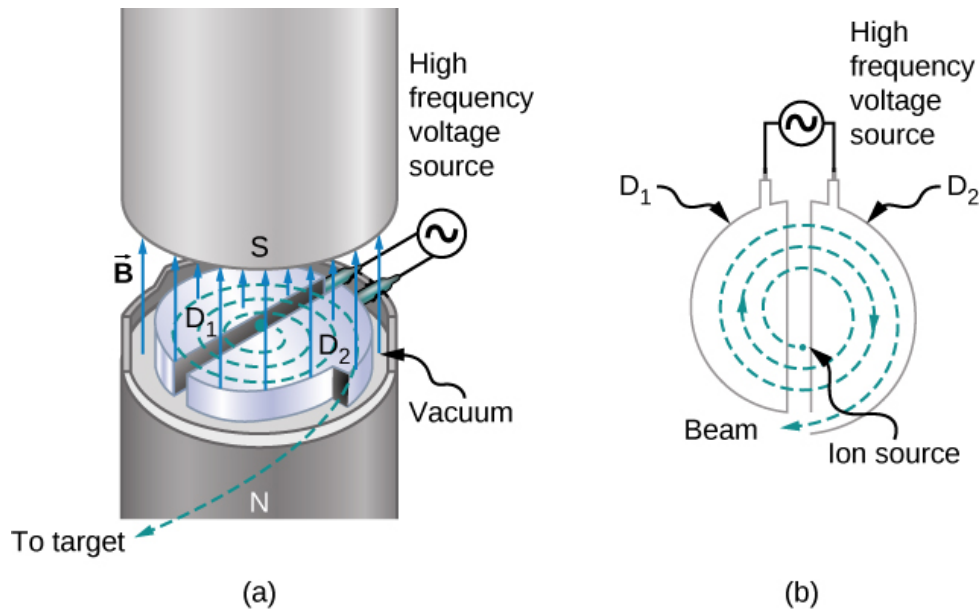
An interesting use of a spectrometer is as part of a system for detecting very small leaks in a research apparatus. In low-temperature physics laboratories, a device known as a dilution refrigerator uses a mixture of He-3, He-4, and other cryogenics to reach temperatures well below 1 K. The performance of the refrigerator is severely hampered if even a minute leak between its various components occurs. Consequently, before it is cooled down to the desired temperature, the refrigerator is subjected to a leak test. A small quantity of gaseous helium is injected into one of its compartments, while an adjacent, but supposedly isolated, compartment is connected to a high-vacuum pump to which a mass spectrometer is attached. A heated filament ionizes any helium atoms evacuated by the pump. The detection of these ions by the spectrometer then indicates a leak between the two compartments of the dilution refrigerator.

In conjunction with gas chromatography, mass spectrometers are used widely to identify unknown substances. While the gas chromatography portion breaks down the substance, the mass spectrometer separates the resulting ionized molecules. This technique is used with fire debris to ascertain the cause, in law enforcement to identify illegal drugs, in security to identify explosives, and in many medicinal applications.

Cyclotron

The **cyclotron** was developed by E.O. Lawrence to accelerate charged particles (usually protons, deuterons, or alpha-particles) to large kinetic energies. These particles are then used for nuclear-collision experiments to produce radioactive isotopes. A cyclotron is illustrated in [\[link\]](#). The particles move between two flat, semi-cylindrical metallic containers D1 and D2, called **dees**. The dees are enclosed in a larger metal container, and the apparatus is placed between the poles of an

electromagnet that provides a uniform magnetic field. Air is removed from the large container so that the particles neither lose energy nor are deflected because of collisions with air molecules. The dees are connected to a high-frequency voltage source that provides an alternating electric field in the small region between them. Because the dees are made of metal, their interiors are shielded from the electric field.



The inside of a cyclotron. A uniform magnetic field is applied as circulating protons travel through the dees, gaining energy as they traverse through the gap between the dees.

Suppose a positively charged particle is injected into the gap between the dees when D2 is at a positive potential relative to D1. The particle is then accelerated across the gap and enters D1 after gaining kinetic energy qV , where V is the average potential difference the particle experiences between the dees. When the particle is inside D1, only the uniform magnetic field \vec{B} of the electromagnet acts on it, so the particle moves in a circle of radius

Equation:

$$r = \frac{mv}{qB}$$

with a period of

Equation:

$$T = \frac{2\pi m}{qB}.$$

The period of the alternating voltage course is set at T , so while the particle is inside D1, moving along its semicircular orbit in a time $T/2$, the polarity of the dees is reversed. When the particle reenters the gap, D1 is positive with respect to D2, and the particle is again accelerated across the gap, thereby gaining a kinetic energy qV . The particle then enters D2, circulates in a slightly larger circle, and emerges from D2 after spending a time $T/2$ in this dee. This process repeats until the orbit of the particle reaches the boundary of the dees. At that point, the particle (actually, a beam of particles) is extracted from the cyclotron and used for some experimental purpose.

The operation of the cyclotron depends on the fact that, in a uniform magnetic field, a particle's orbital period is independent of its radius and its kinetic energy. Consequently, the period of the alternating voltage source need only be set at the one value given by [\[link\]](#). With that setting, the electric field accelerates particles every time they are between the dees.

If the maximum orbital radius in the cyclotron is R , then from [\[link\]](#), the maximum speed of a circulating particle of mass m and charge q is

Note:
Equation:

$$v_{\max} = \frac{qBR}{m}.$$

Thus, its kinetic energy when ejected from the cyclotron is

Equation:

$$\frac{1}{2}mv_{\max}^2 = \frac{q^2B^2R^2}{2m}.$$

The maximum kinetic energy attainable with this type of cyclotron is approximately 30 MeV. Above this energy, relativistic effects become important, which causes the orbital period to increase with the radius. Up to energies of several hundred MeV, the relativistic effects can be compensated for by making the magnetic field gradually increase with the radius of the orbit. However, for higher energies, much more elaborate methods must be used to accelerate particles.

Particles are accelerated to very high energies with either linear accelerators or synchrotrons. The linear accelerator accelerates particles continuously with the electric field of an electromagnetic wave that travels down a long evacuated tube. The Stanford Linear Accelerator (SLAC) is about 3.3 km long and accelerates electrons and positrons (positively charged electrons) to energies of 50 GeV. The synchrotron is constructed so that its bending magnetic field increases with particle speed in such a way that the particles stay in an orbit of fixed radius. The world's highest-energy synchrotron is located at CERN, which is on the Swiss-French border near Geneva. CERN has been of recent interest with the verified discovery of the Higgs Boson (see [Particle Physics and Cosmology](#)). This synchrotron can accelerate beams of approximately 10^{13} protons to energies of about 10^3 GeV.

Example:**Accelerating Alpha-Particles in a Cyclotron**

A cyclotron used to accelerate alpha-particles ($m = 6.64 \times 10^{-27} \text{ kg}$, $q = 3.2 \times 10^{-19} \text{ C}$) has a radius of 0.50 m and a magnetic field of 1.8 T. (a) What is the period of revolution of the alpha-particles? (b) What is their maximum kinetic energy?

Strategy

- The period of revolution is approximately the distance traveled in a circle divided by the speed. Identifying that the magnetic force applied is the centripetal force, we can derive the period formula.
- The kinetic energy can be found from the maximum speed of the beam, corresponding to the maximum radius within the cyclotron.

Solution

- By identifying the mass, charge, and magnetic field in the problem, we can calculate the period:

Equation:

$$T = \frac{2\pi m}{qB} = \frac{2\pi (6.64 \times 10^{-27} \text{ kg})}{(3.2 \times 10^{-19} \text{ C})(1.8 \text{ T})} = 7.3 \times 10^{-8} \text{ s}.$$

- By identifying the charge, magnetic field, radius of path, and the mass, we can calculate the maximum kinetic energy:

Equation:

$$\frac{1}{2}mv_{\text{max}}^2 = \frac{q^2 B^2 R^2}{2m} = \frac{(3.2 \times 10^{-19} \text{ C})^2 (1.8 \text{ T})^2 (0.50 \text{ m})^2}{2(6.65 \times 10^{-27} \text{ kg})} = 6.2 \times 10^{-12} \text{ J} = 39 \text{ MeV}.$$

Note:**Exercise:****Problem:**

Check Your Understanding A cyclotron is to be designed to accelerate protons to kinetic energies of 20 MeV using a magnetic field of 2.0 T. What is the required radius of the cyclotron?

Solution:

0.32 m

Summary

- A mass spectrometer is a device that separates ions according to their charge-to-mass ratios by first sending them through a velocity selector, then a uniform magnetic field.
- Cyclotrons are used to accelerate charged particles to large kinetic energies through applied electric and magnetic fields.

Key Equations

Force on a charge in a magnetic field	$\vec{\mathbf{F}} = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}$
Magnitude of magnetic force	$F = qvB\sin\theta$
Radius of a particle's path in a magnetic field	$r = \frac{mv}{qB}$
Period of a particle's motion in a magnetic field	$T = \frac{2\pi m}{qB}$
Force on a current-carrying wire in a uniform magnetic field	$\vec{\mathbf{F}} = I\vec{\mathbf{l}} \times \vec{\mathbf{B}}$
Magnetic dipole moment	$\vec{\mu} = NIA\hat{\mathbf{n}}$
Torque on a current loop	$\vec{\tau} = \vec{\mu} \times \vec{\mathbf{B}}$
Energy of a magnetic dipole	$U = -\vec{\mu} \cdot \vec{\mathbf{B}}$
Drift velocity in crossed electric and magnetic fields	$v_d = \frac{E}{B}$
Hall potential	$V = \frac{IBl}{neA}$
Hall potential in terms of drift velocity	$V = Blv_d$
Charge-to-mass ratio in a mass spectrometer	$\frac{q}{m} = \frac{E}{BB_0R}$
Maximum speed of a particle in a cyclotron	$v_{\max} = \frac{qBR}{m}$

Conceptual Questions

Exercise:

Problem:

Describe the primary function of the electric field and the magnetic field in a cyclotron.

Problems

Exercise:

Problem:

A physicist is designing a cyclotron to accelerate protons to one-tenth the speed of light. The magnetic field will have a strength of 1.5 T. Determine (a) the rotational period of the circulating protons and (b) the maximum radius of the protons' orbit.

Solution:

a. 4.4×10^{-8} s; b. 0.21 m

Exercise:

Problem:

The strengths of the fields in the velocity selector of a Bainbridge mass spectrometer are $B = 0.500$ T and $E = 1.2 \times 10^5$ V/m, and the strength of the magnetic field that separates the ions is $B_o = 0.750$ T. A stream of singly charged Li ions is found to bend in a circular arc of radius 2.32 cm. What is the mass of the Li ions?

Exercise:

Problem:

The magnetic field in a cyclotron is 1.25 T, and the maximum orbital radius of the circulating protons is 0.40 m. (a) What is the kinetic energy of the protons when they are ejected from the cyclotron? (b) What is this energy in MeV? (c) Through what potential difference would a proton have to be accelerated to acquire this kinetic energy? (d) What is the period of the voltage source used to accelerate the protons? (e) Repeat the calculations for alpha-particles.

Solution:

a. 1.92×10^{-12} J; b. 12 MeV; c. 12 MV; d. 5.2×10^{-8} s; e. 1.92×10^{-12} J, 12 MeV, 12 V, 10.4×10^{-8} s

Exercise:

Problem:

A mass spectrometer is being used to separate common oxygen-16 from the much rarer oxygen-18, taken from a sample of old glacial ice. (The relative abundance of these oxygen isotopes is related to climatic temperature at the time the ice was deposited.) The ratio of the masses of these two ions is 16 to 18, the mass of oxygen-16 is 2.66×10^{-26} kg, and they are singly charged and travel at 5.00×10^6 m/s in a 1.20-T magnetic field. What is the separation between their paths when they hit a target after traversing a semicircle?

Exercise:

Problem:

(a) Triply charged uranium-235 and uranium-238 ions are being separated in a mass spectrometer. (The much rarer uranium-235 is used as reactor fuel.) The masses of the ions are $3.90 \times 10^{-25} \text{ kg}$ and $3.95 \times 10^{-25} \text{ kg}$, respectively, and they travel at $3.0 \times 10^5 \text{ m/s}$ in a 0.250-T field. What is the separation between their paths when they hit a target after traversing a semicircle? (b) Discuss whether this distance between their paths seems to be big enough to be practical in the separation of uranium-235 from uranium-238.

Solution:

a. $2.50 \times 10^{-2} \text{ m}$; b. Yes, this distance between their paths is clearly big enough to separate the U-235 from the U-238, since it is a distance of 2.5 cm.

Additional Problems**Exercise:****Problem:**

Calculate the magnetic force on a hypothetical particle of charge $1.0 \times 10^{-19} \text{ C}$ moving with a velocity of $6.0 \times 10^4 \hat{\mathbf{i}} \text{ m/s}$ in a magnetic field of $1.2 \hat{\mathbf{k}} \text{ T}$.

Exercise:

Problem: Repeat the previous problem with a new magnetic field of $(0.4 \hat{\mathbf{i}} + 1.2 \hat{\mathbf{k}}) \text{ T}$.

Solution:

$$-7.2 \times 10^{-15} \text{ N} \hat{\mathbf{j}}$$

Exercise:**Problem:**

An electron is projected into a uniform magnetic field $(0.5 \hat{\mathbf{i}} + 0.8 \hat{\mathbf{k}}) \text{ T}$ with a velocity of $(3.0 \hat{\mathbf{i}} + 4.0 \hat{\mathbf{j}}) \times 10^6 \text{ m/s}$. What is the magnetic force on the electron?

Exercise:**Problem:**

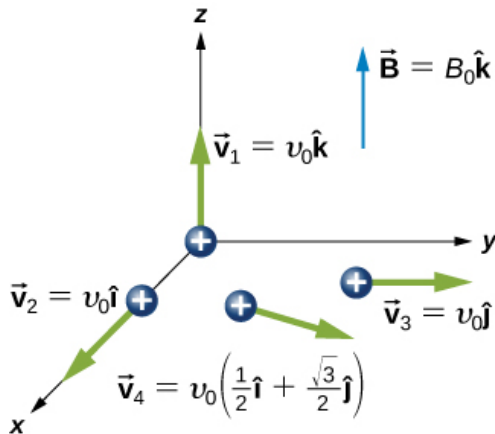
The mass and charge of a water droplet are $1.0 \times 10^{-4} \text{ g}$ and $2.0 \times 10^{-8} \text{ C}$, respectively. If the droplet is given an initial horizontal velocity of $5.0 \times 10^5 \hat{\mathbf{i}} \text{ m/s}$, what magnetic field will keep it moving in this direction? Why must gravity be considered here?

Solution:

$9.8 \times 10^{-5} \hat{\mathbf{j}} \text{ T}$; the magnetic and gravitational forces must balance to maintain dynamic equilibrium

Exercise:**Problem:**

Four different proton velocities are given. For each case, determine the magnetic force on the proton in terms of e , v_0 , and B_0 .

**Exercise:****Problem:**

An electron of kinetic energy 2000 eV passes between parallel plates that are 1.0 cm apart and kept at a potential difference of 300 V. What is the strength of the uniform magnetic field B that will allow the electron to travel undeflected through the plates? Assume E and B are perpendicular.

Solution:

$$1.13 \times 10^{-3} \text{T}$$

Exercise:**Problem:**

An alpha-particle ($m = 6.64 \times 10^{-27} \text{kg}$, $q = 3.2 \times 10^{-19} \text{C}$) moving with a velocity $\vec{v} = (2.0\hat{i} - 4.0\hat{k}) \times 10^6 \text{m/s}$ enters a region where $\vec{E} = (5.0\hat{i} - 2.0\hat{j}) \times 10^4 \text{V/m}$ and $\vec{B} = (1.0\hat{i} + 4.0\hat{k}) \times 10^{-2} \text{T}$. What is the initial force on it?

Exercise:**Problem:**

An electron moving with a velocity $\vec{v} = (4.0\hat{i} + 3.0\hat{j} + 2.0\hat{k}) \times 10^6 \text{m/s}$ enters a region where there is a uniform electric field and a uniform magnetic field. The magnetic field is given by $\vec{B} = (1.0\hat{i} - 2.0\hat{j} + 4.0\hat{k}) \times 10^{-2} \text{T}$. If the electron travels through a region without being deflected, what is the electric field?

Solution:

$$(1.6\hat{\mathbf{i}} - 1.4\hat{\mathbf{j}} - 1.1\hat{\mathbf{k}}) \times 10^5 \text{ V/m}$$

Exercise:**Problem:**

At a particular instant, an electron is traveling west to east with a kinetic energy of 10 keV. Earth's magnetic field has a horizontal component of $1.8 \times 10^{-5} \text{ T}$ north and a vertical component of $5.0 \times 10^{-5} \text{ T}$ down. (a) What is the path of the electron? (b) What is the radius of curvature of the path?

Exercise:**Problem:**

What is the (a) path of a proton and (b) the magnetic force on the proton that is traveling west to east with a kinetic energy of 10 keV in Earth's magnetic field that has a horizontal component of $1.8 \times 10^{-5} \text{ T}$ north and a vertical component of $5.0 \times 10^{-5} \text{ T}$ down?

Solution:

a. circular motion in a north, down plane; b. $(1.61\hat{\mathbf{j}} - 0.58\hat{\mathbf{k}}) \times 10^{-14} \text{ N}$

Exercise:**Problem:**

What magnetic field is required in order to confine a proton moving with a speed of $4.0 \times 10^6 \text{ m/s}$ to a circular orbit of radius 10 cm?

Exercise:**Problem:**

An electron and a proton move with the same speed in a plane perpendicular to a uniform magnetic field. Compare the radii and periods of their orbits.

Solution:

The proton has more mass than the electron; therefore, its radius and period will be larger.

Exercise:**Problem:**

A proton and an alpha-particle have the same kinetic energy and both move in a plane perpendicular to a uniform magnetic field. Compare the periods of their orbits.

Exercise:**Problem:**

A singly charged ion takes $2.0 \times 10^{-3} \text{ s}$ to complete eight revolutions in a uniform magnetic field of magnitude $2.0 \times 10^{-2} \text{ T}$. What is the mass of the ion?

Solution:

$$1.3 \times 10^{-25} \text{ kg}$$

Exercise:**Problem:**

A particle moving downward at a speed of $6.0 \times 10^6 \text{ m/s}$ enters a uniform magnetic field that is horizontal and directed from east to west. (a) If the particle is deflected initially to the north in a circular arc, is its charge positive or negative? (b) If $B = 0.25 \text{ T}$ and the charge-to-mass ratio (q/m) of the particle is $4.0 \times 10^7 \text{ C/kg}$, what is the radius of the path? (c) What is the speed of the particle after it has moved in the field for $1.0 \times 10^{-5} \text{ s}$? for 2.0 s ?

Exercise:**Problem:**

A proton, deuteron, and an alpha-particle are all accelerated from rest through the same potential difference. They then enter the same magnetic field, moving perpendicular to it. Compute the ratios of the radii of their circular paths. Assume that $m_d = 2m_p$ and $m_\alpha = 4m_p$.

Solution:

$$1:0.707:1$$

Exercise:**Problem:**

A singly charged ion is moving in a uniform magnetic field of $7.5 \times 10^{-2} \text{ T}$ completes 10 revolutions in $3.47 \times 10^{-4} \text{ s}$. Identify the ion.

Exercise:**Problem:**

Two particles have the same linear momentum, but particle A has four times the charge of particle B. If both particles move in a plane perpendicular to a uniform magnetic field, what is the ratio R_A/R_B of the radii of their circular orbits?

Solution:

$$1/4$$

Exercise:**Problem:**

A uniform magnetic field of magnitude B is directed parallel to the z -axis. A proton enters the field with a velocity $\vec{v} = (4\hat{j} + 3\hat{k}) \times 10^6 \text{ m/s}$ and travels in a helical path with a radius of 5.0 cm . (a) What is the value of B ? (b) What is the time required for one trip around the helix? (c) Where is the proton $5.0 \times 10^{-7} \text{ s}$ after entering the field?

Exercise:**Problem:**

An electron moving along the $+x$ -axis at $5.0 \times 10^6 \text{ m/s}$ enters a magnetic field that makes a 75° angle with the x -axis of magnitude 0.20 T. Calculate the (a) pitch and (b) radius of the trajectory.

Solution:

a. $2.3 \times 10^{-4} \text{ m}$; b. $1.37 \times 10^{-4} \text{ m}$

Exercise:**Problem:**

(a) A 0.750-m-long section of cable carrying current to a car starter motor makes an angle of 60° with Earth's $5.5 \times 10^{-5} \text{ T}$ field. What is the current when the wire experiences a force of $7.0 \times 10^{-3} \text{ N}$? (b) If you run the wire between the poles of a strong horseshoe magnet, subjecting 5.00 cm of it to a 1.75-T field, what force is exerted on this segment of wire?

Exercise:**Problem:**

(a) What is the angle between a wire carrying an 8.00-A current and the 1.20-T field it is in if 50.0 cm of the wire experiences a magnetic force of 2.40 N? (b) What is the force on the wire if it is rotated to make an angle of 90° with the field?

Solution:

a. 30.0° ; b. 4.80 N

Exercise:**Problem:**

A 1.0-m-long segment of wire lies along the x -axis and carries a current of 2.0 A in the positive x -direction. Around the wire is the magnetic field of $(3.0\hat{i} \times 4.0\hat{k}) \times 10^{-3} \text{ T}$. Find the magnetic force on this segment.

Exercise:**Problem:**

A 5.0-m section of a long, straight wire carries a current of 10 A while in a uniform magnetic field of magnitude $8.0 \times 10^{-3} \text{ T}$. Calculate the magnitude of the force on the section if the angle between the field and the direction of the current is (a) 45° ; (b) 90° ; (c) 0° ; or (d) 180° .

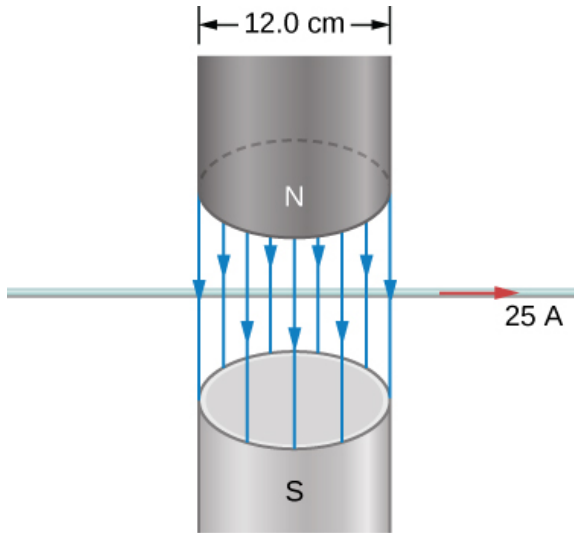
Solution:

a. 0.283 N; b. 0.4 N; c. 0 N; d. 0 N

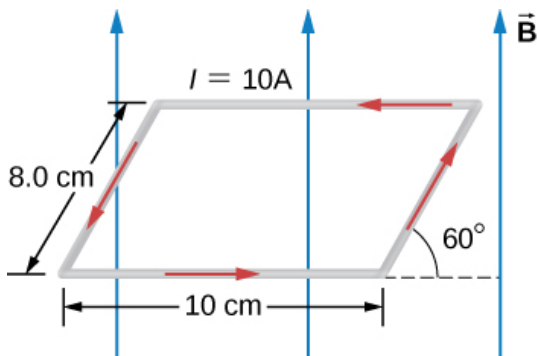
Exercise:

Problem:

An electromagnet produces a magnetic field of magnitude 1.5 T throughout a cylindrical region of radius 6.0 cm. A straight wire carrying a current of 25 A passes through the field as shown in the accompanying figure. What is the magnetic force on the wire?

**Exercise:****Problem:**

The current loop shown in the accompanying figure lies in the plane of the page, as does the magnetic field. Determine the net force and the net torque on the loop if $I = 10\text{ A}$ and $B = 1.5\text{ T}$.

**Solution:**

0 N and 0.012 Nm

Exercise:**Problem:**

A circular coil of radius 5.0 cm is wound with five turns and carries a current of 5.0 A. If the coil is placed in a uniform magnetic field of strength 5.0 T, what is the maximum torque on it?

Exercise:**Problem:**

A circular coil of wire of radius 5.0 cm has 20 turns and carries a current of 2.0 A. The coil lies in a magnetic field of magnitude 0.50 T that is directed parallel to the plane of the coil. (a) What is the magnetic dipole moment of the coil? (b) What is the torque on the coil?

Solution:

a. 0.31 Am^2 ; b. 0.16 Nm

Exercise:**Problem:**

A current-carrying coil in a magnetic field experiences a torque that is 75% of the maximum possible torque. What is the angle between the magnetic field and the normal to the plane of the coil?

Exercise:**Problem:**

A 4.0-cm by 6.0-cm rectangular current loop carries a current of 10 A. What is the magnetic dipole moment of the loop?

Solution:

0.024 Am^2

Exercise:**Problem:**

A circular coil with 200 turns has a radius of 2.0 cm. (a) What current through the coil results in a magnetic dipole moment of 3.0 Am^2 ? (b) What is the maximum torque that the coil will experience in a uniform field of strength $5.0 \times 10^{-2} \text{ T}$? (c) If the angle between μ and B is 45° , what is the magnitude of the torque on the coil? (d) What is the magnetic potential energy of coil for this orientation?

Exercise:**Problem:**

The current through a circular wire loop of radius 10 cm is 5.0 A. (a) Calculate the magnetic dipole moment of the loop. (b) What is the torque on the loop if it is in a uniform 0.20-T magnetic field such that μ and B are directed at 30° to each other? (c) For this position, what is the potential energy of the dipole?

Solution:

a. 0.16 Am^2 ; b. 0.016 Nm; c. 0.028 J

Exercise:

Problem:

A wire of length 1.0 m is wound into a single-turn planar loop. The loop carries a current of 5.0 A, and it is placed in a uniform magnetic field of strength 0.25 T. (a) What is the maximum torque that the loop will experience if it is square? (b) If it is circular? (c) At what angle relative to B would the normal to the circular coil have to be oriented so that the torque on it would be the same as the maximum torque on the square coil?

Exercise:**Problem:**

Consider an electron rotating in a circular orbit of radius r . Show that the magnitudes of the magnetic dipole moment μ and the angular momentum L of the electron are related by:

Equation:

$$\frac{\mu}{L} = \frac{e}{2m}.$$

Solution:

(Proof)

Exercise:**Problem:**

The Hall effect is to be used to find the sign of charge carriers in a semiconductor sample. The probe is placed between the poles of a magnet so that magnetic field is pointed up. A current is passed through a rectangular sample placed horizontally. As current is passed through the sample in the east direction, the north side of the sample is found to be at a higher potential than the south side. Decide if the number density of charge carriers is positively or negatively charged.

Exercise:**Problem:**

The density of charge carriers for copper is 8.47×10^{28} electrons per cubic meter. What will be the Hall voltage reading from a probe made up of $3 \text{ cm} \times 2 \text{ cm} \times 1 \text{ cm}$ ($L \times W \times T$) copper plate when a current of 1.5 A is passed through it in a magnetic field of 2.5 T perpendicular to the $3 \text{ cm} \times 2 \text{ cm}$.

Solution:

$$4.65 \times 10^{-7} \text{ V}$$

Exercise:

Problem:

The Hall effect is to be used to find the density of charge carriers in an unknown material. A Hall voltage $40\text{ }\mu\text{V}$ for 3-A current is observed in a 3-T magnetic field for a rectangular sample with length 2 cm, width 1.5 cm, and height 0.4 cm. Determine the density of the charge carriers.

Exercise:**Problem:**

Show that the Hall voltage across wires made of the same material, carrying identical currents, and subjected to the same magnetic field is inversely proportional to their diameters. (Hint: Consider how drift velocity depends on wire diameter.)

Solution:

Since $E = Blv$, where the width is twice the radius, $I = 2r$, $I = nqAv_d$,
 $v_d = \frac{I}{nqA} = \frac{I}{nq\pi r^2}$ so $E = B \times 2r \times \frac{I}{nq\pi r^2} = \frac{2IB}{nq\pi r} \propto \frac{1}{r} \propto \frac{1}{d}$.

The Hall voltage is inversely proportional to the diameter of the wire.

Exercise:**Problem:**

A velocity selector in a mass spectrometer uses a 0.100-T magnetic field. (a) What electric field strength is needed to select a speed of $4.0 \times 10^6\text{ m/s}$? (b) What is the voltage between the plates if they are separated by 1.00 cm?

Exercise:**Problem:**

Find the radius of curvature of the path of a 25.0-MeV proton moving perpendicularly to the 1.20-T field of a cyclotron.

Solution:

$6.92 \times 10^7\text{ m/s}$; 0.602 m

Exercise:**Problem:**

Unreasonable results To construct a non-mechanical water meter, a 0.500-T magnetic field is placed across the supply water pipe to a home and the Hall voltage is recorded. (a) Find the flow rate through a 3.00-cm-diameter pipe if the Hall voltage is 60.0 mV. (b) What would the Hall voltage be for the same flow rate through a 10.0-cm-diameter pipe with the same field applied?

Exercise:

Problem:

Unreasonable results A charged particle having mass $6.64 \times 10^{-27} \text{ kg}$ (that of a helium atom) moving at $8.70 \times 10^5 \text{ m/s}$ perpendicular to a 1.50-T magnetic field travels in a circular path of radius 16.0 mm. (a) What is the charge of the particle? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

a. $2.4 \times 10^{-19} \text{ C}$; b. not an integer multiple of e ; c. need to assume all charges have multiples of e , could be other forces not accounted for

Exercise:**Problem:**

Unreasonable results An inventor wants to generate 120-V power by moving a 1.00-m-long wire perpendicular to Earth's $5.00 \times 10^{-5} \text{ T}$ field. (a) Find the speed with which the wire must move. (b) What is unreasonable about this result? (c) Which assumption is responsible?

Exercise:**Problem:**

Unreasonable results Frustrated by the small Hall voltage obtained in blood flow measurements, a medical physicist decides to increase the applied magnetic field strength to get a 0.500-V output for blood moving at 30.0 cm/s in a 1.50-cm-diameter vessel. (a) What magnetic field strength is needed? (b) What is unreasonable about this result? (c) Which premise is responsible?

Solution:

a. $B = 5 \text{ T}$; b. very large magnet; c. applying such a large voltage

Challenge Problems**Exercise:****Problem:**

A particle of charge $+q$ and mass m moves with velocity \vec{v}_0 pointed in the $+y$ -direction as it crosses the x -axis at $x = R$ at a particular time. There is a negative charge $-Q$ fixed at the origin, and there exists a uniform magnetic field \vec{B}_0 pointed in the $+z$ -direction. It is found that the particle describes a circle of radius R about $-Q$. Find \vec{B}_0 in terms of the given quantities.

Exercise:**Problem:**

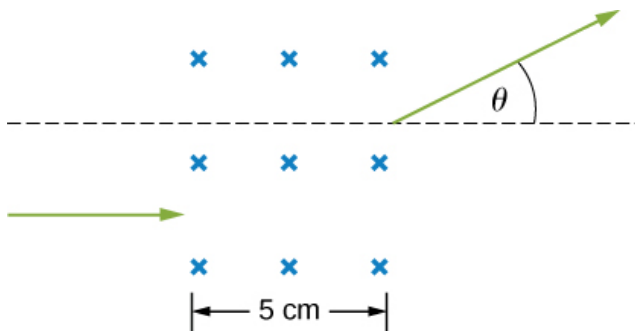
A proton of speed $v = 6 \times 10^5 \text{ m/s}$ enters a region of uniform magnetic field of $B = 0.5 \text{ T}$ at an angle of $q = 30^\circ$ to the magnetic field. In the region of magnetic field proton describes a helical path with radius R and pitch p (distance between loops). Find R and p .

Solution:

$$R = (mv \sin \theta) / qB; p = \left(\frac{2\pi m}{eB} \right) v \cos \theta$$

Exercise:**Problem:**

A particle's path is bent when it passes through a region of non-zero magnetic field although its speed remains unchanged. This is very useful for "beam steering" in particle accelerators. Consider a proton of speed 4×10^6 m/s entering a region of uniform magnetic field 0.2 T over a 5-cm-wide region. Magnetic field is perpendicular to the velocity of the particle. By how much angle will the path of the proton be bent? (Hint: The particle comes out tangent to a circle.)

**Exercise:****Problem:**

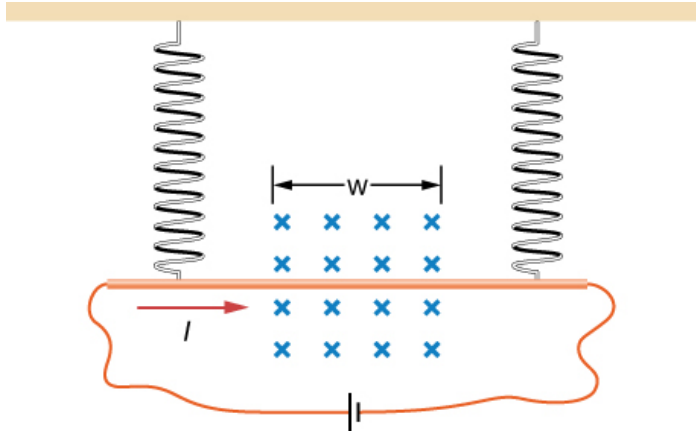
In a region a non-uniform magnetic field exists such that $B_x = 0$, $B_y = 0$, and $B_z = ax$, where a is a constant. At some time t , a wire of length L is carrying a current I is located along the x -axis from origin to $x = L$. Find the magnetic force on the wire at this instant in time.

Solution:

$$IaL^2/2$$

Exercise:**Problem:**

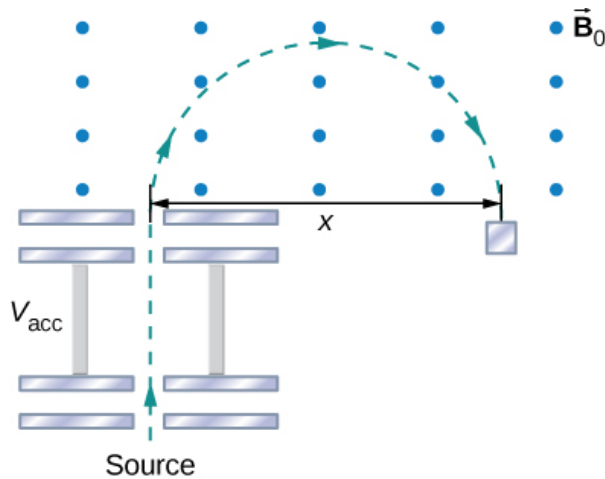
A copper rod of mass m and length L is hung from the ceiling using two springs of spring constant k . A uniform magnetic field of magnitude B_0 pointing perpendicular to the rod and spring (coming into the page in the figure) exists in a region of space covering a length w of the copper rod. The ends of the rod are then connected by flexible copper wire across the terminals of a battery of voltage V . Determine the change in the length of the springs when a current I runs through the copper rod in the direction shown in figure. (Ignore any force by the flexible wire.)



Exercise:

Problem:

The accompanied figure shows an arrangement for measuring mass of ions by an instrument called the mass spectrometer. An ion of mass m and charge $+q$ is produced essentially at rest in source S , a chamber in which a gas discharge is taking place. The ion is accelerated by a potential difference V_{acc} and allowed to enter a region of constant magnetic field \vec{B}_0 . In the uniform magnetic field region, the ion moves in a semicircular path striking a photographic plate at a distance x from the entry point. Derive a formula for mass m in terms of B_0 , q , V_{acc} , and x .



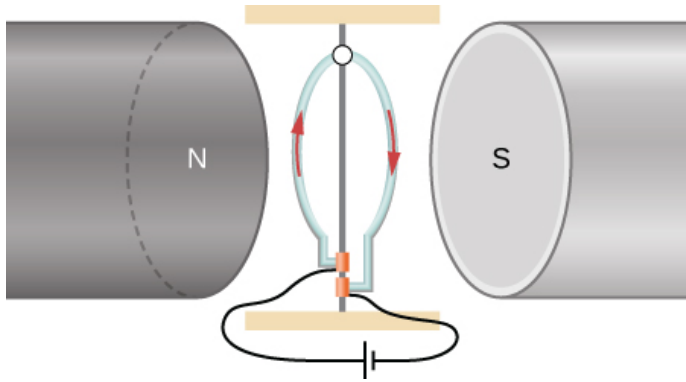
Solution:

$$m = \frac{qB_0^2}{8V_{\text{acc}}} x^2$$

Exercise:

Problem:

A wire is made into a circular shape of radius R and pivoted along a central support. The two ends of the wire are touching a brush that is connected to a dc power source. The structure is between the poles of a magnet such that we can assume there is a uniform magnetic field on the wire. In terms of a coordinate system with origin at the center of the ring, magnetic field is $B_x = B_0, B_y = B_z = 0$, and the ring rotates about the z -axis. Find the torque on the ring when it is not in the xz -plane.

**Exercise:****Problem:**

A long-rigid wire lies along the x -axis and carries a current of 2.5 A in the positive x -direction. Around the wire is the magnetic field $\vec{B} = 2.0\hat{i} + 5.0x^2\hat{j}$, with x in meters and B in millitesla. Calculate the magnetic force on the segment of wire between $x = 2.0$ m and $x = 4.0$ m.

Solution:

0.23 N

Exercise:**Problem:**

A circular loop of wire of area 10 cm^2 carries a current of 25 A. At a particular instant, the loop lies in the xy -plane and is subjected to a magnetic field $\vec{B} = (2.0\hat{i} + 6.0\hat{j} + 8.0\hat{k}) \times 10^{-3} \text{ T}$. As viewed from above the xy -plane, the current is circulating clockwise. (a) What is the magnetic dipole moment of the current loop? (b) At this instant, what is the magnetic torque on the loop?

Glossary

cyclotron

device used to accelerate charged particles to large kinetic energies

dees

large metal containers used in cyclotrons that serve contain a stream of charged particles as their speed is increased

mass spectrometer

device that separates ions according to their charge-to-mass ratios

Introduction

class="introduction"

An external hard drive attached to a computer works by magnetically encoding information that can be stored or retrieved quickly. A key idea in the development of digital devices is the ability to produce and use magnetic fields in this way. (credit: modification of work by “Miss Karen”/Flickr)



In the preceding chapter, we saw that a moving charged particle produces a magnetic field. This connection between electricity and magnetism is exploited in electromagnetic devices, such as a computer hard drive. In fact, it is the underlying principle behind most of the technology in modern society, including telephones, television, computers, and the internet.

In this chapter, we examine how magnetic fields are created by arbitrary distributions of electric current, using the Biot-Savart law. Then we look at how current-carrying wires create magnetic fields and deduce the forces that arise between two current-carrying wires due to these magnetic fields. We also study the torques produced by the magnetic fields of current loops. We then generalize these results to an important law of electromagnetism, called Ampère's law.

We examine some devices that produce magnetic fields from currents in geometries based on loops, known as solenoids and toroids. Finally, we look at how materials behave in magnetic fields and categorize materials based on their responses to magnetic fields.

The Biot-Savart Law

By the end of this section, you will be able to:

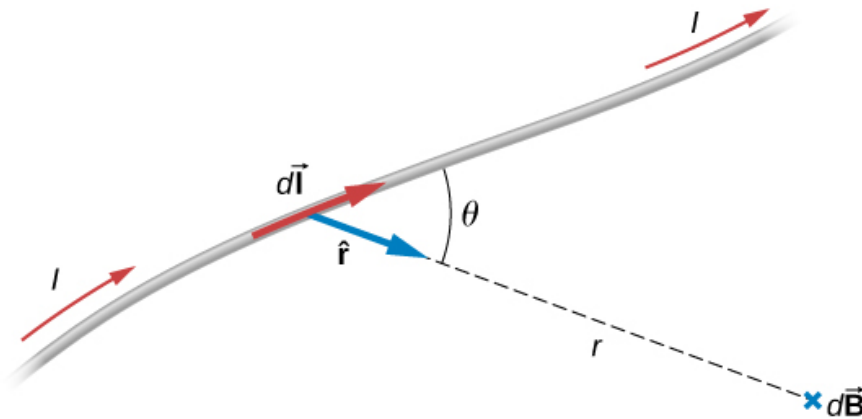
- Explain how to derive a magnetic field from an arbitrary current in a line segment
- Calculate magnetic field from the Biot-Savart law in specific geometries, such as a current in a line and a current in a circular arc

We have seen that mass produces a gravitational field and also interacts with that field. Charge produces an electric field and also interacts with that field. Since moving charge (that is, current) interacts with a magnetic field, we might expect that it also creates that field—and it does.

The equation used to calculate the magnetic field produced by a current is known as the Biot-Savart law. It is an empirical law named in honor of two scientists who investigated the interaction between a straight, current-carrying wire and a permanent magnet. This law enables us to calculate the magnitude and direction of the magnetic field produced by a current in a wire. The **Biot-Savart law** states that at any point P ([link](#)), the magnetic field $d\vec{B}$ due to an element $d\vec{l}$ of a current-carrying wire is given by

Equation:

$$d\vec{B} = \frac{\mu_0}{4\pi} \frac{I d\vec{l} \times \hat{r}}{r^2}.$$



A current element $I d\vec{l}$ produces a magnetic field at point P given by the Biot-Savart law.

The constant μ_0 is known as the **permeability of free space** and is exactly

Note:

Equation:

$$\mu_0 = 4\pi \times 10^{-7} \text{T} \cdot \text{m/A}$$

in the SI system. The infinitesimal wire segment $d\vec{l}$ is in the same direction as the current I (assumed positive), r is the distance from $d\vec{l}$ to P and \hat{r} is a unit vector that points from $d\vec{l}$ to P , as shown in the figure.

The direction of $d\vec{B}$ is determined by applying the right-hand rule to the vector product $d\vec{l} \times \hat{r}$. The magnitude of $d\vec{B}$ is

Note:

Equation:

$$dB = \frac{\mu_0}{4\pi} \frac{I dl \sin \theta}{r^2}$$

where θ is the angle between $d\vec{l}$ and \hat{r} . Notice that if $\theta = 0$, then $d\vec{B} = \vec{0}$. The field produced by a current element $I d\vec{l}$ has no component parallel to $d\vec{l}$.

The magnetic field due to a finite length of current-carrying wire is found by integrating [\[link\]](#) along the wire, giving us the usual form of the Biot-Savart law.

Note:

Biot-Savart law

The magnetic field \vec{B} due to an element $d\vec{l}$ of a current-carrying wire is given by

Equation:

$$\vec{B} = \frac{\mu_0}{4\pi} \int_{\text{wire}} \frac{I d\vec{l} \times \hat{r}}{r^2}.$$

Since this is a vector integral, contributions from different current elements may not point in the same direction. Consequently, the integral is often difficult to evaluate, even for fairly simple geometries. The following strategy may be helpful.

Note:

Solving Biot-Savart Problems

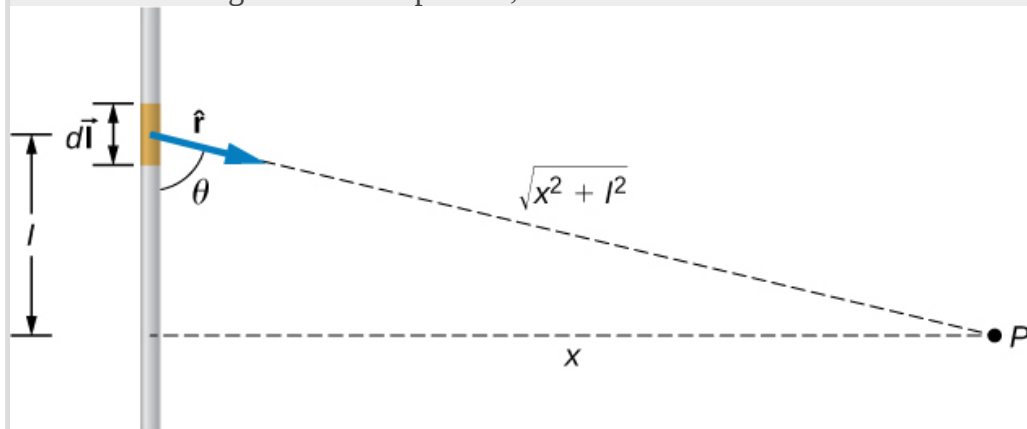
To solve Biot-Savart law problems, the following steps are helpful:

1. Identify that the Biot-Savart law is the chosen method to solve the given problem. If there is symmetry in the problem comparing \vec{B} and $d\vec{l}$, Ampère's law may be the preferred method to solve the question, which will be discussed in [Ampère's Law](#).
2. Draw the current element length $d\vec{l}$ and the unit vector \hat{r} , noting that $d\vec{l}$ points in the direction of the current and \hat{r} points from the current element toward the point where the field is desired.
3. Calculate the cross product $d\vec{l} \times \hat{r}$. The resultant vector gives the direction of the magnetic field according to the Biot-Savart law.
4. Use [\[link\]](#) and substitute all given quantities into the expression to solve for the magnetic field. Note all variables that remain constant over the entire length of the wire may be factored out of the integration.
5. Use the right-hand rule to verify the direction of the magnetic field produced from the current or to write down the direction of the magnetic field if only the magnitude was solved for in the previous part.

Example:

Calculating Magnetic Fields of Short Current Segments

A short wire of length 1.0 cm carries a current of 2.0 A in the vertical direction ([\[link\]](#)). The rest of the wire is shielded so it does not add to the magnetic field produced by the wire. Calculate the magnetic field at point P , which is 1 meter from the wire in the x -direction.



A small line segment carries a current I in the vertical direction. What is the magnetic field at a distance x from the segment?

Strategy

We can determine the magnetic field at point P using the Biot-Savart law. Since the current segment is much smaller than the distance x , we can drop the integral from the expression. The integration is converted back into a summation, but only for small dl , which we now write as Δl . Another way to think about it is that each of the radius values is nearly the same, no matter where the current element is on the line segment, if Δl is small compared to x . The angle θ is calculated using a tangent function. Using the numbers given, we can calculate the magnetic field at P .

Solution

The angle between $\Delta \vec{l}$ and \hat{r} is calculated from trigonometry, knowing the distances l and x from the problem:

Equation:

$$\theta = \tan^{-1} \frac{1 \text{ m}}{0.01 \text{ m}} = 89.4^\circ.$$

The magnetic field at point P is calculated by the Biot-Savart law:

Equation:

$$B = \frac{\mu_0}{4\pi} \frac{I \Delta l \sin \theta}{r^2} = (1 \times 10^{-7} \text{ T} \cdot \text{m/A}) \frac{2 \text{ A}(0.01 \text{ m}) \sin(89.4^\circ)}{(1 \text{ m})^2} = 2.0 \times 10^{-9} \text{ T}.$$

From the right-hand rule and the Biot-Savart law, the field is directed into the page.

Significance

This approximation is only good if the length of the line segment is very small compared to the distance from the current element to the point. If not, the integral form of the Biot-Savart law must be used over the entire line segment to calculate the magnetic field.

Note:

Exercise:

Problem:

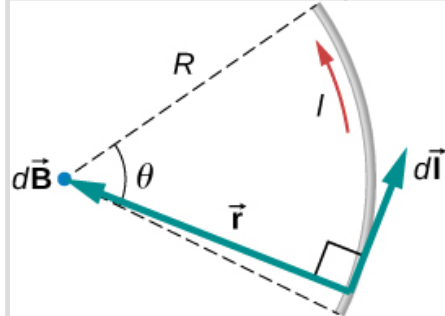
Check Your Understanding Using [\[link\]](#), at what distance would P have to be to measure a magnetic field half of the given answer?

Solution:

1.41 meters

Example:**Calculating Magnetic Field of a Circular Arc of Wire**

A wire carries a current I in a circular arc with radius R swept through an arbitrary angle θ ([link](#)). Calculate the magnetic field at the center of this arc at point P .



A wire segment carrying a current I . The path $d\vec{l}$ and radial direction \hat{r} are indicated.

Strategy

We can determine the magnetic field at point P using the Biot-Savart law. The radial and path length directions are always at a right angle, so the cross product turns into multiplication. We also know that the distance along the path dl is related to the radius times the angle θ (in radians). Then we can pull all constants out of the integration and solve for the magnetic field.

Solution

The Biot-Savart law starts with the following equation:

Equation:

$$\vec{B} = \frac{\mu_0}{4\pi} \int_{\text{wire}} \frac{I d\vec{l} \times \hat{r}}{r^2}.$$

As we integrate along the arc, all the contributions to the magnetic field are in the same direction (out of the page), so we can work with the magnitude of the field. The cross product turns into multiplication because the path dl and the radial direction are perpendicular. We can also substitute the arc length formula, $dl = r d\theta$:

Equation:

$$B = \frac{\mu_0}{4\pi} \int_{\text{wire}} \frac{I r d\theta}{r^2}.$$

The current and radius can be pulled out of the integral because they are the same regardless of where we are on the path. This leaves only the integral over the angle,

Equation:

$$B = \frac{\mu_0 I}{4\pi r} \int_{\text{wire}} d\theta.$$

The angle varies on the wire from 0 to θ ; hence, the result is

Equation:

$$B = \frac{\mu_0 I \theta}{4\pi r}.$$

Significance

The direction of the magnetic field at point P is determined by the right-hand rule, as shown in the previous chapter. If there are other wires in the diagram along with the arc, and you are asked to find the net magnetic field, find each contribution from a wire or arc and add the results by superposition of vectors. Make sure to pay attention to the direction of each contribution. Also note that in a symmetric situation, like a straight or circular wire, contributions from opposite sides of point P cancel each other.

Note:**Exercise:****Problem:**

Check Your Understanding The wire loop forms a full circle of radius R and current I . What is the magnitude of the magnetic field at the center?

Solution:

$$\frac{\mu_0 I}{2R}$$

Summary

- The magnetic field created by a current-carrying wire is found by the Biot-Savart law.
- The current element $I d\vec{l}$ produces a magnetic field a distance r away.

Conceptual Questions**Exercise:**

Problem:

For calculating magnetic fields, what are the advantages and disadvantages of the Biot-Savart law?

Solution:

Biot-Savart law's advantage is that it works with any magnetic field produced by a current loop. The disadvantage is that it can take a long time.

Exercise:**Problem:**

Describe the magnetic field due to the current in two wires connected to the two terminals of a source of emf and twisted tightly around each other.

Exercise:

Problem: How can you decide if a wire is infinite?

Solution:

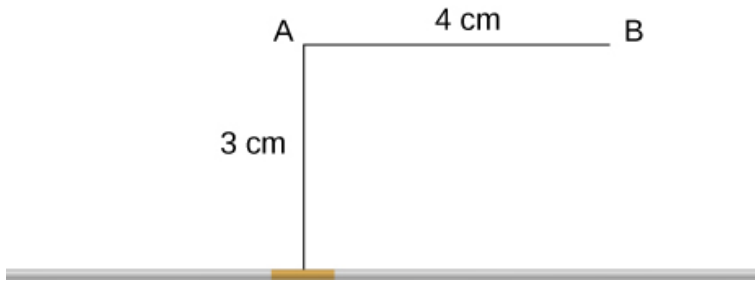
If you were to go to the start of a line segment and calculate the angle θ to be approximately 0° , the wire can be considered infinite. This judgment is based also on the precision you need in the result.

Exercise:**Problem:**

Identical currents are carried in two circular loops; however, one loop has twice the diameter as the other loop. Compare the magnetic fields created by the loops at the center of each loop.

Problems**Exercise:****Problem:**

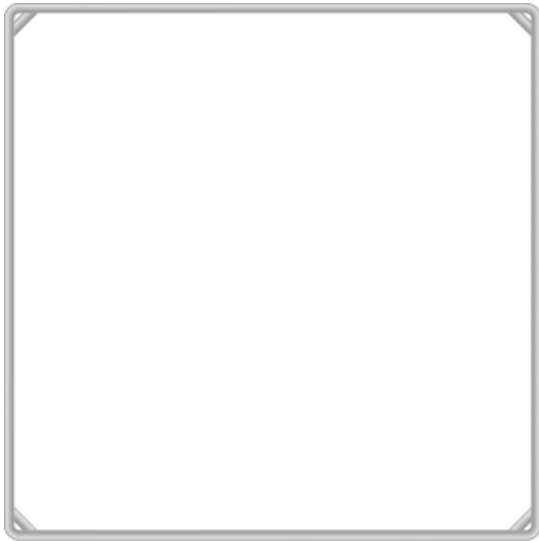
A 10-A current flows through the wire shown. What is the magnitude of the magnetic field due to a 0.5-mm segment of wire as measured at (a) point A and (b) point B?



Exercise:

Problem:

Ten amps flow through a square loop where each side is 20 cm in length. At each corner of the loop is a 0.01-cm segment that connects the longer wires as shown. Calculate the magnitude of the magnetic field at the center of the loop.

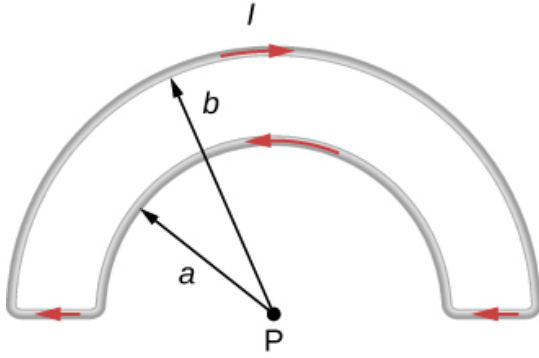


Solution:

$$5.66 \times 10^{-5} \text{T}$$

Exercise:

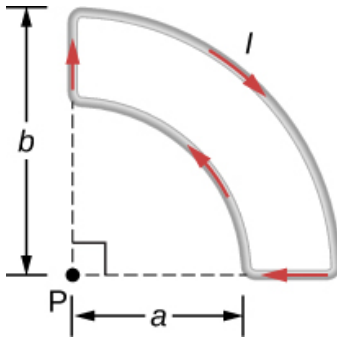
Problem: What is the magnetic field at P due to the current I in the wire shown?



Exercise:

Problem:

The accompanying figure shows a current loop consisting of two concentric circular arcs and two perpendicular radial lines. Determine the magnetic field at point P.



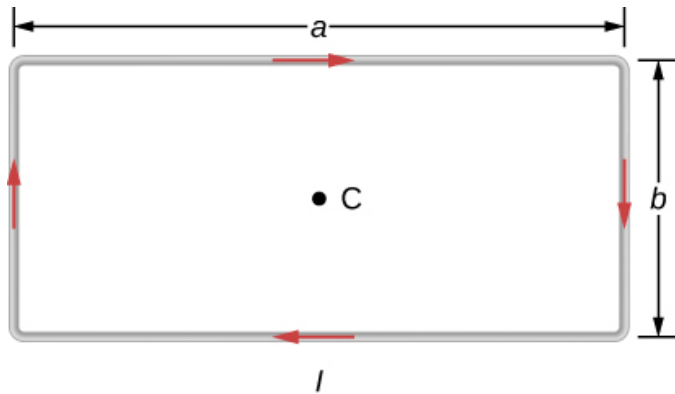
Solution:

$$B = \frac{\mu_0 I}{8} \left(\frac{1}{a} - \frac{1}{b} \right) \text{ out of the page}$$

Exercise:

Problem:

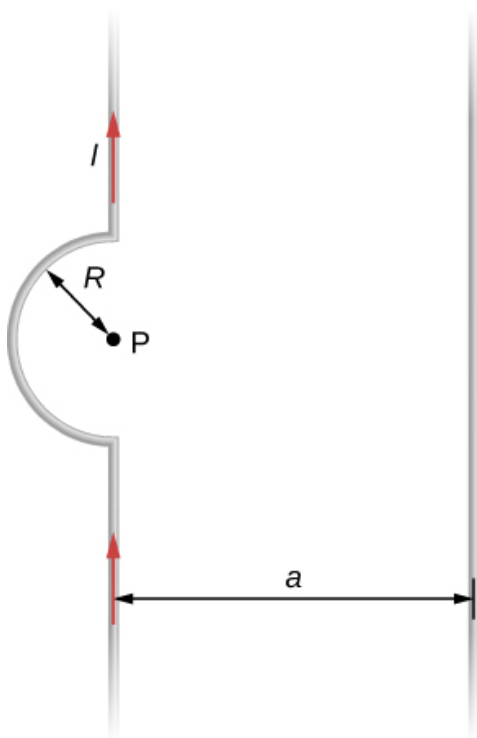
Find the magnetic field at the center C of the rectangular loop of wire shown in the accompanying figure.



Exercise:

Problem:

Two long wires, one of which has a semicircular bend of radius R , are positioned as shown in the accompanying figure. If both wires carry a current I , how far apart must their parallel sections be so that the net magnetic field at P is zero? Does the current in the straight wire flow up or down?



Solution:

$a = \frac{2R}{\pi}$; the current in the wire to the right must flow up the page.

Glossary

Biot-Savart law

an equation giving the magnetic field at a point produced by a current-carrying wire

permeability of free space

μ_0 , measure of the ability of a material, in this case free space, to support a magnetic field

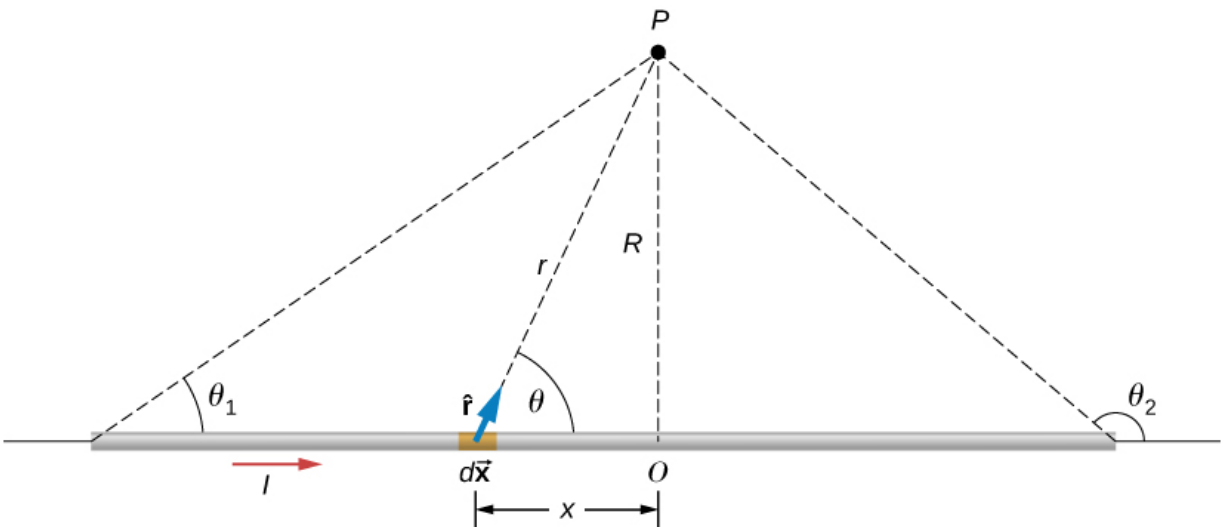
Magnetic Field Due to a Thin Straight Wire

By the end of this section, you will be able to:

- Explain how the Biot-Savart law is used to determine the magnetic field due to a thin, straight wire.
- Determine the dependence of the magnetic field from a thin, straight wire based on the distance from it and the current flowing in the wire.
- Sketch the magnetic field created from a thin, straight wire by using the second right-hand rule.

How much current is needed to produce a significant magnetic field, perhaps as strong as Earth's field? Surveyors will tell you that overhead electric power lines create magnetic fields that interfere with their compass readings. Indeed, when Oersted discovered in 1820 that a current in a wire affected a compass needle, he was not dealing with extremely large currents. How does the shape of wires carrying current affect the shape of the magnetic field created? We noted in [Chapter 11](#) that a current loop created a magnetic field similar to that of a bar magnet, but what about a straight wire? We can use the Biot-Savart law to answer all of these questions, including determining the magnetic field of a long straight wire.

[\[link\]](#) shows a section of an infinitely long, straight wire that carries a current I . What is the magnetic field at a point P , located a distance R from the wire?



A section of a thin, straight current-carrying wire. The independent variable θ has the limits θ_1 and θ_2 .

Let's begin by considering the magnetic field due to the current element $I d\vec{x}$ located at the position x . Using the right-hand rule 1 from the previous chapter, $d\vec{x} \times \hat{r}$ points out of the page for any element along the wire. At point P , therefore, the magnetic fields due to all current elements have the same direction. This means that we can calculate the net field there by evaluating the scalar sum of the contributions of the elements. With $|d\vec{x} \times \hat{r}| = (dx)(1) \sin \theta$, we have from the Biot-Savart law

Equation:

$$B = \frac{\mu_0}{4\pi} \int_{\text{wire}} \frac{I \sin \theta dx}{r^2}.$$

The wire is symmetrical about point O , so we can set the limits of the integration from zero to infinity and double the answer, rather than integrate from negative infinity to positive infinity. Based on the picture and geometry, we can write expressions for r and $\sin \theta$ in terms of x and R , namely:

Equation:

$$\begin{aligned} r &= \sqrt{x^2 + R^2} \\ \sin \theta &= \frac{R}{\sqrt{x^2 + R^2}}. \end{aligned}$$

Substituting these expressions into [\[link\]](#), the magnetic field integration becomes

Equation:

$$B = \frac{\mu_o I}{2\pi} \int_0^\infty \frac{R dx}{(x^2 + R^2)^{3/2}}.$$

Evaluating the integral yields

Equation:

$$B = \frac{\mu_o I}{2\pi R} \left[\frac{x}{(x^2 + R^2)^{1/2}} \right]_0^\infty.$$

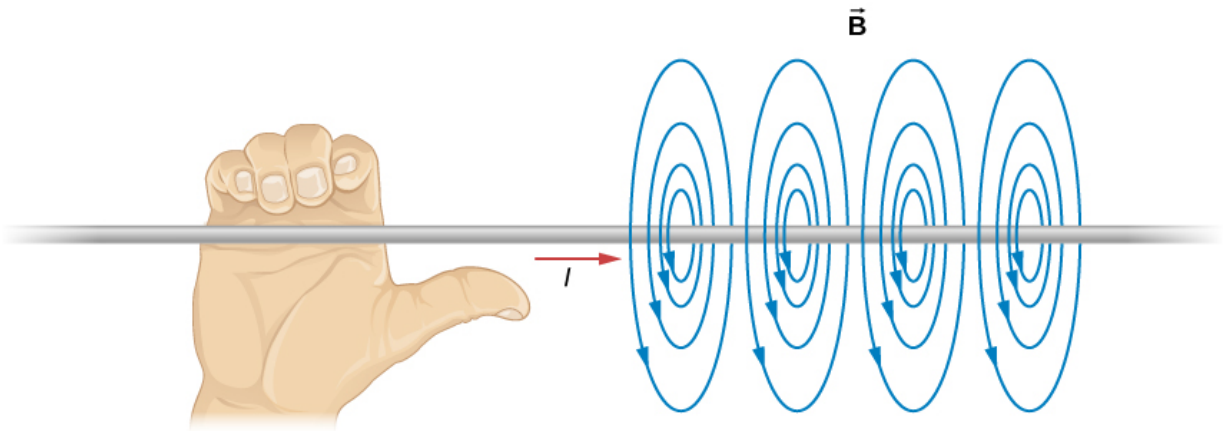
Substituting the limits gives us the solution

Note:

Equation:

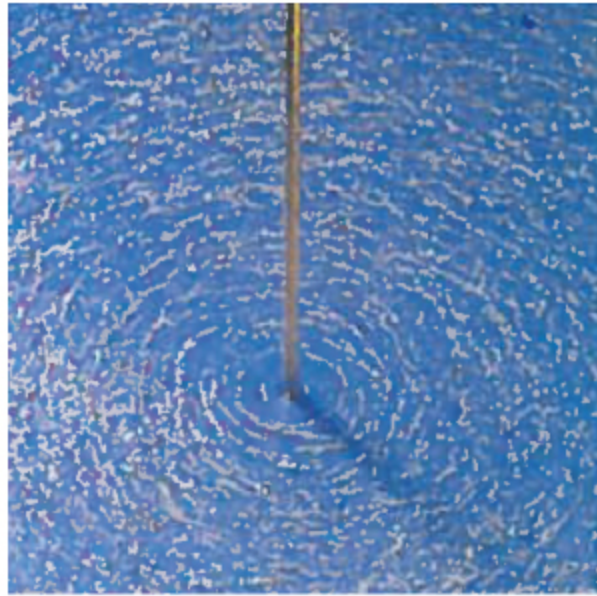
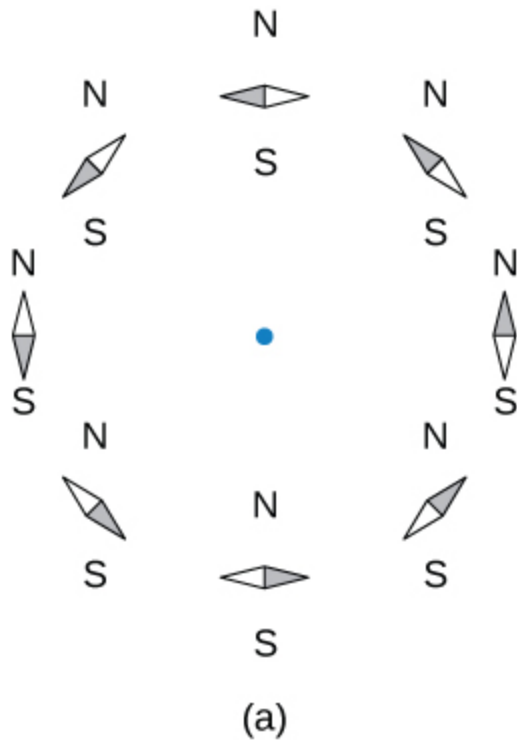
$$B = \frac{\mu_o I}{2\pi R}.$$

The magnetic field lines of the infinite wire are circular and centered at the wire ([\[link\]](#)), and they are identical in every plane perpendicular to the wire. Since the field decreases with distance from the wire, the spacing of the field lines must increase correspondingly with distance. The direction of this magnetic field may be found with a second form of the right-hand rule (illustrated in [\[link\]](#)). If you hold the wire with your right hand so that your thumb points along the current, then your fingers wrap around the wire in the same sense as \vec{B} .



Some magnetic field lines of an infinite wire. The direction of \vec{B} can be found with a form of the right-hand rule.

The direction of the field lines can be observed experimentally by placing several small compass needles on a circle near the wire, as illustrated in [\[link\]](#). When there is no current in the wire, the needles align with Earth's magnetic field. However, when a large current is sent through the wire, the compass needles all point tangent to the circle. Iron filings sprinkled on a horizontal surface also delineate the field lines, as shown in [\[link\]](#).

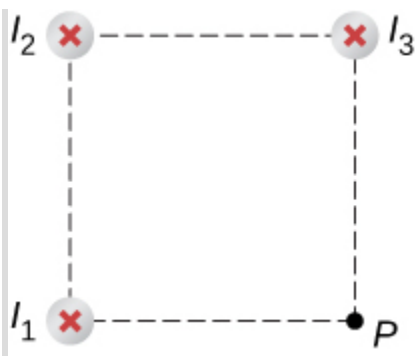


The shape of the magnetic field lines of a long wire can be seen using (a) small compass needles and (b) iron filings.

Example:

Calculating Magnetic Field Due to Three Wires

Three wires sit at the corners of a square, all carrying currents of 2 amps into the page as shown in [\[link\]](#). Calculate the magnitude of the magnetic field at the other corner of the square, point P , if the length of each side of the square is 1 cm.



Three wires have current flowing into the page. The magnetic field is determined at the fourth corner of the square.

Strategy

The magnetic field due to each wire at the desired point is calculated. The diagonal distance is calculated using the Pythagorean theorem. Next, the direction of each magnetic field's contribution is determined by drawing a circle centered at the point of the wire and out toward the desired point. The direction of the magnetic field contribution from that wire is tangential to the curve. Lastly, working with these vectors, the resultant is calculated.

Solution

Wires 1 and 3 both have the same magnitude of magnetic field contribution at point P :

Equation:

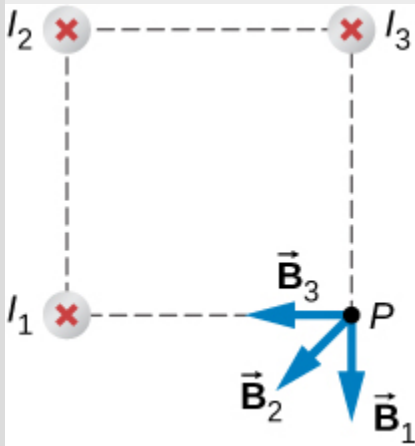
$$B_1 = B_3 = \frac{\mu_o I}{2\pi R} = \frac{(4\pi \times 10^{-7} \text{ T} \cdot \text{m/A})(2 \text{ A})}{2\pi(0.01 \text{ m})} = 4 \times 10^{-5} \text{ T}.$$

Wire 2 has a longer distance and a magnetic field contribution at point P of:

Equation:

$$B_2 = \frac{\mu_o I}{2\pi R} = \frac{(4\pi \times 10^{-7} \text{T} \cdot \text{m/A})(2 \text{ A})}{2\pi(0.01414 \text{ m})} = 3 \times 10^{-5} \text{T}.$$

The vectors for each of these magnetic field contributions are shown.



The magnetic field in the x-direction has contributions from wire 3 and the x-component of wire 2:

Equation:

$$B_{\text{net } x} = -4 \times 10^{-5} \text{T} - 2.83 \times 10^{-5} \text{T} \cos(45^\circ) = -6 \times 10^{-5} \text{T}.$$

The y-component is similarly the contributions from wire 1 and the y-component of wire 2:

Equation:

$$B_{\text{net } y} = -4 \times 10^{-5} \text{T} - 2.83 \times 10^{-5} \text{T} \sin(45^\circ) = -6 \times 10^{-5} \text{T}.$$

Therefore, the net magnetic field is the resultant of these two components:

Equation:

$$\begin{aligned} B_{\text{net}} &= \sqrt{B_{\text{net } x}^2 + B_{\text{net } y}^2} \\ B_{\text{net}} &= \sqrt{(-6 \times 10^{-5} \text{T})^2 + (-6 \times 10^{-5} \text{T})^2} \\ B_{\text{net}} &= 8 \times 10^{-5} \text{T}. \end{aligned}$$

Significance

The geometry in this problem results in the magnetic field contributions in the x - and y -directions having the same magnitude. This is not necessarily the case if the currents were different values or if the wires were located in different positions. Regardless of the numerical results, working on the components of the vectors will yield the resulting magnetic field at the point in need.

Note:

Exercise:

Problem:

Check Your Understanding Using [\[link\]](#), keeping the currents the same in wires 1 and 3, what should the current be in wire 2 to counteract the magnetic fields from wires 1 and 3 so that there is no net magnetic field at point P?

Solution:

4 amps flowing out of the page

Summary

- The strength of the magnetic field created by current in a long straight wire is given by $B = \frac{\mu_0 I}{2\pi R}$ (long straight wire) where I is the current, R is the shortest distance to the wire, and the constant $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/s}$ is the permeability of free space.
- The direction of the magnetic field created by a long straight wire is given by right-hand rule 2 (RHR-2): Point the thumb of the right hand in the direction of current, and the fingers curl in the direction of the magnetic field loops created by it.

Conceptual Questions

Exercise:

Problem:

How would you orient two long, straight, current-carrying wires so that there is no net magnetic force between them? (*Hint: What orientation would lead to one wire not experiencing a magnetic field from the other?*)

Solution:

You would make sure the currents flow perpendicular to one another.

Problems

Exercise:

Problem:

A typical current in a lightning bolt is 10^4 A. Estimate the magnetic field 1 m from the bolt.

Exercise:

Problem:

The magnitude of the magnetic field 50 cm from a long, thin, straight wire is $8.0 \mu\text{T}$. What is the current through the long wire?

Solution:

20 A

Exercise:

Problem:

A transmission line strung 7.0 m above the ground carries a current of 500 A. What is the magnetic field on the ground directly below the wire? Compare your answer with the magnetic field of Earth.

Exercise:**Problem:**

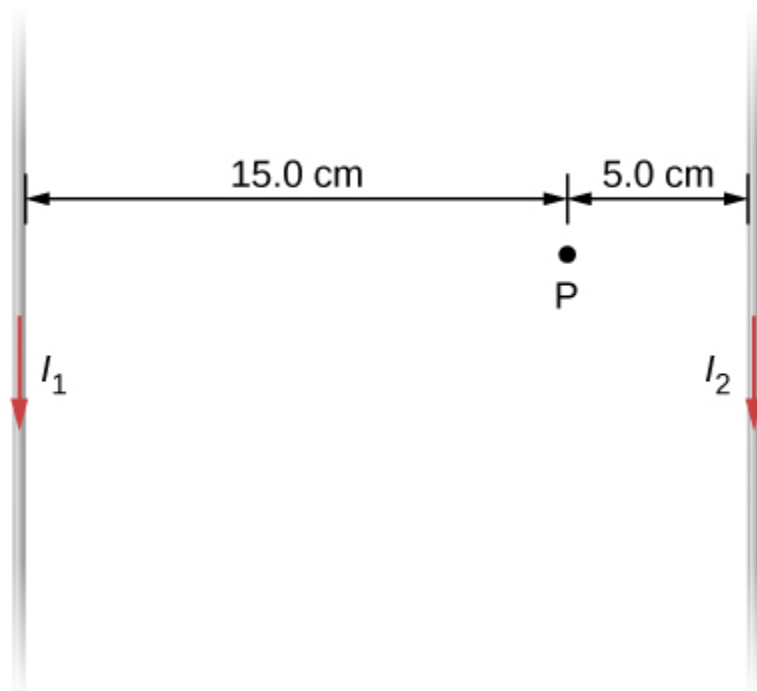
A long, straight, horizontal wire carries a left-to-right current of 20 A. If the wire is placed in a uniform magnetic field of magnitude $4.0 \times 10^{-5} \text{ T}$ that is directed vertically downward, what is the resultant magnitude of the magnetic field 20 cm above the wire? 20 cm below the wire?

Solution:

Both answers have the magnitude of magnetic field of $4.5 \times 10^{-5} \text{ T}$.

Exercise:**Problem:**

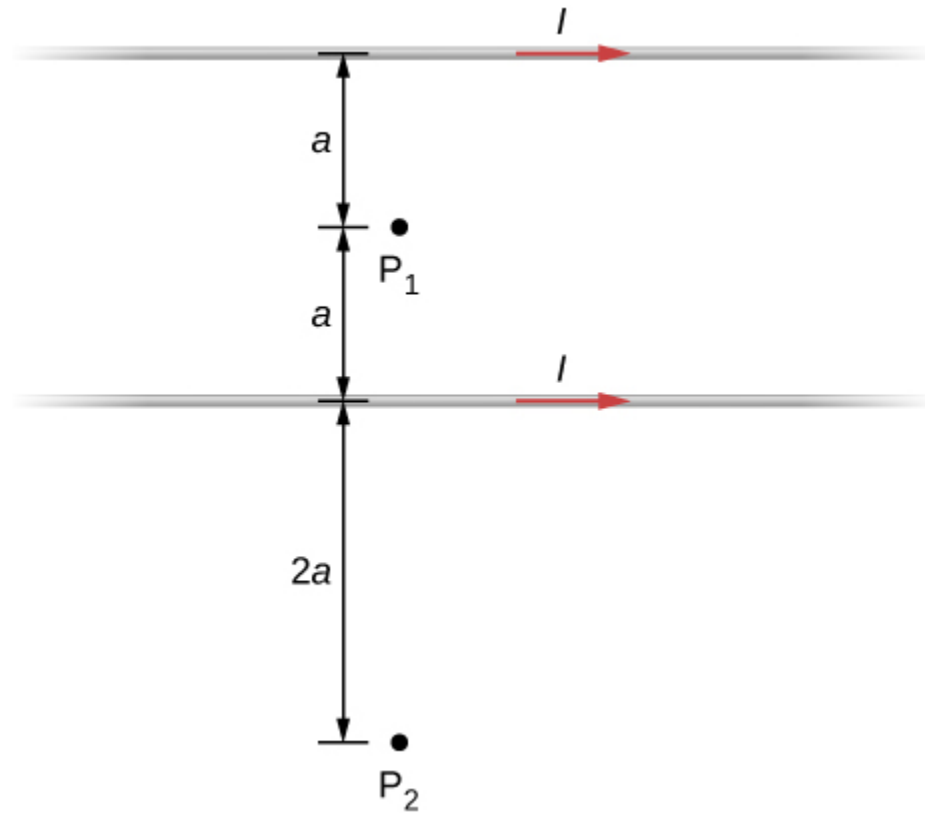
The two long, parallel wires shown in the accompanying figure carry currents in the same direction. If $I_1 = 10 \text{ A}$ and $I_2 = 20 \text{ A}$, what is the magnetic field at point P?



Exercise:

Problem:

The accompanying figure shows two long, straight, horizontal wires that are parallel and a distance $2a$ apart. If both wires carry current I in the same direction, (a) what is the magnetic field at P_1 ? (b) P_2 ?



Solution:

At P_1 , the net magnetic field is zero. At P_2 , $B = \frac{3\mu_0 I}{8\pi a}$ into the page.

Exercise:

Problem:

Repeat the calculations of the preceding problem with the direction of the current in the lower wire reversed.

Exercise:

Problem:

Consider the area between the wires of the preceding problem. At what distance from the top wire is the net magnetic field a minimum?

Assume that the currents are equal and flow in opposite directions.

Solution:

The magnetic field is at a minimum at distance a from the top wire, or half-way between the wires.

Magnetic Force between Two Parallel Currents

By the end of this section, you will be able to:

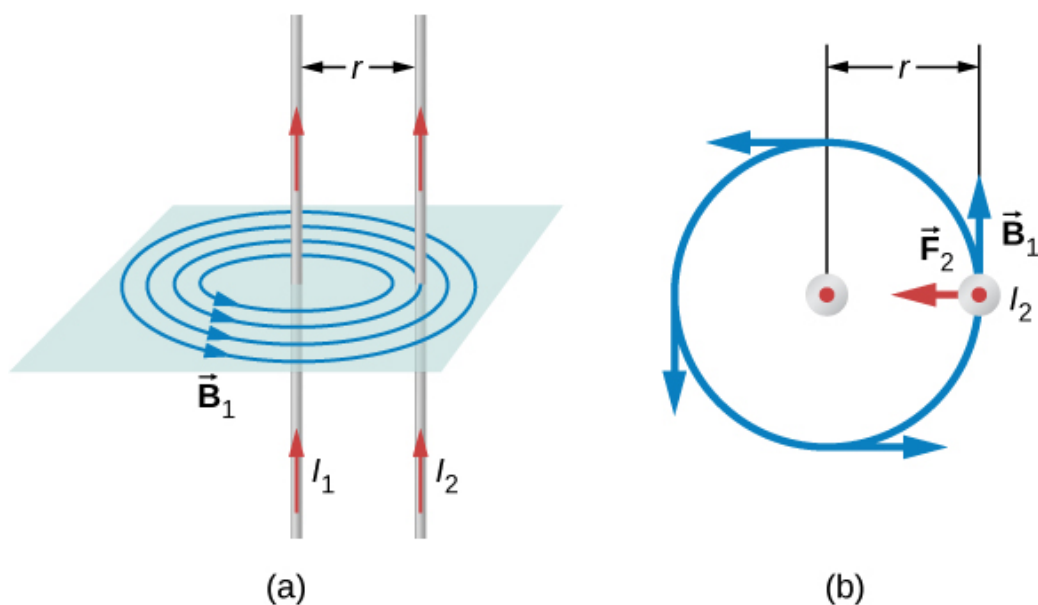
- Explain how parallel wires carrying currents can attract or repel each other
- Define the ampere and describe how it is related to current-carrying wires
- Calculate the force of attraction or repulsion between two current-carrying wires

You might expect that two current-carrying wires generate significant forces between them, since ordinary currents produce magnetic fields and these fields exert significant forces on ordinary currents. But you might not expect that the force between wires is used to define the ampere. It might also surprise you to learn that this force has something to do with why large circuit breakers burn up when they attempt to interrupt large currents.

The force between two long, straight, and parallel conductors separated by a distance r can be found by applying what we have developed in the preceding sections. [\[link\]](#) shows the wires, their currents, the field created by one wire, and the consequent force the other wire experiences from the created field. Let us consider the field produced by wire 1 and the force it exerts on wire 2 (call the force F_2). The field due to I_1 at a distance r is

Equation:

$$B_1 = \frac{\mu_0 I_1}{2\pi r}$$



(a) The magnetic field produced by a long straight conductor is perpendicular to a parallel conductor, as indicated by right hand

perpendicular to a parallel conductor, as indicated by right-hand rule (RHR)-2. (b) A view from above of the two wires shown in (a), with one magnetic field line shown for wire 1. RHR-1 shows that the force between the parallel conductors is attractive when the currents are in the same direction. A similar analysis shows that the force is repulsive between currents in opposite directions.

This field is uniform from the wire 1 and perpendicular to it, so the force F_2 it exerts on a length l of wire 2 is given by $F = IlB \sin \theta$ with $\sin \theta = 1$:

Equation:

$$F_2 = I_2 l B_1.$$

The forces on the wires are equal in magnitude, so we just write F for the magnitude of F_2 . (Note that $\vec{F}_1 = -\vec{F}_2$.) Since the wires are very long, it is convenient to think in terms of F/l , the force per unit length. Substituting the expression for B_1 into [\[link\]](#) and rearranging terms gives

Note:

Equation:

$$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}.$$

The ratio F/l is the force per unit length between two parallel currents I_1 and I_2 separated by a distance r . The force is attractive if the currents are in the same direction and repulsive if they are in opposite directions.

This force is responsible for the *pinch effect* in electric arcs and other plasmas. The force exists whether the currents are in wires or not. It is only apparent if the overall charge density is zero; otherwise, the Coulomb repulsion overwhelms the magnetic attraction. In an electric arc, where charges are moving parallel to one another, an attractive force squeezes currents into a smaller tube. In large circuit breakers, such as those used in neighborhood power distribution systems, the pinch effect can concentrate an arc between plates of a switch trying to break a large current, burn holes, and even ignite the equipment. Another example of the pinch effect is found in

the solar plasma, where jets of ionized material, such as solar flares, are shaped by magnetic forces.

The definition of the ampere is based on the force between current-carrying wires. Note that for long, parallel wires separated by 1 meter with each carrying 1 ampere, the force per meter is

Equation:

$$\frac{F}{l} = \frac{(4\pi \times 10^{-7} \text{ T} \cdot \text{m/A})(1 \text{ A})^2}{(2\pi)(1 \text{ m})} = 2 \times 10^{-7} \text{ N/m}.$$

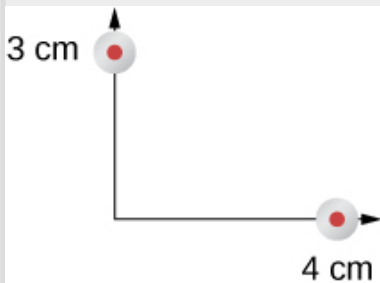
Since μ_0 is exactly $4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ by definition, and because $1 \text{ T} = 1 \text{ N}/(\text{A} \cdot \text{m})$, the force per meter is exactly $2 \times 10^{-7} \text{ N/m}$. This is the basis of the definition of the ampere.

Infinite-length wires are impractical, so in practice, a current balance is constructed with coils of wire separated by a few centimeters. Force is measured to determine current. This also provides us with a method for measuring the coulomb. We measure the charge that flows for a current of one ampere in one second. That is, $1 \text{ C} = 1 \text{ A} \cdot \text{s}$. For both the ampere and the coulomb, the method of measuring force between conductors is the most accurate in practice.

Example:

Calculating Forces on Wires

Two wires, both carrying current out of the page, have a current of magnitude 5.0 mA. The first wire is located at (0.0 cm, 3.0 cm) while the other wire is located at (4.0 cm, 0.0 cm) as shown in [\[link\]](#). What is the magnetic force per unit length of the first wire on the second and the second wire on the first?



Two current-carrying
wires at given
locations with

currents out of the
page.

Strategy

Each wire produces a magnetic field felt by the other wire. The distance along the hypotenuse of the triangle between the wires is the radial distance used in the calculation to determine the force per unit length. Since both wires have currents flowing in the same direction, the direction of the force is toward each other.

Solution

The distance between the wires results from finding the hypotenuse of a triangle:

Equation:

$$r = \sqrt{(3.0 \text{ cm})^2 + (4.0 \text{ cm})^2} = 5.0 \text{ cm}.$$

The force per unit length can then be calculated using the known currents in the wires:

Equation:

$$\frac{F}{l} = \frac{(4\pi \times 10^{-7} \text{ T} \cdot \text{m/A})(5 \times 10^{-3} \text{ A})^2}{(2\pi)(5 \times 10^{-2} \text{ m})} = 1 \times 10^{-10} \text{ N/m}.$$

The force from the first wire pulls the second wire. The angle between the radius and the x-axis is

Equation:

$$\theta = \tan^{-1} \left(\frac{3 \text{ cm}}{4 \text{ cm}} \right) = 36.9^\circ.$$

The unit vector for this is calculated by

Equation:

$$-\cos(36.9^\circ)\hat{\mathbf{i}} + \sin(36.9^\circ)\hat{\mathbf{j}} = -0.8\hat{\mathbf{i}} + 0.6\hat{\mathbf{j}}.$$

Therefore, the force per unit length from wire one on wire 2 is

Equation:

$$\frac{\vec{F}}{l} = (1 \times 10^{-10} \text{ N/m}) \times (-0.8\hat{\mathbf{i}} + 0.6\hat{\mathbf{j}}) = (-8 \times 10^{-11}\hat{\mathbf{i}} + 6 \times 10^{-11}\hat{\mathbf{j}}) \text{ N/m}.$$

The force per unit length from wire 2 on wire 1 is the negative of the previous answer:

Equation:

$$\frac{\vec{F}}{l} = (8 \times 10^{-11} \hat{i} - 6 \times 10^{-11} \hat{j}) \text{N/m}.$$

Significance

These wires produced magnetic fields of equal magnitude but opposite directions at each other's locations. Whether the fields are identical or not, the forces that the wires exert on each other are always equal in magnitude and opposite in direction (Newton's third law).

Note:

Exercise:

Problem:

Check Your Understanding Two wires, both carrying current out of the page, have a current of magnitude 2.0 mA and 3.0 mA, respectively. The first wire is located at (0.0 cm, 5.0 cm) while the other wire is located at (12.0 cm, 0.0 cm). What is the magnitude of the magnetic force per unit length of the first wire on the second and the second wire on the first?

Solution:

Both have a force per unit length of $9.23 \times 10^{-12} \text{ N/m}$

Summary

- The force between two parallel currents I_1 and I_2 , separated by a distance r , has a magnitude per unit length given by $\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}$.
- The force is attractive if the currents are in the same direction, repulsive if they are in opposite directions.

Conceptual Questions

Exercise:

Problem:

Compare and contrast the electric field of an infinite line of charge and the magnetic field of an infinite line of current.

Exercise:

Problem: Is \vec{B} constant in magnitude for points that lie on a magnetic field line?

Solution:

A magnetic field line gives the direction of the magnetic field at any point in space. The density of magnetic field lines indicates the strength of the magnetic field.

Problems**Exercise:****Problem:**

Two long, straight wires are parallel and 25 cm apart. (a) If each wire carries a current of 50 A in the same direction, what is the magnetic force per meter exerted on each wire? (b) Does the force pull the wires together or push them apart? (c) What happens if the currents flow in opposite directions?

Exercise:**Problem:**

Two long, straight wires are parallel and 10 cm apart. One carries a current of 2.0 A, the other a current of 5.0 A. (a) If the two currents flow in opposite directions, what is the magnitude and direction of the force per unit length of one wire on the other? (b) What is the magnitude and direction of the force per unit length if the currents flow in the same direction?

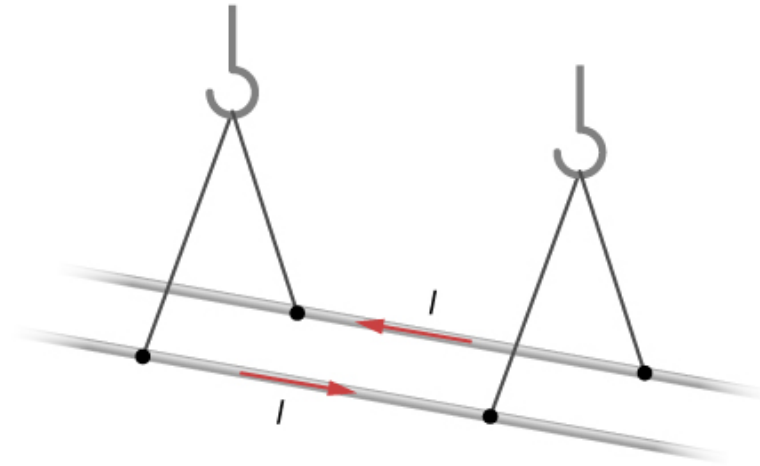
Solution:

a. $F/l = 8 \times 10^{-6} \text{ N/m}$ away from the other wire; b. $F/l = 8 \times 10^{-6} \text{ N/m}$ toward the other wire

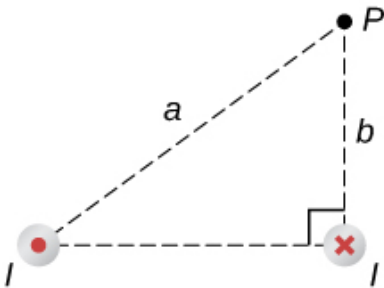
Exercise:

Problem:

Two long, parallel wires are hung by cords of length 5.0 cm, as shown in the accompanying figure. Each wire has a mass per unit length of 30 g/m, and they carry the same current in opposite directions. What is the current if the cords hang at 6.0° with respect to the vertical?

**Exercise:****Problem:**

A circuit with current I has two long parallel wire sections that carry current in opposite directions. Find magnetic field at a point P near these wires that is a distance a from one wire and b from the other wire as shown in the figure.

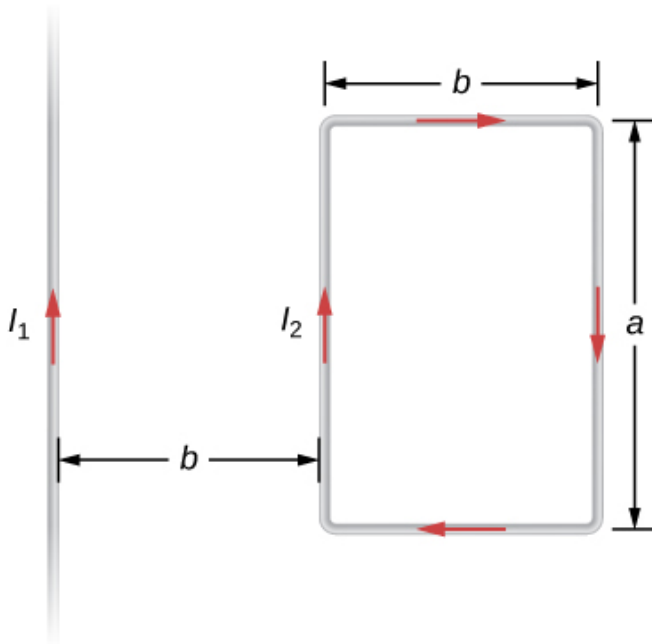
**Solution:**

$$B = \frac{\mu_o I}{2\pi a^2 b} \left((a_2 + b_2) \hat{\mathbf{i}} + b \sqrt{(a^2 - b^2)} \hat{\mathbf{j}} \right)$$

Exercise:

Problem:

The infinite, straight wire shown in the accompanying figure carries a current I_1 . The rectangular loop, whose long sides are parallel to the wire, carries a current I_2 . What are the magnitude and direction of the force on the rectangular loop due to the magnetic field of the wire?

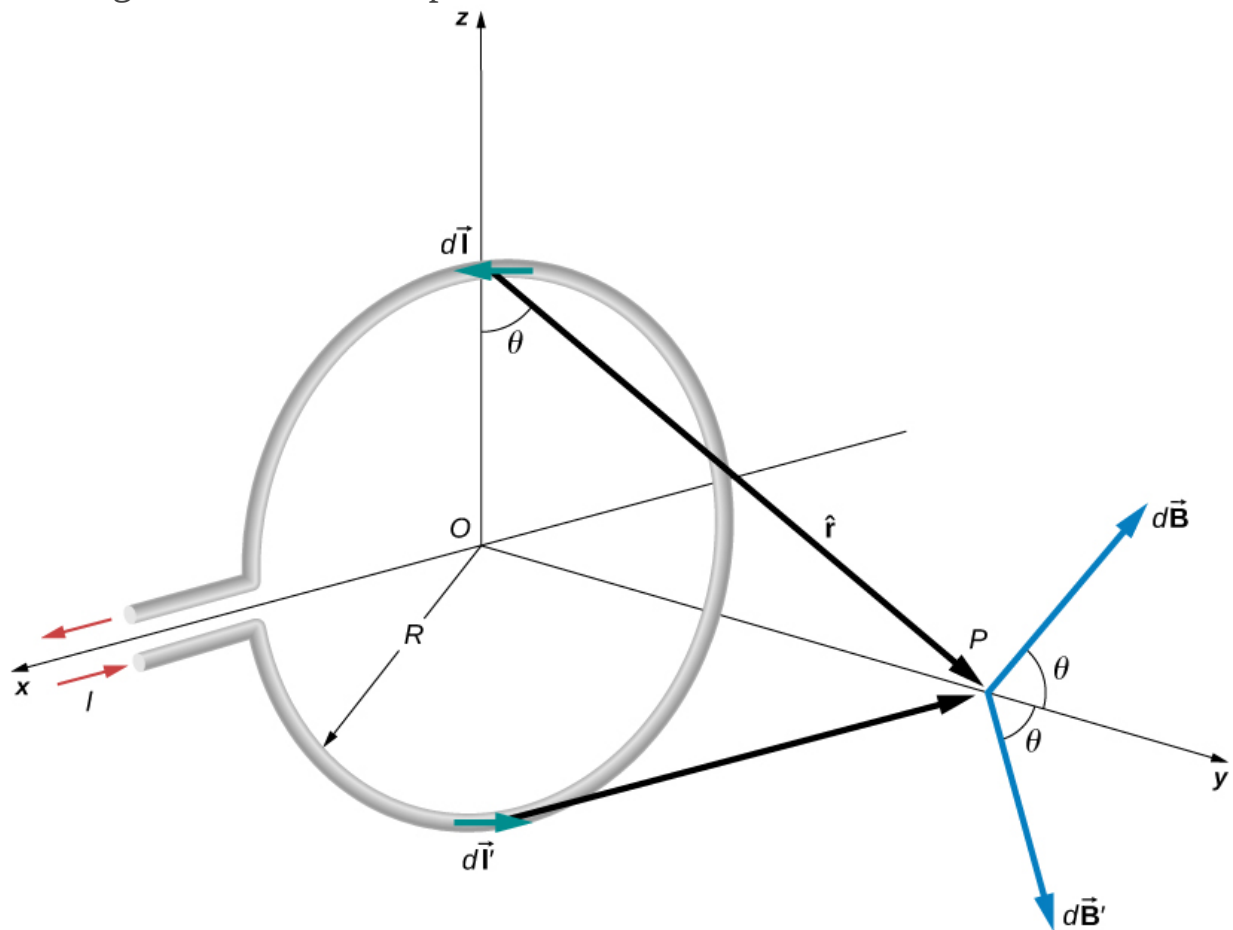


Magnetic Field of a Current Loop

By the end of this section, you will be able to:

- Explain how the Biot-Savart law is used to determine the magnetic field due to a current in a loop of wire at a point along a line perpendicular to the plane of the loop.
- Determine the magnetic field of an arc of current.

The circular loop of [\[link\]](#) has a radius R , carries a current I , and lies in the xz -plane. What is the magnetic field due to the current at an arbitrary point P along the axis of the loop?



Determining the magnetic field at point P along the axis of a current-carrying loop of wire.

We can use the Biot-Savart law to find the magnetic field due to a current. We first consider arbitrary segments on opposite sides of the loop to qualitatively show by the vector results that the net magnetic field direction is along the central axis from the loop. From there, we can use the Biot-Savart law to derive the expression for magnetic field.

Let P be a distance y from the center of the loop. From the right-hand rule, the magnetic field $d\vec{B}$ at P , produced by the current element $I d\vec{l}$, is directed at an angle θ above the y -axis as shown. Since $d\vec{l}$ is parallel along the x -axis and \hat{r} is in the yz -plane, the two vectors are perpendicular, so we have

Equation:

$$dB = \frac{\mu_0}{4\pi} \frac{I dl \sin \pi/2}{r^2} = \frac{\mu_0}{4\pi} \frac{I dl}{y^2 + R^2}$$

where we have used $r^2 = y^2 + R^2$.

Now consider the magnetic field $d\vec{B}'$ due to the current element $I d\vec{l}'$, which is directly opposite $I d\vec{l}$ on the loop. The magnitude of $d\vec{B}'$ is also given by [\[link\]](#), but it is directed at an angle θ below the y -axis. The components of $d\vec{B}$ and $d\vec{B}'$ perpendicular to the y -axis therefore cancel, and in calculating the net magnetic field, only the components along the y -axis need to be considered. The components perpendicular to the axis of the loop sum to zero in pairs. Hence at point P :

Equation:

$$\vec{B} = \hat{j} \int_{\text{loop}} dB \cos \theta = \hat{j} \frac{\mu_0 I}{4\pi} \int_{\text{loop}} \frac{\cos \theta dl}{y^2 + R^2}.$$

For all elements $d\vec{l}$ on the wire, y , R , and $\cos \theta$ are constant and are related by

Equation:

$$\cos \theta = \frac{R}{\sqrt{y^2 + R^2}}.$$

Now from [\[link\]](#), the magnetic field at P is

Equation:

$$\vec{\mathbf{B}} = \hat{\mathbf{j}} \frac{\mu_0 I R}{4\pi(y^2 + R^2)^{3/2}} \int_{\text{loop}} dl = \frac{\mu_0 I R^2}{2(y^2 + R^2)^{3/2}} \hat{\mathbf{j}}$$

where we have used $\int_{\text{loop}} dl = 2\pi R$. As discussed in the previous chapter,

the closed current loop is a magnetic dipole of moment $\mu = IA\hat{\mathbf{n}}$. For this example, $A = \pi R^2$ and $\hat{\mathbf{n}} = \hat{\mathbf{j}}$, so the magnetic field at P can also be written as

Equation:

$$\vec{\mathbf{B}} = \frac{\mu_0 \mu \hat{\mathbf{j}}}{2\pi(y^2 + R^2)^{3/2}}.$$

By setting $y = 0$ in [\[link\]](#), we obtain the magnetic field at the center of the loop:

Note:

Equation:

$$\vec{\mathbf{B}} = \frac{\mu_0 I}{2R} \hat{\mathbf{j}}.$$

This equation becomes $B = \mu_0 n I / (2R)$ for a flat coil of n loops per length. It can also be expressed as

Equation:

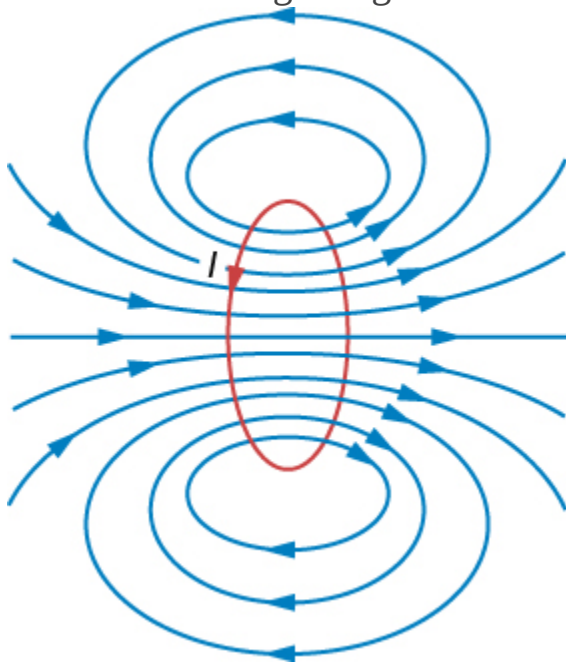
$$\vec{B} = \frac{\mu_0 \mu}{2\pi R^3}.$$

If we consider $y \gg R$ in [\[link\]](#), the expression reduces to an expression known as the magnetic field from a dipole:

Equation:

$$\vec{B} = \frac{\mu_0 \mu}{2\pi y^3}.$$

The calculation of the magnetic field due to the circular current loop at points off-axis requires rather complex mathematics, so we'll just look at the results. The magnetic field lines are shaped as shown in [\[link\]](#). Notice that one field line follows the axis of the loop. This is the field line we just found. Also, very close to the wire, the field lines are almost circular, like the lines of a long straight wire.

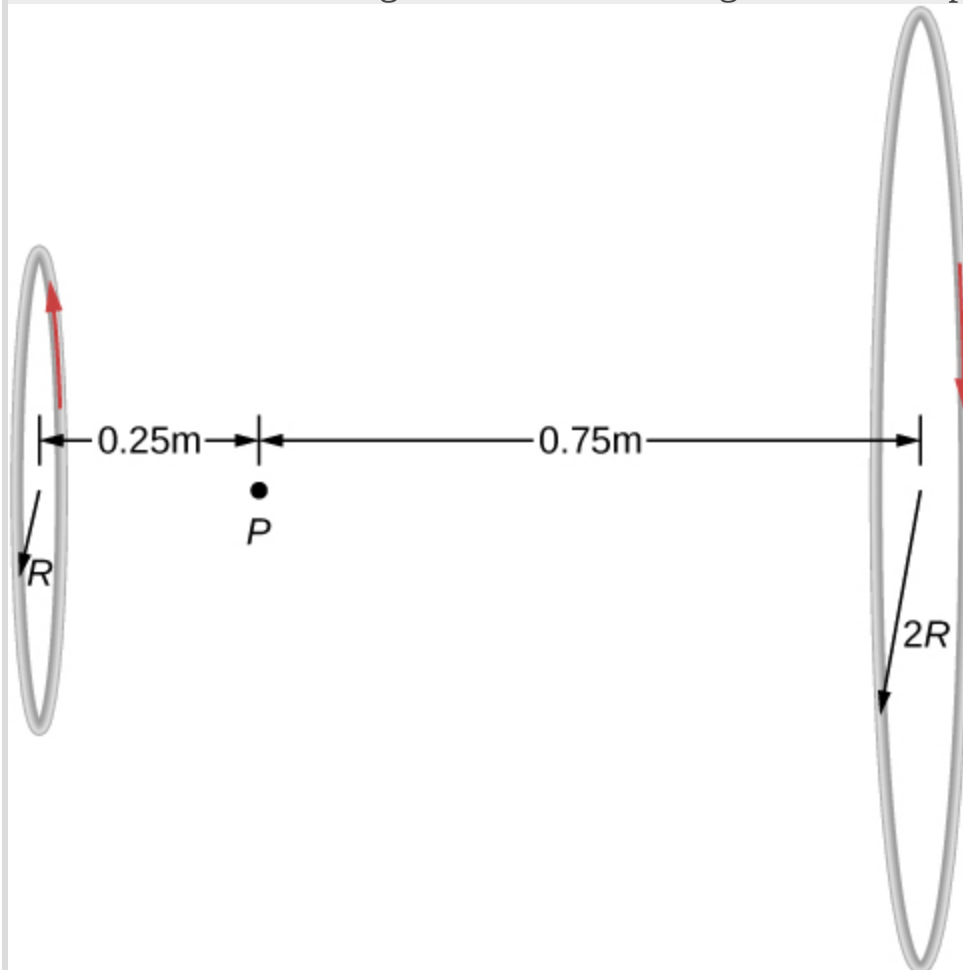


Sketch of the magnetic field lines of a circular current loop.

Example:

Magnetic Field between Two Loops

Two loops of wire carry the same current of 10 mA, but flow in opposite directions as seen in [\[link\]](#). One loop is measured to have a radius of $R = 50$ cm while the other loop has a radius of $2R = 100$ cm. The distance from the first loop to the point where the magnetic field is measured is 0.25 m, and the distance from that point to the second loop is 0.75 m. What is the magnitude of the net magnetic field at point P ?



Two loops of different radii have the same current but flowing in opposite directions. The magnetic field at point P is measured to be zero.

Strategy

The magnetic field at point P has been determined in [\[link\]](#). Since the currents are flowing in opposite directions, the net magnetic field is the difference between the two fields generated by the coils. Using the given quantities in the problem, the net magnetic field is then calculated.

Solution

Solving for the net magnetic field using [\[link\]](#) and the given quantities in the problem yields

Equation:

$$\begin{aligned} B &= \frac{\mu_0 I R_1^2}{2(y_1^2 + R_1^2)^{3/2}} - \frac{\mu_0 I R_2^2}{2(y_2^2 + R_2^2)^{3/2}} \\ B &= \frac{(4\pi \times 10^{-7} \text{ T}\cdot\text{m/A})(0.010 \text{ A})(0.5 \text{ m})^2}{2((0.25 \text{ m})^2 + (0.5 \text{ m})^2)^{3/2}} - \frac{(4\pi \times 10^{-7} \text{ T}\cdot\text{m/A})(0.010 \text{ A})(1.0 \text{ m})^2}{2((0.75 \text{ m})^2 + (1.0 \text{ m})^2)^{3/2}} \\ B &= 5.77 \times 10^{-9} \text{ T to the right.} \end{aligned}$$

Significance

Helmholtz coils typically have loops with equal radii with current flowing in the same direction to have a strong uniform field at the midpoint between the loops. A similar application of the magnetic field distribution created by Helmholtz coils is found in a magnetic bottle that can temporarily trap charged particles. See [Magnetic Forces and Fields](#) for a discussion on this.

Note:

Exercise:

Problem:

Check Your Understanding Using [\[link\]](#), at what distance would you have to move the first coil to have zero measurable magnetic field at point P ?

Solution:

0.608 meters

Summary

- The magnetic field strength at the center of a circular loop is given by $B = \frac{\mu_0 I}{2R}$ (at center of loop), where R is the radius of the loop. RHR-2 gives the direction of the field about the loop.

Conceptual Questions

Exercise:

Problem: Is the magnetic field of a current loop uniform?

Exercise:**Problem:**

What happens to the length of a suspended spring when a current passes through it?

Solution:

The spring reduces in length since each coil will have a north pole-produced magnetic field next to a south pole of the next coil.

Exercise:

Problem:

Two concentric circular wires with different diameters carry currents in the same direction. Describe the force on the inner wire.

Problems**Exercise:****Problem:**

When the current through a circular loop is 6.0 A, the magnetic field at its center is $2.0 \times 10^{-4} \text{T}$. What is the radius of the loop?

Solution:

0.019 m

Exercise:**Problem:**

How many turns must be wound on a flat, circular coil of radius 20 cm in order to produce a magnetic field of magnitude $4.0 \times 10^{-5} \text{T}$ at the center of the coil when the current through it is 0.85 A?

Exercise:**Problem:**

A flat, circular loop has 20 turns. The radius of the loop is 10.0 cm and the current through the wire is 0.50 A. Determine the magnitude of the magnetic field at the center of the loop.

Solution:

$N \times 6.28 \times 10^{-5} \text{T}$

Exercise:

Problem:

A circular loop of radius R carries a current I . At what distance along the axis of the loop is the magnetic field one-half its value at the center of the loop?

Exercise:**Problem:**

Two flat, circular coils, each with a radius R and wound with N turns, are mounted along the same axis so that they are parallel a distance d apart. What is the magnetic field at the midpoint of the common axis if a current I flows in the same direction through each coil?

Solution:

$$B = \frac{\mu_o I R^2}{\left(\left(\frac{d}{2}\right)^2 + R^2\right)^{3/2}}$$

Exercise:**Problem:**

For the coils in the preceding problem, what is the magnetic field at the center of either coil?

Ampère's Law

By the end of this section, you will be able to:

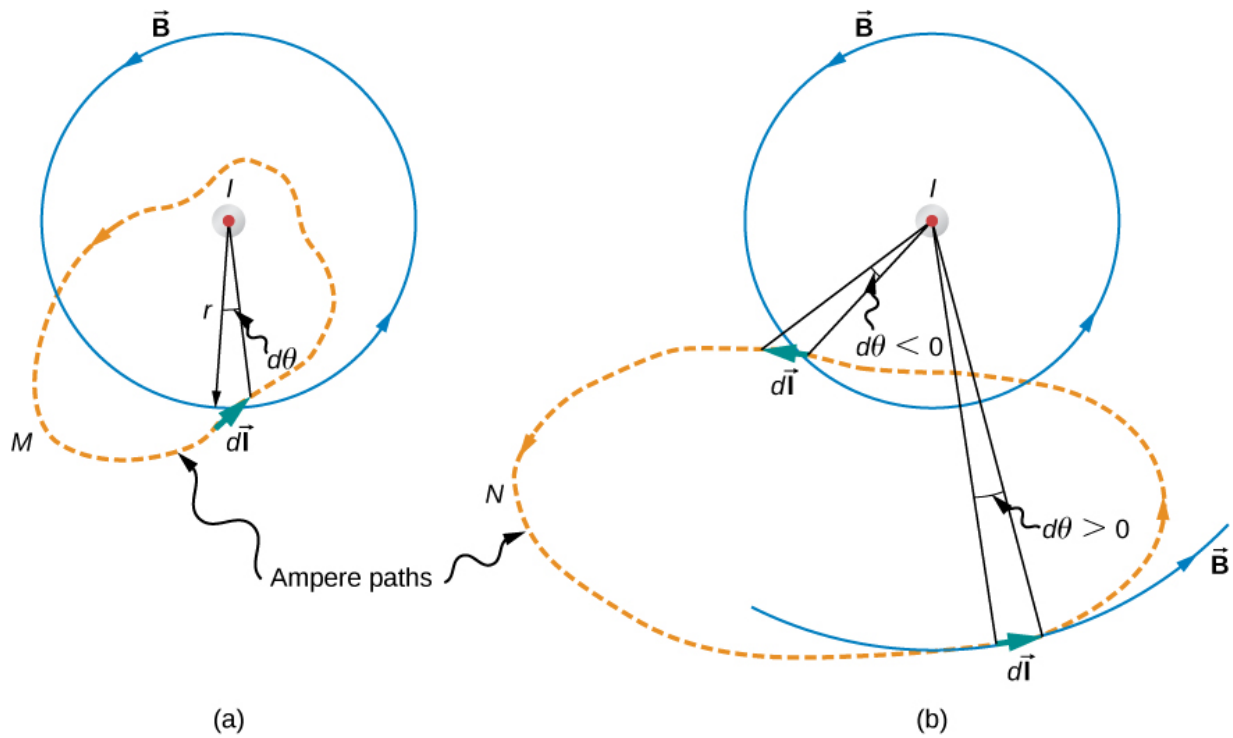
- Explain how Ampère's law relates the magnetic field produced by a current to the value of the current
- Calculate the magnetic field from a long straight wire, either thin or thick, by Ampère's law

A fundamental property of a static magnetic field is that, unlike an electrostatic field, it is not conservative. A conservative field is one that does the same amount of work on a particle moving between two different points regardless of the path chosen. Magnetic fields do not have such a property. Instead, there is a relationship between the magnetic field and its source, electric current. It is expressed in terms of the line integral of \vec{B} and is known as **Ampère's law**. This law can also be derived directly from the Biot-Savart law. We now consider that derivation for the special case of an infinite, straight wire.

[\[link\]](#) shows an arbitrary plane perpendicular to an infinite, straight wire whose current I is directed out of the page. The magnetic field lines are circles directed counterclockwise and centered on the wire. To begin, let's consider $\oint \vec{B} \cdot d\vec{l}$ over the closed paths M and N . Notice that one path (M) encloses the wire, whereas the other (N) does not. Since the field lines are circular, $\vec{B} \cdot d\vec{l}$ is the product of B and the projection of dl onto the circle passing through $d\vec{l}$. If the radius of this particular circle is r , the projection is $r d\theta$, and

Equation:

$$\vec{B} \cdot d\vec{l} = Br d\theta.$$



The current I of a long, straight wire is directed out of the page. The integral $\oint d\theta$ equals 2π and 0 , respectively, for paths M and N .

With \vec{B} given by [\[link\]](#),
Equation:

$$\oint \vec{B} \cdot d\vec{l} = \oint \frac{\mu_0 I}{2\pi r} r d\theta = \frac{\mu_0 I}{2\pi} \oint d\theta.$$

For path M , which circulates around the wire, $\oint_M d\theta = 2\pi$ and

Equation:

$$\oint_M \vec{B} \cdot d\vec{l} = \mu_0 I.$$

Path N , on the other hand, circulates through both positive (counterclockwise) and negative (clockwise) $d\theta$ (see [\[link\]](#)), and since it is closed, $\oint_N d\theta = 0$. Thus for path N ,

Equation:

$$\oint_N \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = 0.$$

The extension of this result to the general case is Ampère's law.

Note:

Ampère's law

Over an arbitrary closed path,

Equation:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = \mu_0 I$$

where I is the total current passing through any open surface S whose perimeter is the path of integration. Only currents inside the path of integration need be considered.

To determine whether a specific current I is positive or negative, curl the fingers of your right hand in the direction of the path of integration, as shown in [\[link\]](#). If I passes through S in the same direction as your extended thumb, I is positive; if I passes through S in the direction opposite to your extended thumb, it is negative.

Note:

Ampère's Law

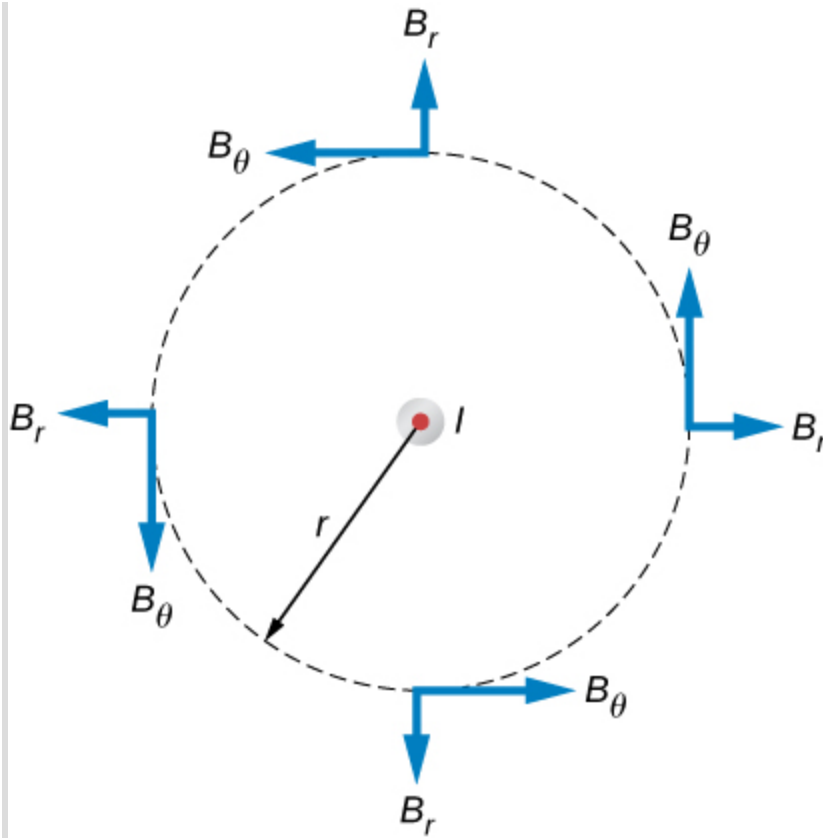
To calculate the magnetic field created from current in wire(s), use the following steps:

1. Identify the symmetry of the current in the wire(s). If there is no symmetry, use the Biot-Savart law to determine the magnetic field.
2. Determine the direction of the magnetic field created by the wire(s) by right-hand rule 2.
3. Chose a path loop where the magnetic field is either constant or zero.
4. Calculate the current inside the loop.
5. Calculate the line integral $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}}$ around the closed loop.
6. Equate $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}}$ with $\mu_0 I_{\text{enc}}$ and solve for $\vec{\mathbf{B}}$.

Example:

Using Ampère's Law to Calculate the Magnetic Field Due to a Wire

Use Ampère's law to calculate the magnetic field due to a steady current I in an infinitely long, thin, straight wire as shown in [\[link\]](#).



The possible components of the magnetic field B due to a current I , which is directed out of the page. The radial component is zero because the angle between the magnetic field and the path is at a right angle.

Strategy

Consider an arbitrary plane perpendicular to the wire, with the current directed out of the page. The possible magnetic field components in this plane, B_r and B_θ , are shown at arbitrary points on a circle of radius r centered on the wire. Since the field is cylindrically symmetric, neither B_r nor B_θ varies with the position on this circle. Also from symmetry, the radial lines, if they exist, must be directed either all inward or all outward from the wire. This means, however, that there must be a net magnetic flux across an arbitrary cylinder concentric with the wire. The radial component

of the magnetic field must be zero because $\vec{\mathbf{B}}_r \cdot d\vec{\mathbf{l}} = 0$. Therefore, we can apply Ampère's law to the circular path as shown.

Solution

Over this path $\vec{\mathbf{B}}$ is constant and parallel to $d\vec{\mathbf{l}}$, so

Equation:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = B_\theta \oint dl = B_\theta(2\pi r).$$

Thus Ampère's law reduces to

Equation:

$$B_\theta(2\pi r) = \mu_0 I.$$

Finally, since B_θ is the only component of $\vec{\mathbf{B}}$, we can drop the subscript and write

Equation:

$$B = \frac{\mu_0 I}{2\pi r}.$$

This agrees with the Biot-Savart calculation above.

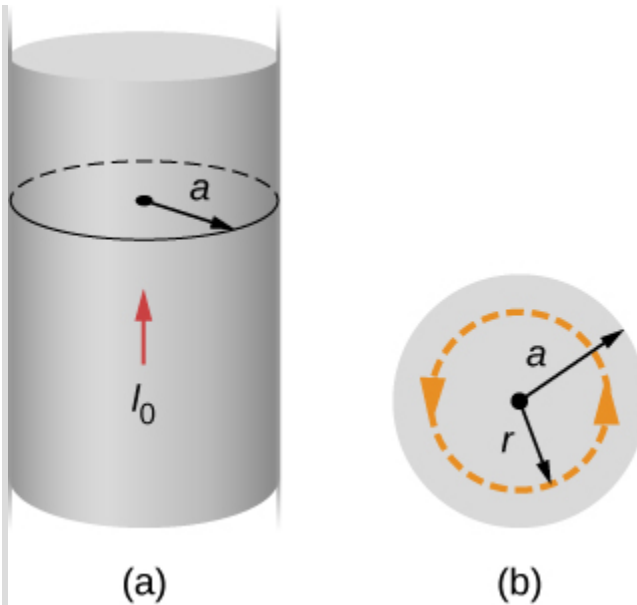
Significance

Ampère's law works well if you have a path to integrate over which $\vec{\mathbf{B}} \cdot d\vec{\mathbf{l}}$ has results that are easy to simplify. For the infinite wire, this works easily with a path that is circular around the wire so that the magnetic field factors out of the integration. If the path dependence looks complicated, you can always go back to the Biot-Savart law and use that to find the magnetic field.

Example:

Calculating the Magnetic Field of a Thick Wire with Ampère's Law

The radius of the long, straight wire of [link](#) is a , and the wire carries a current I_0 that is distributed uniformly over its cross-section. Find the magnetic field both inside and outside the wire.



(a) A model of a current-carrying wire of radius a and current I_0 . (b) A cross-section of the same wire showing the radius a and the Ampère's loop of radius r .

Strategy

This problem has the same geometry as [\[link\]](#), but the enclosed current changes as we move the integration path from outside the wire to inside the wire, where it doesn't capture the entire current enclosed (see [\[link\]](#)).

Solution

For any circular path of radius r that is centered on the wire,

Equation:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = \oint B dl = B \oint dl = B(2\pi r).$$

From Ampère's law, this equals the total current passing through any surface bounded by the path of integration.

Consider first a circular path that is inside the wire ($r \leq a$) such as that shown in part (a) of [\[link\]](#). We need the current I passing through the area enclosed by the path. It's equal to the current density J times the area

enclosed. Since the current is uniform, the current density inside the path equals the current density in the whole wire, which is $I_0/\pi a^2$. Therefore the current I passing through the area enclosed by the path is

Equation:

$$I = \frac{\pi r^2}{\pi a^2} I_0 = \frac{r^2}{a^2} I_0.$$

We can consider this ratio because the current density J is constant over the area of the wire. Therefore, the current density of a part of the wire is equal to the current density in the whole area. Using Ampère's law, we obtain

Equation:

$$B(2\pi r) = \mu_0 \frac{r^2}{a^2} I_0,$$

and the magnetic field inside the wire is

Equation:

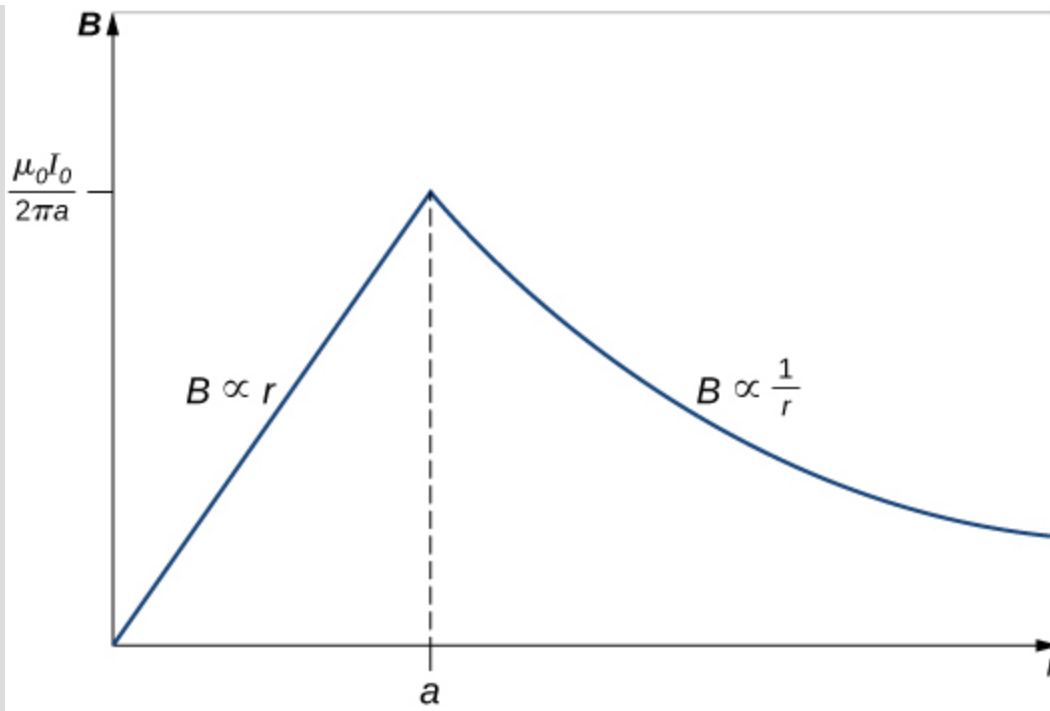
$$B = \frac{\mu_0 I_0}{2\pi} \frac{r}{a^2} \quad (r \leq a).$$

Outside the wire, the situation is identical to that of the infinite thin wire of the previous example; that is,

Equation:

$$B = \frac{\mu_0 I_0}{2\pi r} \quad (r \geq a).$$

The variation of B with r is shown in [\[link\]](#).



Variation of the magnetic field produced by a current I_0 in a long, straight wire of radius a .

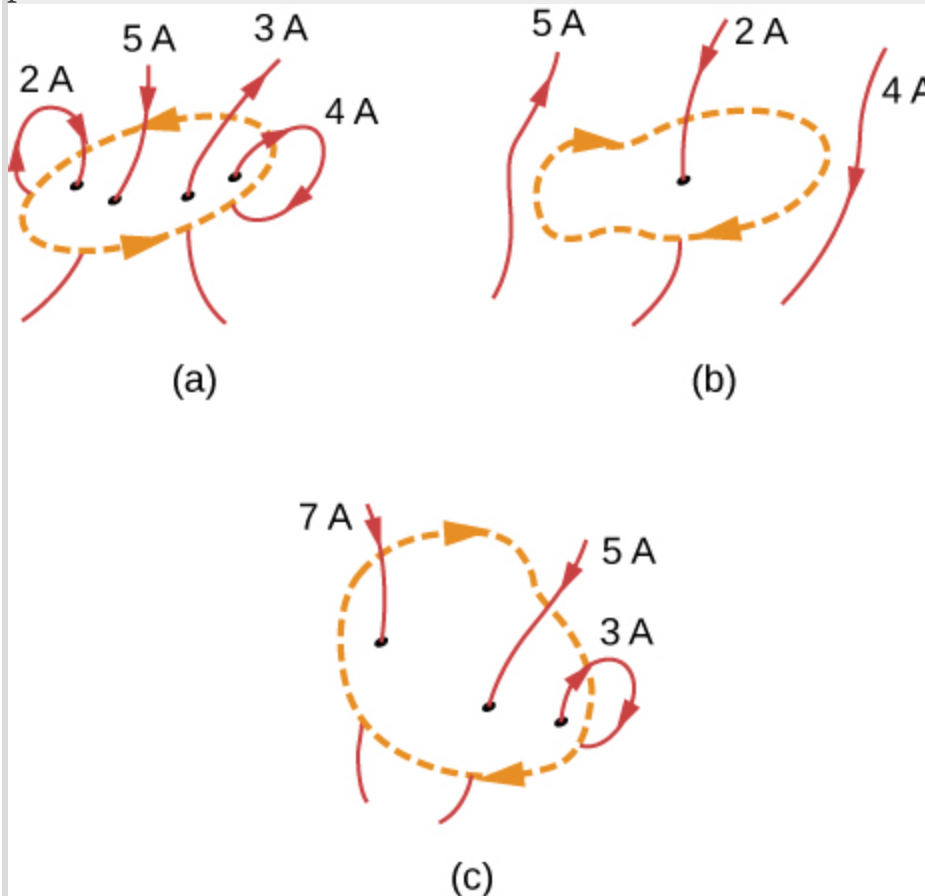
Significance

The results show that as the radial distance increases inside the thick wire, the magnetic field increases from zero to a familiar value of the magnetic field of a thin wire. Outside the wire, the field drops off regardless of whether it was a thick or thin wire.

This result is similar to how Gauss's law for electrical charges behaves inside a uniform charge distribution, except that Gauss's law for electrical charges has a uniform volume distribution of charge, whereas Ampère's law here has a uniform area of current distribution. Also, the drop-off outside the thick wire is similar to how an electric field drops off outside of a linear charge distribution, since the two cases have the same geometry and neither case depends on the configuration of charges or currents once the loop is outside the distribution.

Example:**Using Ampère's Law with Arbitrary Paths**

Use Ampère's law to evaluate $\oint \vec{B} \cdot d\vec{l}$ for the current configurations and paths in [\[link\]](#).



Current configurations and paths for [\[link\]](#).

Strategy

Ampère's law states that $\oint \vec{B} \cdot d\vec{l} = \mu_0 I$ where I is the total current passing through the enclosed loop. The quickest way to evaluate the integral is to calculate $\mu_0 I$ by finding the net current through the loop. Positive currents flow with your right-hand thumb if your fingers wrap around in the direction of the loop. This will tell us the sign of the answer.

Solution

(a) The current going downward through the loop equals the current going out of the loop, so the net current is zero. Thus, $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = 0$.

(b) The only current to consider in this problem is 2A because it is the only current inside the loop. The right-hand rule shows us the current going downward through the loop is in the positive direction. Therefore, the answer is $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = \mu_0(2 \text{ A}) = 2.51 \times 10^{-6} \text{ T} \cdot \text{m}$.

(c) The right-hand rule shows us the current going downward through the loop is in the positive direction. There are $7\text{A} + 5\text{A} = 12\text{A}$ of current going downward and -3 A going upward. Therefore, the total current is 9 A and $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = \mu_0(9 \text{ A}) = 1.13 \times 10^{-5} \text{ T} \cdot \text{m}$.

Significance

If the currents all wrapped around so that the same current went into the loop and out of the loop, the net current would be zero and no magnetic field would be present. This is why wires are very close to each other in an electrical cord. The currents flowing toward a device and away from a device in a wire equal zero total current flow through an Ampère loop around these wires. Therefore, no stray magnetic fields can be present from cords carrying current.

Note:

Exercise:

Problem:

Check Your Understanding Consider using Ampère's law to calculate the magnetic fields of a finite straight wire and of a circular loop of wire. Why is it not useful for these calculations?

Solution:

In these cases the integrals around the Ampèrian loop are very difficult because there is no symmetry, so this method would not be useful.

Summary

- The magnetic field created by current following any path is the sum (or integral) of the fields due to segments along the path (magnitude and direction as for a straight wire), resulting in a general relationship between current and field known as Ampère's law.
- Ampère's law can be used to determine the magnetic field from a thin wire or thick wire by a geometrically convenient path of integration. The results are consistent with the Biot-Savart law.

Conceptual Questions

Exercise:

Problem:

Is Ampère's law valid for all closed paths? Why isn't it normally useful for calculating a magnetic field?

Solution:

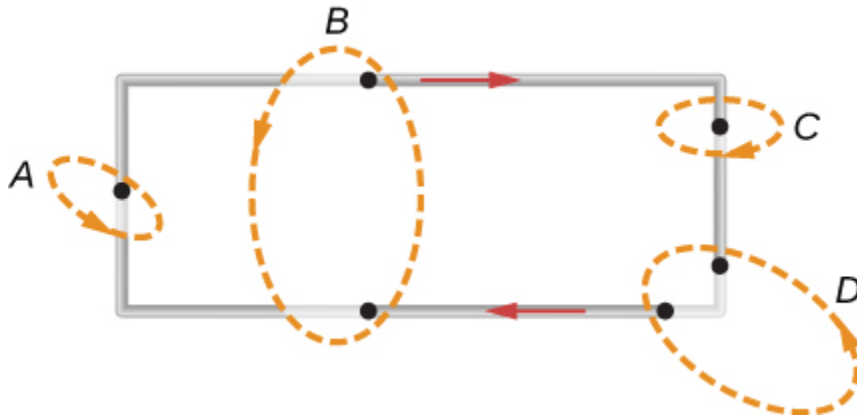
Ampère's law is valid for all closed paths, but it is not useful for calculating fields when the magnetic field produced lacks symmetry that can be exploited by a suitable choice of path.

Problems

Exercise:

Problem:

A current I flows around the rectangular loop shown in the accompanying figure. Evaluate $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}}$ for the paths A , B , C , and D .



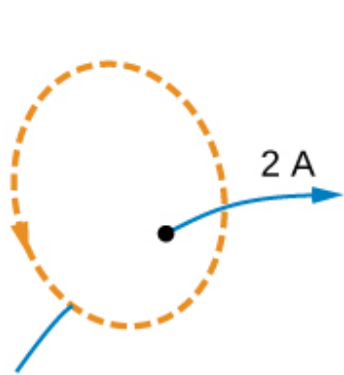
Solution:

a. $\mu_0 I$; b. 0; c. $\mu_0 I$; d. 0

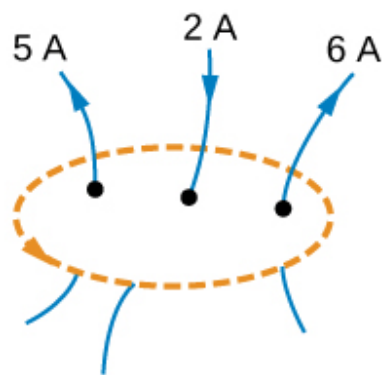
Exercise:

Problem:

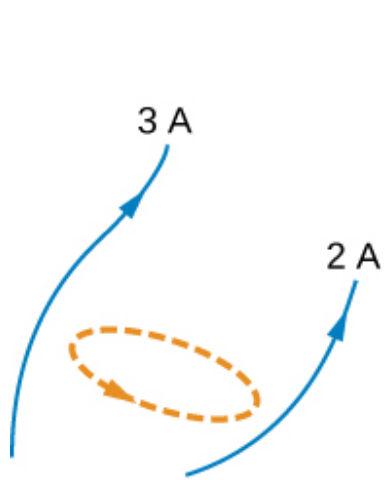
Evaluate $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}}$ for each of the cases shown in the accompanying figure.



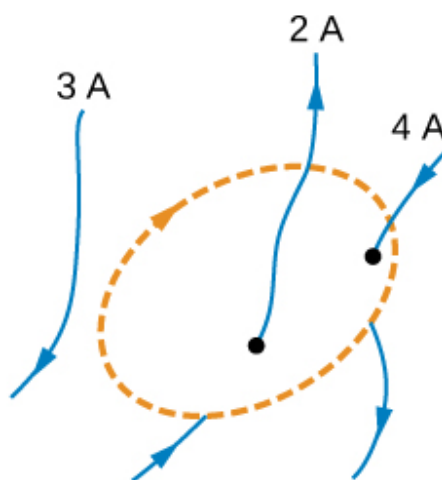
(a)



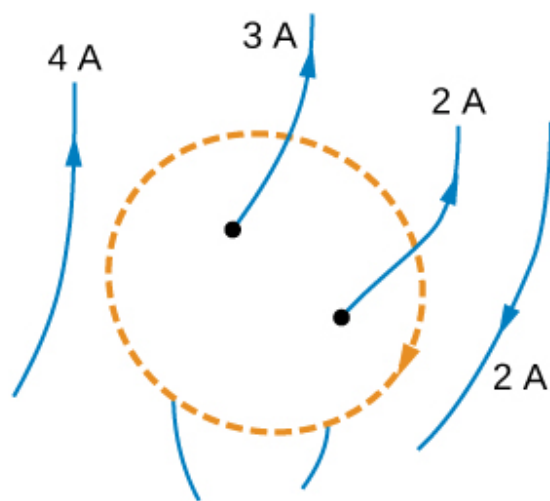
(b)



(c)



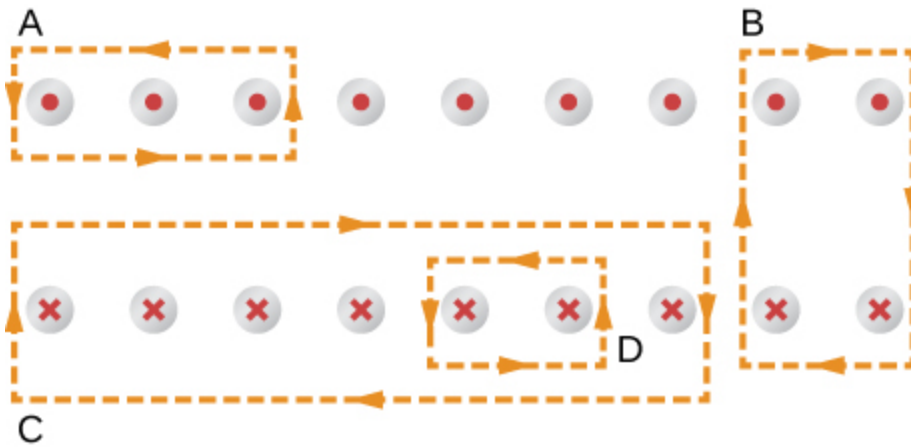
(d)



(e)

Exercise:**Problem:**

The coil whose lengthwise cross section is shown in the accompanying figure carries a current I and has N evenly spaced turns distributed along the length l . Evaluate $\oint \vec{B} \cdot d\vec{l}$ for the paths indicated.

**Solution:**

a. $3\mu_0 I$; b. 0; c. $7\mu_0 I$; d. $-2\mu_0 I$

Exercise:**Problem:**

A superconducting wire of diameter 0.25 cm carries a current of 1000 A. What is the magnetic field just outside the wire?

Exercise:**Problem:**

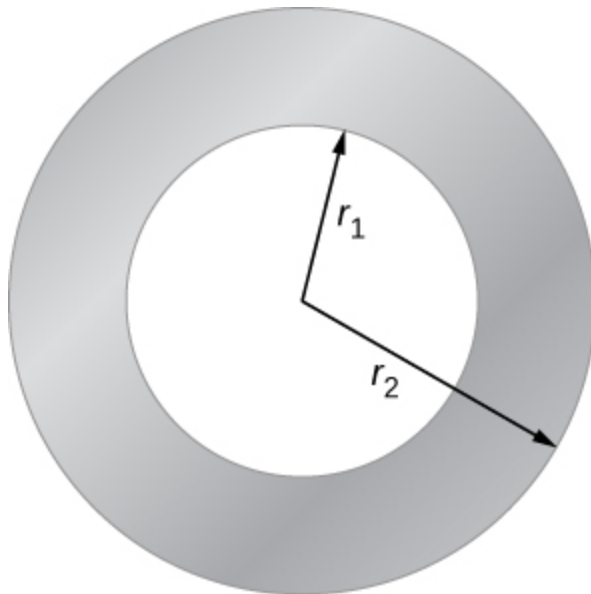
A long, straight wire of radius R carries a current I that is distributed uniformly over the cross-section of the wire. At what distance from the axis of the wire is the magnitude of the magnetic field a maximum?

Solution:

at the radius R

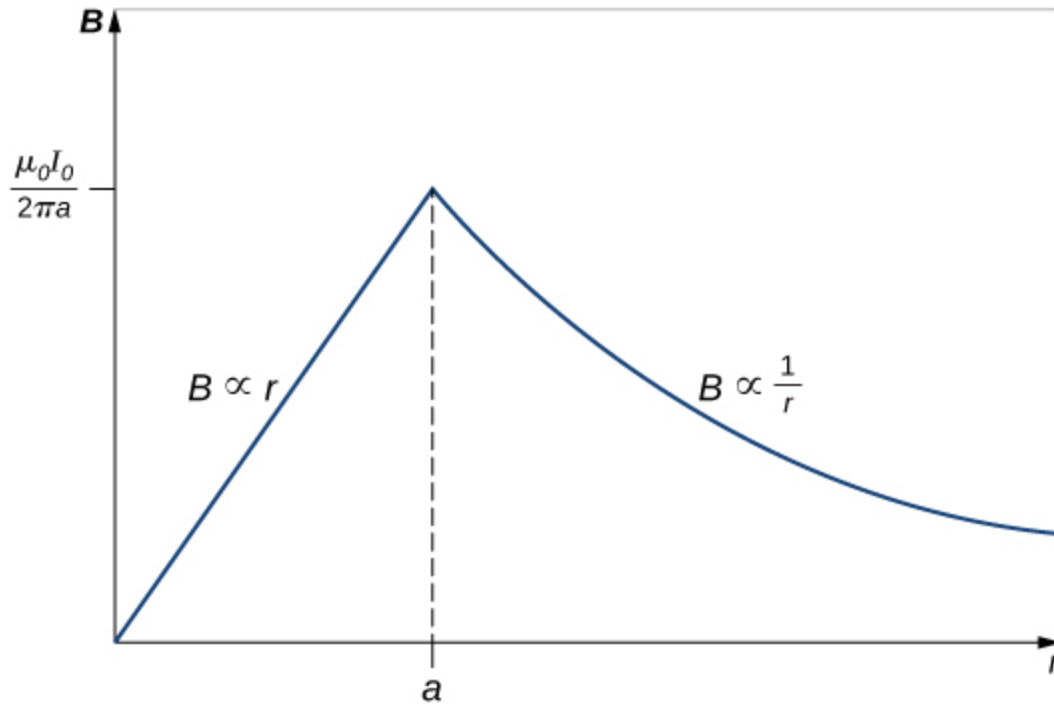
Exercise:**Problem:**

The accompanying figure shows a cross-section of a long, hollow, cylindrical conductor of inner radius $r_1 = 3.0$ cm and outer radius $r_2 = 5.0$ cm. A 50-A current distributed uniformly over the cross-section flows into the page. Calculate the magnetic field at $r = 2.0$ cm, $r = 4.0$ cm, and $r = 6.0$ cm.

**Exercise:****Problem:**

A long, solid, cylindrical conductor of radius 3.0 cm carries a current of 50 A distributed uniformly over its cross-section. Plot the magnetic field as a function of the radial distance r from the center of the conductor.

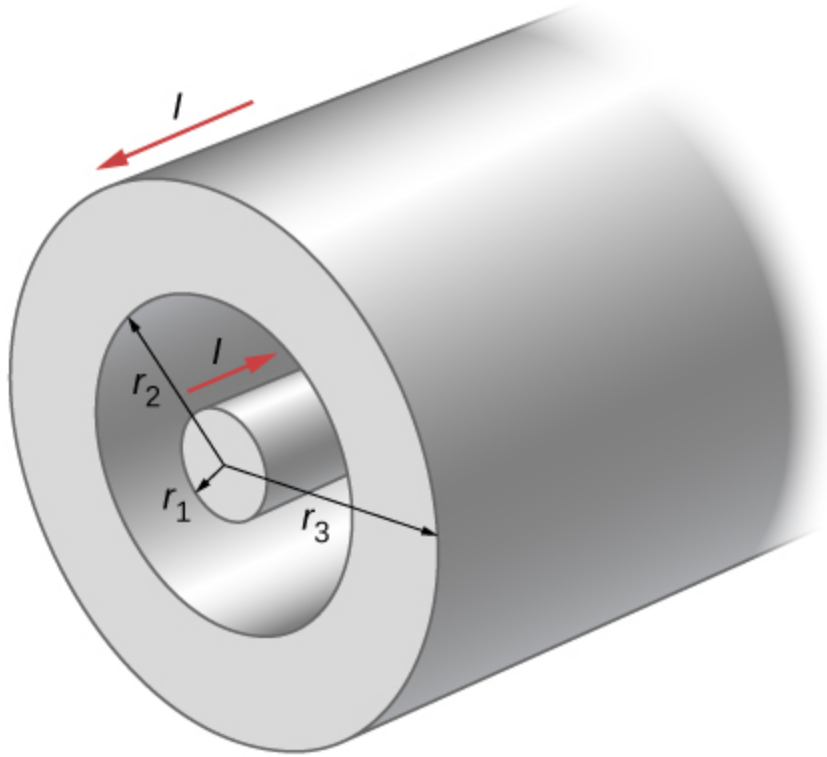
Solution:



Exercise:

Problem:

A portion of a long, cylindrical coaxial cable is shown in the accompanying figure. A current I flows down the center conductor, and this current is returned in the outer conductor. Determine the magnetic field in the regions (a) $r \leq r_1$, (b) $r_2 \geq r \geq r_1$, (c) $r_3 \geq r \geq r_2$, and (d) $r \geq r_3$. Assume that the current is distributed uniformly over the cross sections of the two parts of the cable.



Glossary

Ampère's law

physical law that states that the line integral of the magnetic field around an electric current is proportional to the current

Solenoids and Toroids

By the end of this section, you will be able to:

- Establish a relationship for how the magnetic field of a solenoid varies with distance and current by using both the Biot-Savart law and Ampère's law
- Establish a relationship for how the magnetic field of a toroid varies with distance and current by using Ampère's law

Two of the most common and useful electromagnetic devices are called solenoids and toroids. In one form or another, they are part of numerous instruments, both large and small. In this section, we examine the magnetic field typical of these devices.

Solenoids

A long wire wound in the form of a helical coil is known as a **solenoid**. Solenoids are commonly used in experimental research requiring magnetic fields. A solenoid is generally easy to wind, and near its center, its magnetic field is quite uniform and directly proportional to the current in the wire.

[\[link\]](#) shows a solenoid consisting of N turns of wire tightly wound over a length L . A current I is flowing along the wire of the solenoid. The number of turns per unit length is N/L ; therefore, the number of turns in an infinitesimal length dy are $(N/L)dy$ turns. This produces a current

Equation:

$$dI = \frac{NI}{L} dy.$$

We first calculate the magnetic field at the point P of [\[link\]](#). This point is on the central axis of the solenoid. We are basically cutting the solenoid into thin slices that are dy thick and treating each as a current loop. Thus, dI is the current through each slice. The magnetic field $d\vec{B}$ due to the current dI in dy can be found with the help of [\[link\]](#) and [\[link\]](#):

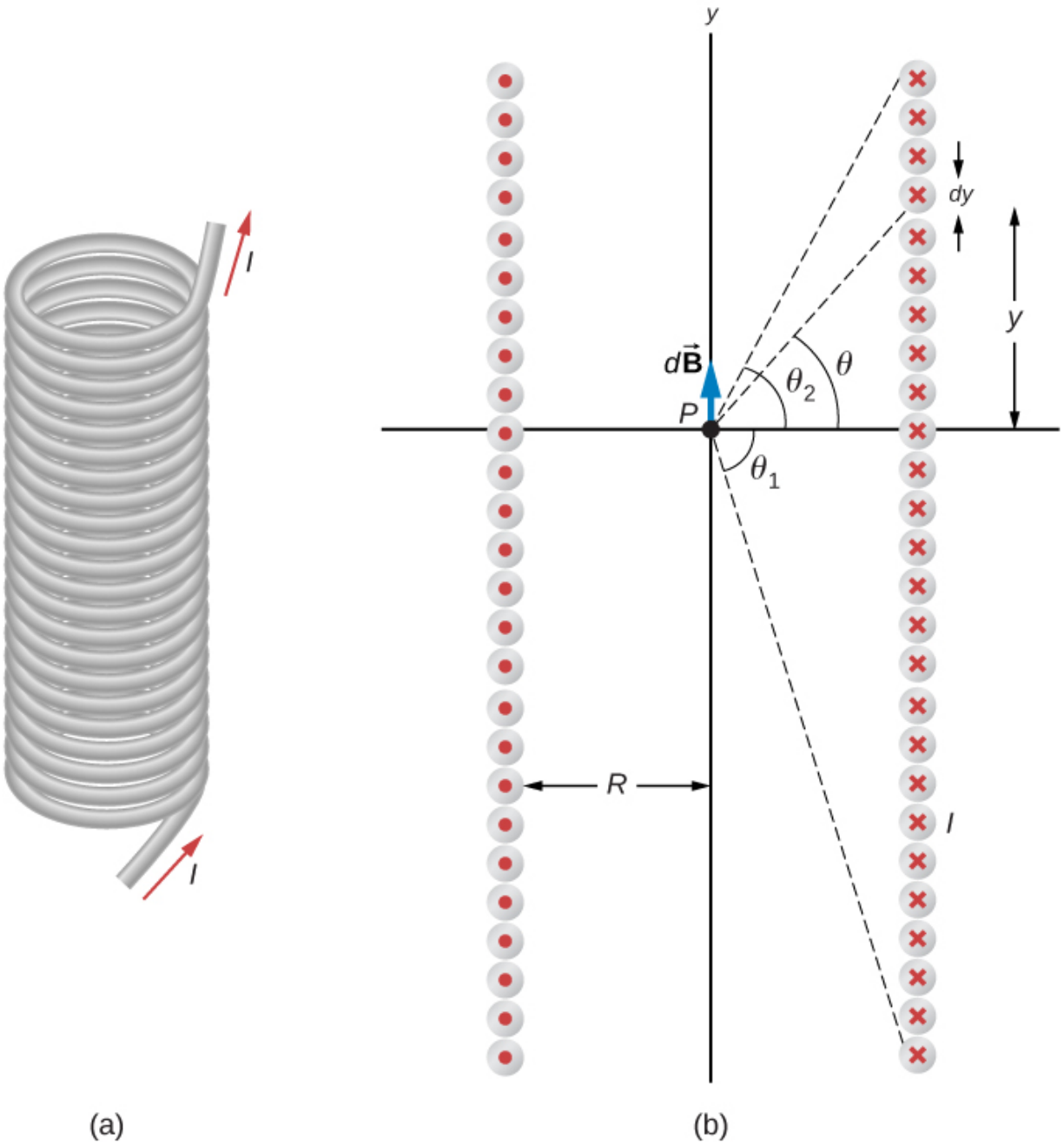
Equation:

$$d\vec{\mathbf{B}} = \frac{\mu_0 R^2 dI}{2(y^2 + R^2)^{3/2}} \hat{\mathbf{j}} = \left(\frac{\mu_0 I R^2 N}{2L} \hat{\mathbf{j}} \right) \frac{dy}{(y^2 + R^2)^{3/2}}$$

where we used [\[link\]](#) to replace dI . The resultant field at P is found by integrating $d\vec{\mathbf{B}}$ along the entire length of the solenoid. It's easiest to evaluate this integral by changing the independent variable from y to θ . From inspection of [\[link\]](#), we have:

Equation:

$$\sin \theta = \frac{y}{\sqrt{y^2 + R^2}}.$$



Taking the differential of both sides of this equation, we obtain

Equation:

$$\begin{aligned}\cos\theta\,d\theta &= \left[-\frac{y^2}{(y^2+R^2)^{3/2}} + \frac{1}{\sqrt{y^2+R^2}} \right] dy \\ &= \frac{R^2 dy}{(y^2+R^2)^{3/2}}.\end{aligned}$$

When this is substituted into the equation for $d\vec{\mathbf{B}}$, we have

Equation:

$$\vec{\mathbf{B}} = \frac{\mu_0 I N}{2L} \hat{\mathbf{j}} \int_{\theta_1}^{\theta_2} \cos\theta\,d\theta = \frac{\mu_0 I N}{2L} (\sin\theta_2 - \sin\theta_1) \hat{\mathbf{j}},$$

which is the magnetic field along the central axis of a finite solenoid.

Of special interest is the infinitely long solenoid, for which $L \rightarrow \infty$. From a practical point of view, the infinite solenoid is one whose length is much larger than its radius ($L \gg R$). In this case, $\theta_1 = -\frac{\pi}{2}$ and $\theta_2 = \frac{\pi}{2}$. Then from [\[link\]](#), the magnetic field along the central axis of an infinite solenoid is

Equation:

$$\vec{\mathbf{B}} = \frac{\mu_0 I N}{2L} \hat{\mathbf{j}} [\sin(\pi/2) - \sin(-\pi/2)] = \frac{\mu_0 I N}{L} \hat{\mathbf{j}}$$

or

Equation:

$$\vec{\mathbf{B}} = \mu_0 n I \hat{\mathbf{j}},$$

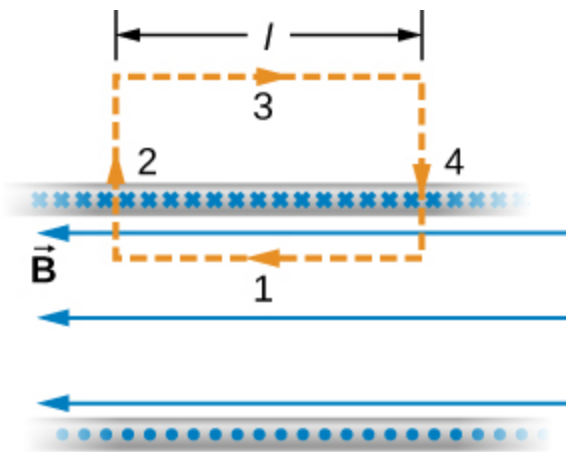
where n is the number of turns per unit length. You can find the direction of $\vec{\mathbf{B}}$ with a right-hand rule: Curl your fingers in the direction of the current,

and your thumb points along the magnetic field in the interior of the solenoid.

We now use these properties, along with Ampère's law, to calculate the magnitude of the magnetic field at any location inside the infinite solenoid. Consider the closed path of [\[link\]](#). Along segment 1, $\vec{\mathbf{B}}$ is uniform and parallel to the path. Along segments 2 and 4, $\vec{\mathbf{B}}$ is perpendicular to part of the path and vanishes over the rest of it. Therefore, segments 2 and 4 do not contribute to the line integral in Ampère's law. Along segment 3, $\vec{\mathbf{B}} = 0$ because the magnetic field is zero outside the solenoid. If you consider an Ampère's law loop outside of the solenoid, the current flows in opposite directions on different segments of wire. Therefore, there is no enclosed current and no magnetic field according to Ampère's law. Thus, there is no contribution to the line integral from segment 3. As a result, we find

Equation:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = \int_1 \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = Bl.$$



The path of integration used in Ampère's law to evaluate the magnetic field of an infinite solenoid.

The solenoid has n turns per unit length, so the current that passes through the surface enclosed by the path is nI . Therefore, from Ampère's law,

Equation:

$$Bl = \mu_0 n I l$$

and

Note:

Equation:

$$B = \mu_0 n I$$

within the solenoid. This agrees with what we found earlier for B on the central axis of the solenoid. Here, however, the location of segment 1 is arbitrary, so we have found that this equation gives the magnetic field everywhere inside the infinite solenoid.

Outside the solenoid, one can draw an Ampère's law loop around the entire solenoid. This would enclose current flowing in both directions. Therefore, the net current inside the loop is zero. According to Ampère's law, if the net current is zero, the magnetic field must be zero. Therefore, for locations outside of the solenoid's radius, the magnetic field is zero.

When a patient undergoes a magnetic resonance imaging (MRI) scan, the person lies down on a table that is moved into the center of a large solenoid that can generate very large magnetic fields. The solenoid is capable of these high fields from high currents flowing through superconducting wires. The large magnetic field is used to change the spin of protons in the patient's body. The time it takes for the spins to align or relax (return to

original orientation) is a signature of different tissues that can be analyzed to see if the structures of the tissues is normal ([\[link\]](#)).



In an MRI machine, a large magnetic field is generated by the cylindrical solenoid surrounding the patient. (credit: Liz West)

Example:

Magnetic Field Inside a Solenoid

A solenoid has 300 turns wound around a cylinder of diameter 1.20 cm and length 14.0 cm. If the current through the coils is 0.410 A, what is the magnitude of the magnetic field inside and near the middle of the solenoid?

Strategy

We are given the number of turns and the length of the solenoid so we can find the number of turns per unit length. Therefore, the magnetic field inside and near the middle of the solenoid is given by [\[link\]](#). Outside the solenoid, the magnetic field is zero.

Solution

The number of turns per unit length is

Equation:

$$n = \frac{300 \text{ turns}}{0.140 \text{ m}} = 2.14 \times 10^3 \text{ turns/m.}$$

The magnetic field produced inside the solenoid is

Equation:

$$B = \mu_0 n I = (4\pi \times 10^{-7} \text{ T} \cdot \text{m/A})(2.14 \times 10^3 \text{ turns/m})(0.410 \text{ A})$$

$$B = 1.10 \times 10^{-3} \text{ T.}$$

Significance

This solution is valid only if the length of the solenoid is reasonably large compared with its diameter. This example is a case where this is valid.

Note:**Exercise:****Problem:**

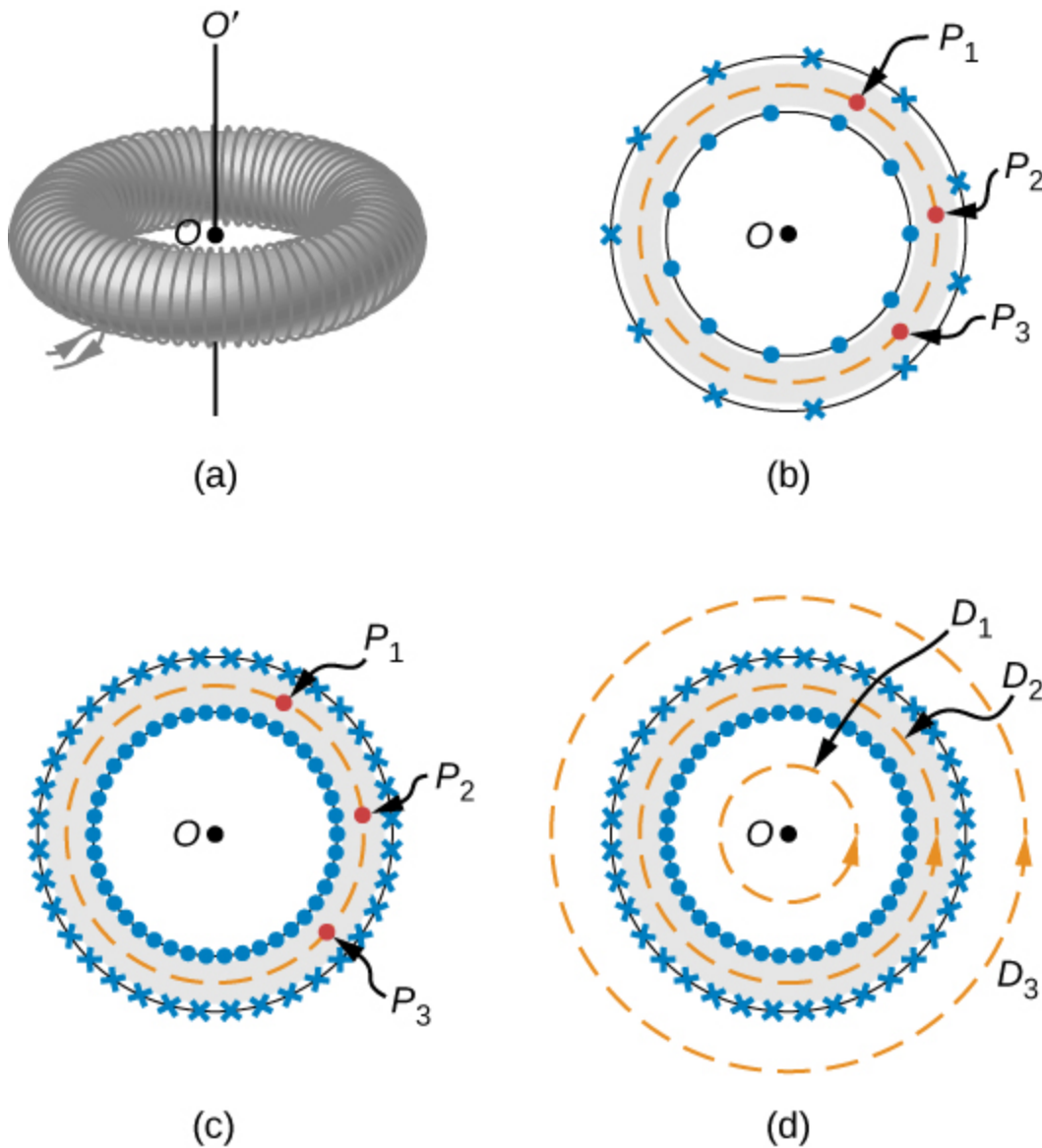
Check Your Understanding What is the ratio of the magnetic field produced from using a finite formula over the infinite approximation for an angle θ of (a) 85° ? (b) 89° ? The solenoid has 1000 turns in 50 cm with a current of 1.0 A flowing through the coils

Solution:

a. 1.00382; b. 1.00015

Toroids

A toroid is a donut-shaped coil closely wound with one continuous wire, as illustrated in part (a) of [\[link\]](#). If the toroid has N windings and the current in the wire is I , what is the magnetic field both inside and outside the toroid?



(a) A toroid is a coil wound into a donut-shaped object. (b) A loosely wound toroid does not have cylindrical symmetry. (c) In a tightly wound toroid, cylindrical symmetry is a very

good approximation. (d) Several paths of integration for Ampère's law.

We begin by assuming cylindrical symmetry around the axis OO' . Actually, this assumption is not precisely correct, for as part (b) of [\[link\]](#) shows, the view of the toroidal coil varies from point to point (for example, P_1 , P_2 , and P_3) on a circular path centered around OO' . However, if the toroid is tightly wound, all points on the circle become essentially equivalent [part (c) of [\[link\]](#)], and cylindrical symmetry is an accurate approximation.

With this symmetry, the magnetic field must be tangent to and constant in magnitude along any circular path centered on OO' . This allows us to write for each of the paths D_1 , D_2 , and D_3 shown in part (d) of [\[link\]](#),

Equation:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}} = B(2\pi r).$$

Ampère's law relates this integral to the net current passing through any surface bounded by the path of integration. For a path that is external to the toroid, either no current passes through the enclosing surface (path D_1), or the current passing through the surface in one direction is exactly balanced by the current passing through it in the opposite direction (path D_3). In either case, there is no net current passing through the surface, so

Equation:

$$\oint B(2\pi r) = 0$$

and

Equation:

$$B = 0 \quad (\text{outside the toroid}).$$

The turns of a toroid form a helix, rather than circular loops. As a result, there is a small field external to the coil; however, the derivation above holds if the coils were circular.

For a circular path within the toroid (path D_2), the current in the wire cuts the surface N times, resulting in a net current NI through the surface. We now find with Ampère's law,

Equation:

$$B(2\pi r) = \mu_0 NI$$

and

Note:

Equation:

$$B = \frac{\mu_0 NI}{2\pi r} \quad (\text{within the toroid}).$$

The magnetic field is directed in the counterclockwise direction for the windings shown. When the current in the coils is reversed, the direction of the magnetic field also reverses.

The magnetic field inside a toroid is not uniform, as it varies inversely with the distance r from the axis OO' . However, if the central radius R (the radius midway between the inner and outer radii of the toroid) is much larger than the cross-sectional diameter of the coils r , the variation is fairly small, and the magnitude of the magnetic field may be calculated by [\[link\]](#) where $r = R$.

Summary

- The magnetic field strength inside a solenoid is

Equation:

$$B = \mu_0 n I \quad (\text{inside a solenoid})$$

where n is the number of loops per unit length of the solenoid. The field inside is very uniform in magnitude and direction.

- The magnetic field strength inside a toroid is

Equation:

$$B = \frac{\mu_0 N I}{2\pi r} \quad (\text{within the toroid})$$

where N is the number of windings. The field inside a toroid is not uniform and varies with the distance as $1/r$.

Conceptual Questions

Exercise:

Problem:

Is the magnetic field inside a toroid completely uniform? Almost uniform?

Exercise:

Problem:

Explain why $\vec{B} = 0$ inside a long, hollow copper pipe that is carrying an electric current parallel to the axis. Is $\vec{B} = 0$ outside the pipe?

Solution:

If there is no current inside the loop, there is no magnetic field (see Ampère's law). Outside the pipe, there may be an enclosed current through the copper pipe, so the magnetic field may not be zero outside the pipe.

Problems

Exercise:

Problem:

A solenoid is wound with 2000 turns per meter. When the current is 5.2 A, what is the magnetic field within the solenoid?

Solution:

$$B = 1.3 \times 10^{-2} \text{T}$$

Exercise:

Problem:

A solenoid has 12 turns per centimeter. What current will produce a magnetic field of $2.0 \times 10^{-2} \text{T}$ within the solenoid?

Exercise:

Problem:

If a current is 2.0 A, how many turns per centimeter must be wound on a solenoid in order to produce a magnetic field of $2.0 \times 10^{-3} \text{T}$ within it?

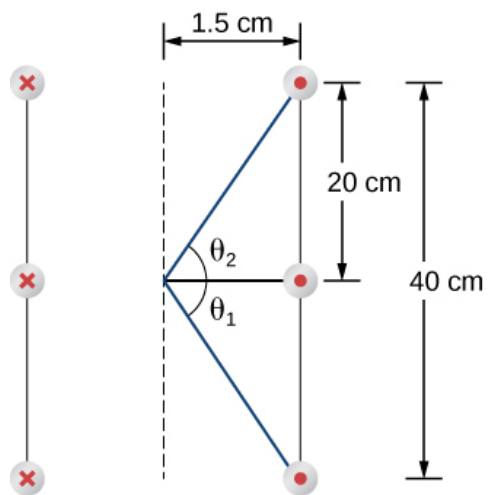
Solution:

roughly eight turns per cm

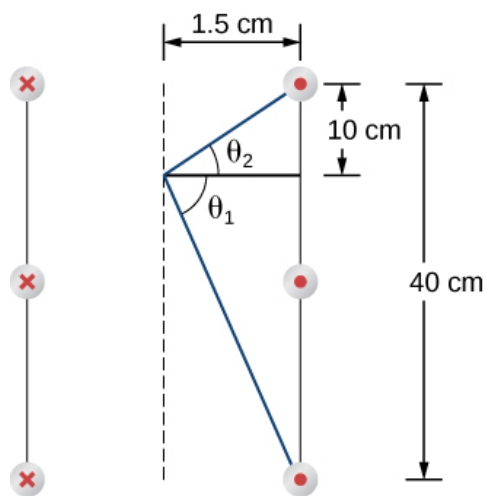
Exercise:

Problem:

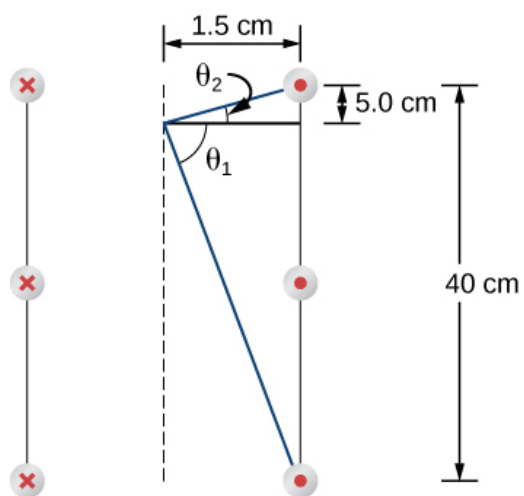
A solenoid is 40 cm long, has a diameter of 3.0 cm, and is wound with 500 turns. If the current through the windings is 4.0 A, what is the magnetic field at a point on the axis of the solenoid that is (a) at the center of the solenoid, (b) 10.0 cm from one end of the solenoid, and (c) 5.0 cm from one end of the solenoid? (d) Compare these answers with the infinite-solenoid case.



(a)



(b)



(c)

Exercise:**Problem:**

Determine the magnetic field on the central axis at the opening of a semi-infinite solenoid. (That is, take the opening to be at $x = 0$ and the other end to be at $x = \infty$.)

Solution:

$$B = \frac{1}{2}\mu_0 nI$$

Exercise:**Problem:**

By how much is the approximation $B = \mu_0 nI$ in error at the center of a solenoid that is 15.0 cm long, has a diameter of 4.0 cm, is wrapped with n turns per meter, and carries a current I ?

Exercise:**Problem:**

A solenoid with 25 turns per centimeter carries a current I . An electron moves within the solenoid in a circle that has a radius of 2.0 cm and is perpendicular to the axis of the solenoid. If the speed of the electron is 2.0×10^5 m/s, what is I ?

Solution:

$$0.0181 \text{ A}$$

Exercise:**Problem:**

A toroid has 250 turns of wire and carries a current of 20 A. Its inner and outer radii are 8.0 and 9.0 cm. What are the values of its magnetic field at $r = 8.1, 8.5$, and 8.9 cm?

Exercise:**Problem:**

A toroid with a square cross section $3.0\text{ cm} \times 3.0\text{ cm}$ has an inner radius of 25.0 cm . It is wound with 500 turns of wire, and it carries a current of 2.0 A . What is the strength of the magnetic field at the center of the square cross section?

Solution:

0.0008 T

Glossary

solenoid

thin wire wound into a coil that produces a magnetic field when an electric current is passed through it

toroid

donut-shaped coil closely wound around that is one continuous wire

Magnetism in Matter

By the end of this section, you will be able to:

- Classify magnetic materials as paramagnetic, diamagnetic, or ferromagnetic, based on their response to a magnetic field
- Sketch how magnetic dipoles align with the magnetic field in each type of substance
- Define hysteresis and magnetic susceptibility, which determines the type of magnetic material

Why are certain materials magnetic and others not? And why do certain substances become magnetized by a field, whereas others are unaffected? To answer such questions, we need an understanding of magnetism on a microscopic level.

Within an atom, every electron travels in an orbit and spins on an internal axis. Both types of motion produce current loops and therefore magnetic dipoles. For a particular atom, the net magnetic dipole moment is the vector sum of the magnetic dipole moments. Values of μ for several types of atoms are given in [\[link\]](#). Notice that some atoms have a zero net dipole moment and that the magnitudes of the nonvanishing moments are typically $10^{-23} \text{ A} \cdot \text{m}^2$.

Atom	Magnetic Moment ($10^{-24} \text{ A} \cdot \text{m}^2$)
H	9.27
He	0
Li	9.27

Atom	Magnetic Moment ($10^{-24} \text{ A} \cdot \text{m}^2$)
O	13.9
Na	9.27
S	13.9

Magnetic Moments of Some Atoms

A handful of matter has approximately 10^{26} atoms and ions, each with its magnetic dipole moment. If no external magnetic field is present, the magnetic dipoles are randomly oriented—as many are pointed up as down, as many are pointed east as west, and so on. Consequently, the net magnetic dipole moment of the sample is zero. However, if the sample is placed in a magnetic field, these dipoles tend to align with the field (see [\[link\]](#)), and this alignment determines how the sample responds to the field. On the basis of this response, a material is said to be either paramagnetic, ferromagnetic, or diamagnetic.

In a **paramagnetic material**, only a small fraction (roughly one-third) of the magnetic dipoles are aligned with the applied field. Since each dipole produces its own magnetic field, this alignment contributes an extra magnetic field, which enhances the applied field. When a **ferromagnetic material** is placed in a magnetic field, its magnetic dipoles also become aligned; furthermore, they become locked together so that a permanent magnetization results, even when the field is turned off or reversed. This permanent magnetization happens in ferromagnetic materials but not paramagnetic materials. **Diamagnetic materials** are composed of atoms that have no net magnetic dipole moment. However, when a diamagnetic material is placed in a magnetic field, a magnetic dipole moment is directed opposite to the applied field and therefore produces a magnetic field that opposes the applied field. We now consider each type of material in greater detail.

Paramagnetic Materials

For simplicity, we assume our sample is a long, cylindrical piece that completely fills the interior of a long, tightly wound solenoid. When there is no current in the solenoid, the magnetic dipoles in the sample are randomly oriented and produce no net magnetic field. With a solenoid current, the magnetic field due to the solenoid exerts a torque on the dipoles that tends to align them with the field. In competition with the aligning torque are thermal collisions that tend to randomize the orientations of the dipoles. The relative importance of these two competing processes can be estimated by comparing the energies involved. From [\[link\]](#), the energy difference between a magnetic dipole aligned with and against a magnetic field is $U_B = 2\mu B$. If $\mu = 9.3 \times 10^{-24} \text{ A} \cdot \text{m}^2$ (the value of atomic hydrogen) and $B = 1.0 \text{ T}$, then

Equation:

$$U_B = 1.9 \times 10^{-23} \text{ J}.$$

At a room temperature of 27°C , the thermal energy per atom is

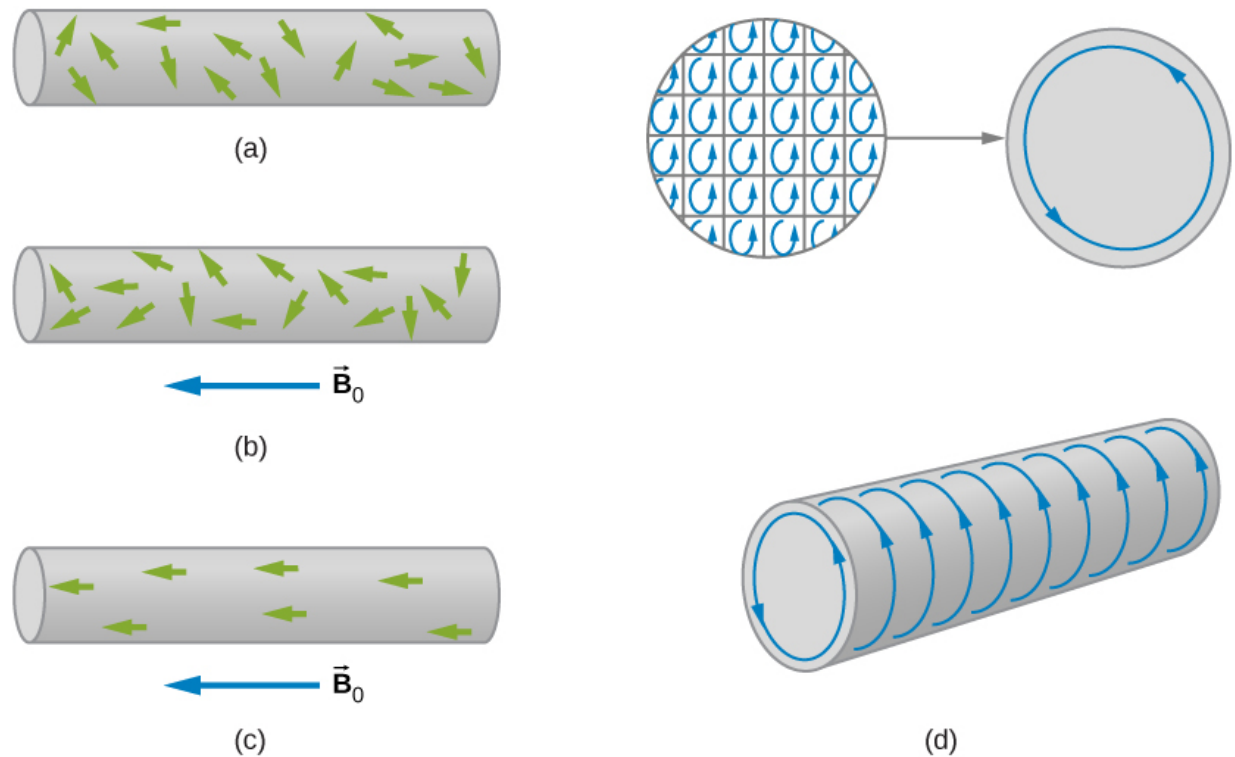
Equation:

$$U_T \approx kT = (1.38 \times 10^{-23} \text{ J/K})(300 \text{ K}) = 4.1 \times 10^{-21} \text{ J},$$

which is about 220 times greater than U_B . Clearly, energy exchanges in thermal collisions can seriously interfere with the alignment of the magnetic dipoles. As a result, only a small fraction of the dipoles is aligned at any instant.

The four sketches of [\[link\]](#) furnish a simple model of this alignment process. In part (a), before the field of the solenoid (not shown) containing the paramagnetic sample is applied, the magnetic dipoles are randomly oriented and there is no net magnetic dipole moment associated with the material. With the introduction of the field, a partial alignment of the dipoles takes place, as depicted in part (b). The component of the net magnetic dipole moment that is perpendicular to the field vanishes. We may then represent the sample by part (c), which shows a collection of magnetic dipoles completely aligned with the field. By treating these dipoles as current loops, we can picture the dipole alignment as equivalent to a current around the

surface of the material, as in part (d). This fictitious surface current produces its own magnetic field, which enhances the field of the solenoid.



The alignment process in a paramagnetic material filling a solenoid (not shown). (a) Without an applied field, the magnetic dipoles are randomly oriented. (b) With a field, partial alignment occurs. (c) An equivalent representation of part (b). (d) The internal currents cancel, leaving an effective surface current that produces a magnetic field similar to that of a finite solenoid.

We can express the total magnetic field \vec{B} in the material as
Equation:

$$\vec{B} = \vec{B}_0 + \vec{B}_m,$$

where $\vec{\mathbf{B}}_0$ is the field due to the current I_0 in the solenoid and $\vec{\mathbf{B}}_m$ is the field due to the surface current I_m around the sample. Now $\vec{\mathbf{B}}_m$ is usually proportional to $\vec{\mathbf{B}}_0$, a fact we express by

Equation:

$$\vec{\mathbf{B}}_m = \chi \vec{\mathbf{B}}_0,$$

where χ is a dimensionless quantity called the **magnetic susceptibility**.

Values of χ for some paramagnetic materials are given in [\[link\]](#). Since the alignment of magnetic dipoles is so weak, χ is very small for paramagnetic materials. By combining [\[link\]](#) and [\[link\]](#), we obtain:

Equation:

$$\vec{\mathbf{B}} = \vec{\mathbf{B}}_0 + \chi \vec{\mathbf{B}}_0 = (1 + \chi) \vec{\mathbf{B}}_0.$$

For a sample within an infinite solenoid, this becomes

Equation:

$$B = (1 + \chi) \mu_0 n I.$$

This expression tells us that the insertion of a paramagnetic material into a solenoid increases the field by a factor of $(1 + \chi)$. However, since χ is so small, the field isn't enhanced very much.

The quantity

Note:

Equation:

$$\mu = (1 + \chi) \mu_0.$$

is called the magnetic permeability of a material. In terms of μ , [\[link\]](#) can be written as

Note:
Equation:

$$B = \mu nI$$

for the filled solenoid.

Paramagnetic Materials	χ	Diamagnetic Materials	χ
Aluminum	2.2×10^{-5}	Bismuth	-1.7×10^{-5}
Calcium	1.4×10^{-5}	Carbon (diamond)	-2.2×10^{-5}
Chromium	3.1×10^{-4}	Copper	-9.7×10^{-6}
Magnesium	1.2×10^{-5}	Lead	-1.8×10^{-5}
Oxygen gas (1 atm)	1.8×10^{-6}	Mercury	-2.8×10^{-5}
Oxygen liquid (90 K)	3.5×10^{-3}	Hydrogen gas (1 atm)	-2.2×10^{-9}

Paramagnetic Materials	χ	Diamagnetic Materials	χ
Tungsten	6.8×10^{-5}	Nitrogen gas (1 atm)	-6.7×10^{-9}
Air (1 atm)	3.6×10^{-7}	Water	-9.1×10^{-6}

Magnetic Susceptibilities*Note: Unless otherwise specified, values given are for room temperature.

Diamagnetic Materials

A magnetic field always induces a magnetic dipole in an atom. This induced dipole points opposite to the applied field, so its magnetic field is also directed opposite to the applied field. In paramagnetic and ferromagnetic materials, the induced magnetic dipole is masked by much stronger permanent magnetic dipoles of the atoms. However, in diamagnetic materials, whose atoms have no permanent magnetic dipole moments, the effect of the induced dipole is observable.

We can now describe the magnetic effects of diamagnetic materials with the same model developed for paramagnetic materials. In this case, however, the fictitious surface current flows opposite to the solenoid current, and the magnetic susceptibility χ is negative. Values of χ for some diamagnetic materials are also given in [\[link\]](#).

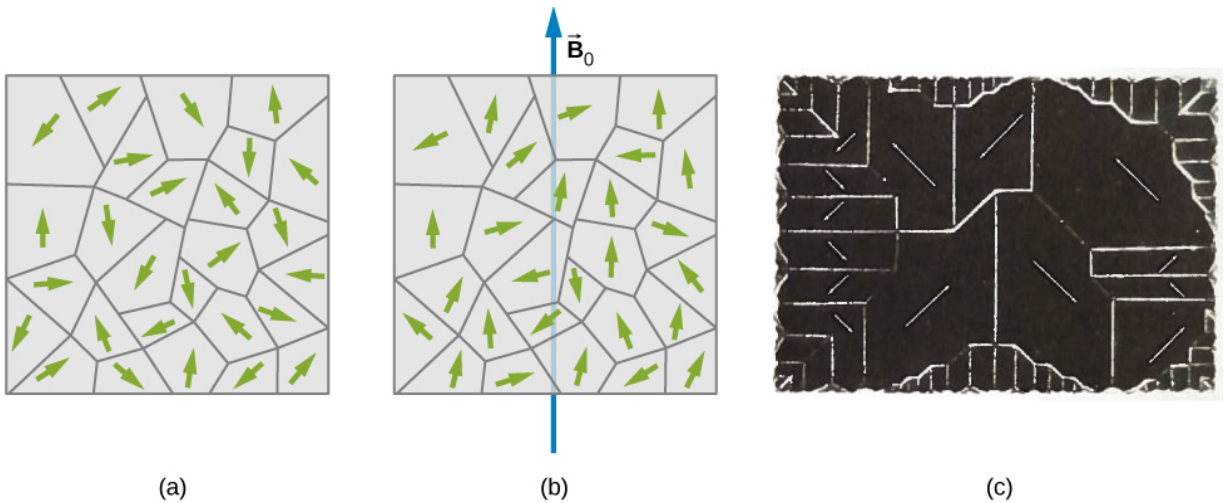
Note:

Water is a common diamagnetic material. Animals are mostly composed of water. Experiments have been performed on [frogs](#) and [mice](#) in diverging magnetic fields. The water molecules are repelled from the applied magnetic field against gravity until the animal reaches an equilibrium. The result is that the animal is levitated by the magnetic field.

Ferromagnetic Materials

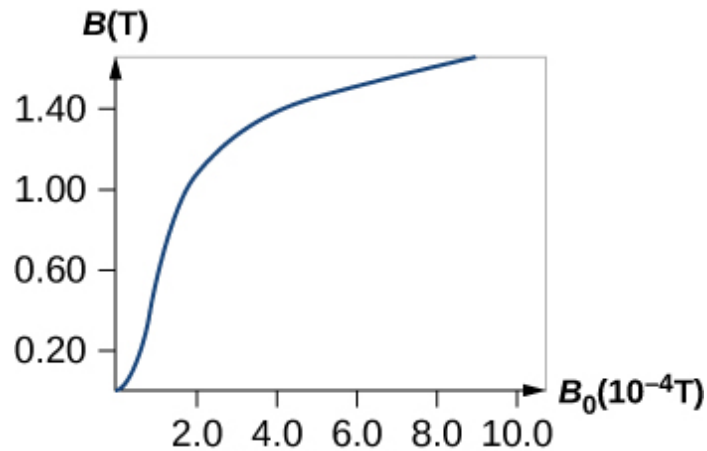
Common magnets are made of a ferromagnetic material such as iron or one of its alloys. Experiments reveal that a ferromagnetic material consists of tiny regions known as **magnetic domains**. Their volumes typically range from 10^{-12} to 10^{-8}m^3 , and they contain about 10^{17} to 10^{21} atoms. Within a domain, the magnetic dipoles are rigidly aligned in the same direction by coupling among the atoms. This coupling, which is due to quantum mechanical effects, is so strong that even thermal agitation at room temperature cannot break it. The result is that each domain has a net dipole moment. Some materials have weaker coupling and are ferromagnetic only at lower temperatures.

If the domains in a ferromagnetic sample are randomly oriented, as shown in [\[link\]](#), the sample has no net magnetic dipole moment and is said to be unmagnetized. Suppose that we fill the volume of a solenoid with an unmagnetized ferromagnetic sample. When the magnetic field \vec{B}_0 of the solenoid is turned on, the dipole moments of the domains rotate so that they align somewhat with the field, as depicted in [\[link\]](#). In addition, the aligned domains tend to increase in size at the expense of unaligned ones. The net effect of these two processes is the creation of a net magnetic dipole moment for the ferromagnet that is directed along the applied magnetic field. This net magnetic dipole moment is much larger than that of a paramagnetic sample, and the domains, with their large numbers of atoms, do not become misaligned by thermal agitation. Consequently, the field due to the alignment of the domains is quite large.



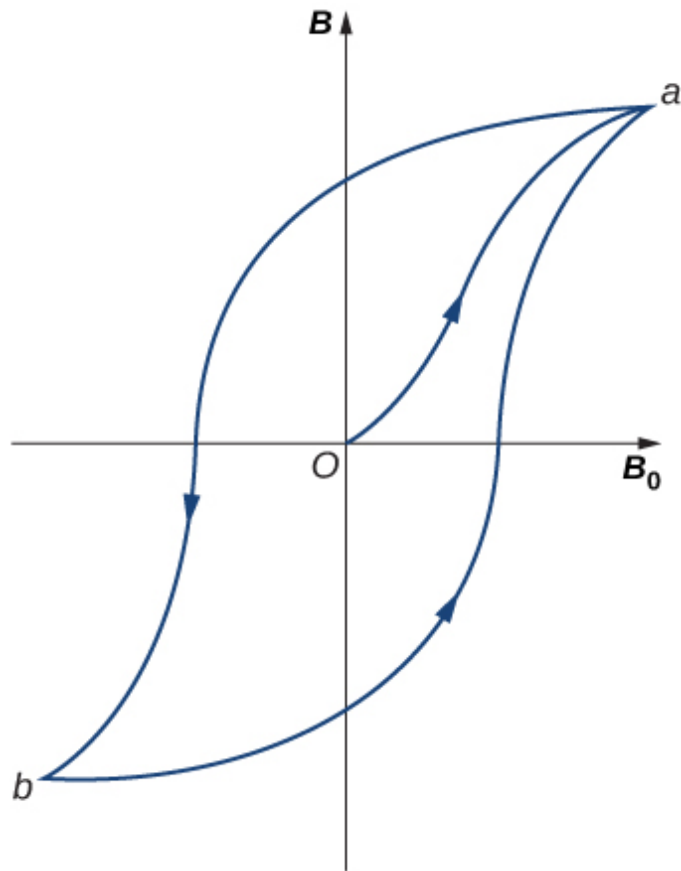
(a) Domains are randomly oriented in an unmagnetized ferromagnetic sample such as iron. The arrows represent the orientations of the magnetic dipoles within the domains. (b) In an applied magnetic field, the domains align somewhat with the field. (c) The domains of a single crystal of nickel. The white lines show the boundaries of the domains. These lines are produced by iron oxide powder sprinkled on the crystal.

Besides iron, only four elements contain the magnetic domains needed to exhibit ferromagnetic behavior: cobalt, nickel, gadolinium, and dysprosium. Many alloys of these elements are also ferromagnetic. Ferromagnetic materials can be described using [\[link\]](#) through [\[link\]](#), the paramagnetic equations. However, the value of χ for ferromagnetic material is usually on the order of 10^3 to 10^4 , and it also depends on the history of the magnetic field to which the material has been subject. A typical plot of B (the total field in the material) versus B_0 (the applied field) for an initially unmagnetized piece of iron is shown in [\[link\]](#). Some sample numbers are (1) for $B_0 = 1.0 \times 10^{-4} \text{ T}$, $B = 0.60 \text{ T}$, and $\chi = (0.60 / 1.0 \times 10^{-4}) - 1 \approx 6.0 \times 10^3$; (2) for $B_0 = 6.0 \times 10^{-4} \text{ T}$, $B = 1.5 \text{ T}$, and $\chi = (1.5 / 6.0 \times 10^{-4}) - 1 \approx 2.5 \times 10^3$.



(a) The magnetic field B in annealed iron as a function of the applied field B_0 .

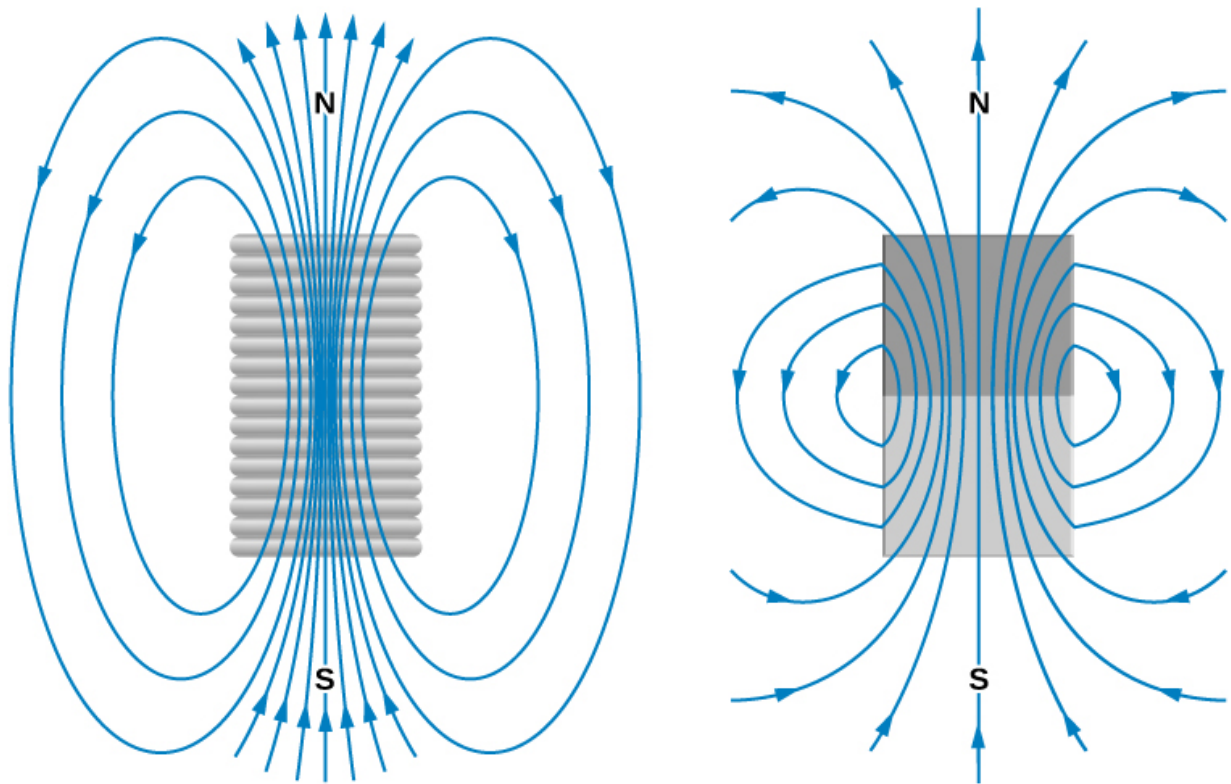
When B_0 is varied over a range of positive and negative values, B is found to behave as shown in [\[link\]](#). Note that the same B_0 (corresponding to the same current in the solenoid) can produce different values of B in the material. The magnetic field B produced in a ferromagnetic material by an applied field B_0 depends on the magnetic history of the material. This effect is called **hysteresis**, and the curve of [\[link\]](#) is called a hysteresis loop. Notice that B does not disappear when $B_0 = 0$ (i.e., when the current in the solenoid is turned off). The iron stays magnetized, which means that it has become a permanent magnet.



A typical hysteresis loop for a ferromagnet. When the material is first magnetized, it follows a curve from 0 to a . When B_0 is reversed, it takes the path shown from a to b . If B_0 is reversed again, the material follows the curve from b to a .

Like the paramagnetic sample of [\[link\]](#), the partial alignment of the domains in a ferromagnet is equivalent to a current flowing around the surface. A bar magnet can therefore be pictured as a tightly wound solenoid with a large current circulating through its coils (the surface current). You can see in [\[link\]](#) that this model fits quite well. The fields of the bar magnet and the finite solenoid are strikingly similar. The figure also shows how the poles of the bar magnet are identified. To form closed loops, the field lines outside

the magnet leave the north (N) pole and enter the south (S) pole, whereas inside the magnet, they leave S and enter N.



Comparison of the magnetic fields of a finite solenoid and a bar magnet.

Ferromagnetic materials are found in computer hard disk drives and permanent data storage devices ([\[link\]](#)). A material used in your hard disk drives is called a spin valve, which has alternating layers of ferromagnetic (aligning with the external magnetic field) and antiferromagnetic (each atom is aligned opposite to the next) metals. It was observed that a significant change in resistance was discovered based on whether an applied magnetic field was on the spin valve or not. This large change in resistance creates a quick and consistent way for recording or reading information by an applied current.



The inside of a hard disk drive. The silver disk contains the information, whereas the thin stylus on top of the disk reads and writes information to the disk.

Example:**Iron Core in a Coil**

A long coil is tightly wound around an iron cylinder whose magnetization curve is shown in [\[link\]](#). (a) If $n = 20$ turns per centimeter, what is the applied field B_0 when $I_0 = 0.20$ A? (b) What is the net magnetic field for this same current? (c) What is the magnetic susceptibility in this case?

Strategy

(a) The magnetic field of a solenoid is calculated using [\[link\]](#). (b) The graph is read to determine the net magnetic field for this same current. (c) The magnetic susceptibility is calculated using [\[link\]](#).

Solution

a. The applied field B_0 of the coil is

Equation:

$$B_0 = \mu_0 n I_0 = (4\pi \times 10^{-7} \text{ T} \cdot \text{m/A})(2000/\text{m})(0.20 \text{ A})$$

$$B_0 = 5.0 \times 10^{-4} \text{ T}.$$

b. From inspection of the magnetization curve of [\[link\]](#), we see that, for this value of B_0 , $B = 1.4 \text{ T}$. Notice that the internal field of the aligned atoms is much larger than the externally applied field.

c. The magnetic susceptibility is calculated to be

Equation:

$$\chi = \frac{B}{B_0} - 1 = \frac{1.4 \text{ T}}{5.0 \times 10^{-4} \text{ T}} - 1 = 2.8 \times 10^3.$$

Significance

Ferromagnetic materials have susceptibilities in the range of 10^3 which compares well to our results here. Paramagnetic materials have fractional susceptibilities, so their applied field of the coil is much greater than the magnetic field generated by the material.

Note:

Exercise:

Problem:

Check Your Understanding Repeat the calculations from the previous example for $I_0 = 0.040 \text{ A}$.

Solution:

a. $1.0 \times 10^{-4} \text{ T}$; b. 0.60 T ; c. 6.0×10^3

Summary

- Materials are classified as paramagnetic, diamagnetic, or ferromagnetic, depending on how they behave in an applied magnetic field.
- Paramagnetic materials have partial alignment of their magnetic dipoles with an applied magnetic field. This is a positive magnetic susceptibility. Only a surface current remains, creating a solenoid-like magnetic field.
- Diamagnetic materials exhibit induced dipoles opposite to an applied magnetic field. This is a negative magnetic susceptibility.
- Ferromagnetic materials have groups of dipoles, called domains, which align with the applied magnetic field. However, when the field is removed, the ferromagnetic material remains magnetized, unlike paramagnetic materials. This magnetization of the material versus the applied field effect is called hysteresis.

Key Equations

Permeability of free space	$\mu_0 = 4\pi \times 10^{-7} \text{T} \cdot \text{m/A}$
Contribution to magnetic field from a current element	$dB = \frac{\mu_0}{4\pi} \frac{I dl \sin \theta}{r^2}$
Biot–Savart law	$\vec{B} = \frac{\mu_0}{4\pi} \int_{\text{wire}} \frac{I d\vec{l} \times \hat{r}}{r^2}$
Magnetic field due to a long straight wire	$B = \frac{\mu_0 I}{2\pi R}$
Force between two parallel	

currents	$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}$
Magnetic field of a current loop	$B = \frac{\mu_0 I}{2R}$ (at center of loop)
Ampère's law	$\oint \vec{B} \cdot d\vec{l} = \mu_0 I$
Magnetic field strength inside a solenoid	$B = \mu_0 n I$
Magnetic field strength inside a toroid	$B = \frac{\mu_0 N I}{2\pi r}$
Magnetic permeability	$\mu = (1 + \chi)\mu_0$
Magnetic field of a solenoid filled with paramagnetic material	$B = \mu n I$

Conceptual Questions

Exercise:

Problem:

A diamagnetic material is brought close to a permanent magnet. What happens to the material?

Exercise:

Problem:

If you cut a bar magnet into two pieces, will you end up with one magnet with an isolated north pole and another magnet with an isolated south pole? Explain your answer.

Solution:

The bar magnet will then become two magnets, each with their own north and south poles. There are no magnetic monopoles or single pole magnets.

Problems

Exercise:

Problem:

The magnetic field in the core of an air-filled solenoid is 1.50 T. By how much will this magnetic field decrease if the air is pumped out of the core while the current is held constant?

Exercise:

Problem:

A solenoid has a ferromagnetic core, $n = 1000$ turns per meter, and $I = 5.0$ A. If B inside the solenoid is 2.0 T, what is χ for the core material?

Solution:

317.31

Exercise:

Problem:

A 20-A current flows through a solenoid with 2000 turns per meter. What is the magnetic field inside the solenoid if its core is (a) a vacuum and (b) filled with liquid oxygen at 90 K?

Exercise:

Problem:

The magnetic dipole moment of the iron atom is about $2.1 \times 10^{-23} \text{ A} \cdot \text{m}^2$. (a) Calculate the maximum magnetic dipole moment of a domain consisting of 10^{19} iron atoms. (b) What current would have to flow through a single circular loop of wire of diameter 1.0 cm to produce this magnetic dipole moment?

Solution:

$$2.1 \times 10^{-4} \text{ A} \cdot \text{m}^2$$
$$2.7 \text{ A}$$

Exercise:**Problem:**

Suppose you wish to produce a 1.2-T magnetic field in a toroid with an iron core for which $\chi = 4.0 \times 10^3$. The toroid has a mean radius of 15 cm and is wound with 500 turns. What current is required?

Exercise:**Problem:**

A current of 1.5 A flows through the windings of a large, thin toroid with 200 turns per meter and a radius of 1 meter. If the toroid is filled with iron for which $\chi = 3.0 \times 10^3$, what is the magnetic field within it?

Solution:

$$0.18 \text{ T}$$

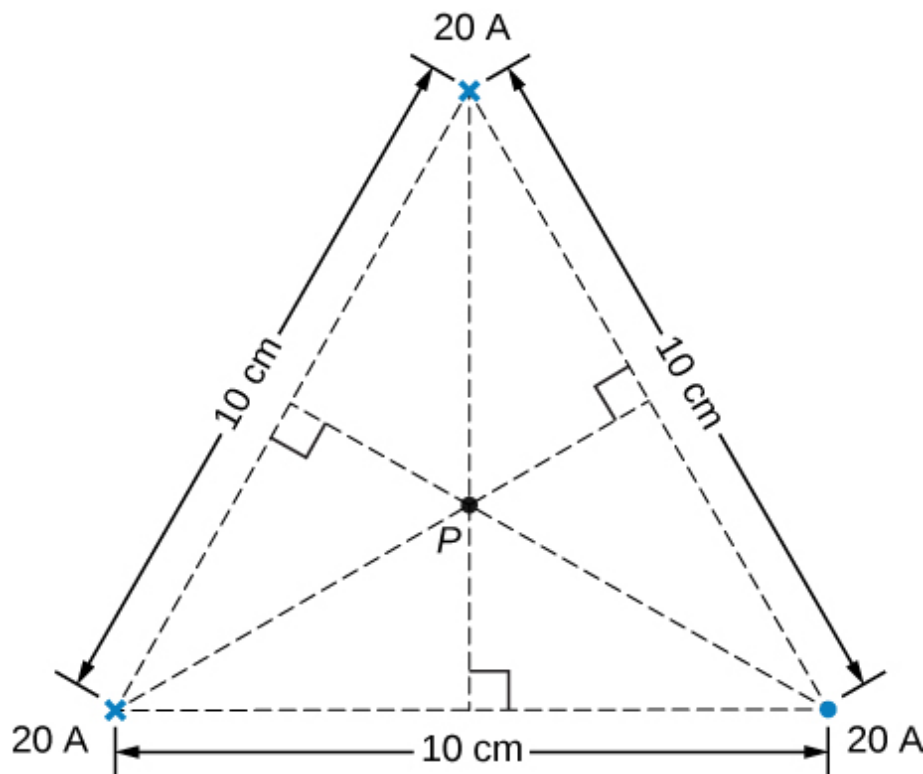
Exercise:

Problem:

A solenoid with an iron core is 25 cm long and is wrapped with 100 turns of wire. When the current through the solenoid is 10 A, the magnetic field inside it is 2.0 T. For this current, what is the permeability of the iron? If the current is turned off and then restored to 10 A, will the magnetic field necessarily return to 2.0 T?

Additional Problems**Exercise:****Problem:**

Three long, straight, parallel wires, all carrying 20 A, are positioned as shown in the accompanying figure. What is the magnitude of the magnetic field at the point P ?

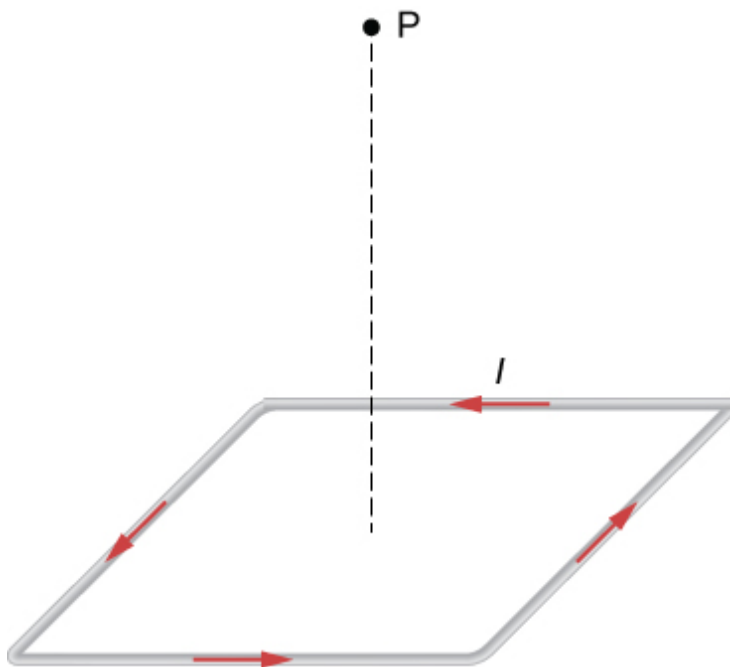


Solution:

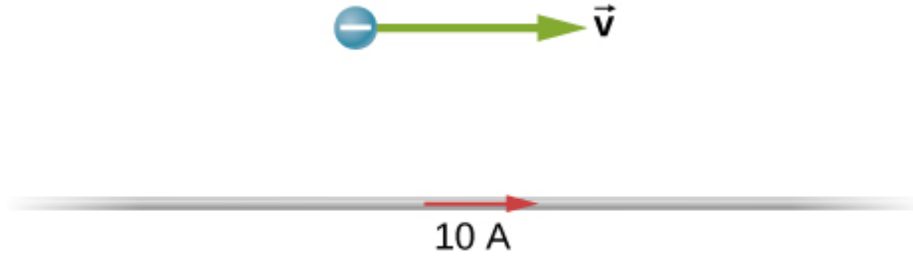
$$B = 1.4 \times 10^{-4} \text{T}$$

Exercise:**Problem:**

A current I flows around a wire bent into the shape of a square of side a . What is the magnetic field at the point P that is a distance z above the center of the square (see the accompanying figure)?

**Exercise:****Problem:**

The accompanying figure shows a long, straight wire carrying a current of 10 A. What is the magnetic force on an electron at the instant it is 20 cm from the wire, traveling parallel to the wire with a speed of $2.0 \times 10^5 \text{ m/s}$? Describe qualitatively the subsequent motion of the electron.



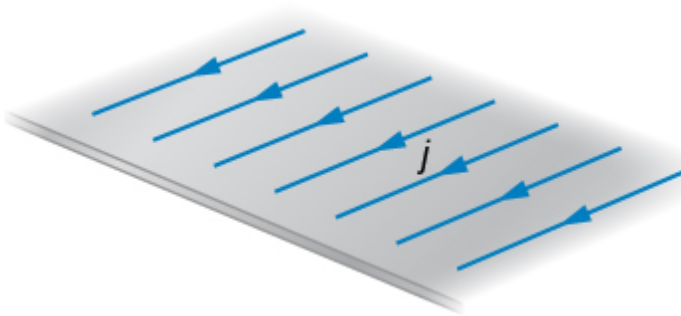
Solution:

$3.2 \times 10^{-19} \text{ N}$ in an arc away from the wire

Exercise:

Problem:

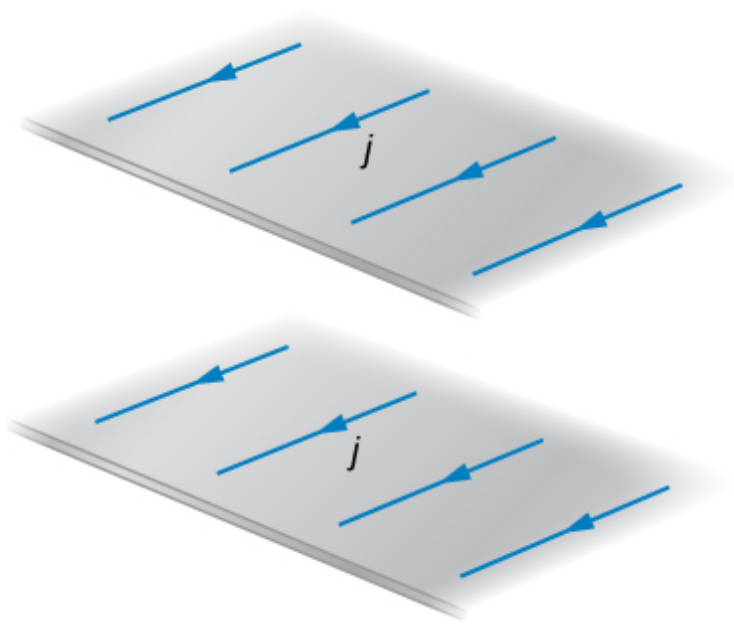
Current flows along a thin, infinite sheet as shown in the accompanying figure. The current per unit length along the sheet is J in amperes per meter. (a) Use the Biot-Savart law to show that $B = \mu_0 J/2$ on either side of the sheet. What is the direction of \vec{B} on each side? (b) Now use Ampère's law to calculate the field.



Exercise:

Problem:

(a) Use the result of the previous problem to calculate the magnetic field between, above, and below the pair of infinite sheets shown in the accompanying figure. (b) Repeat your calculations if the direction of the current in the lower sheet is reversed.

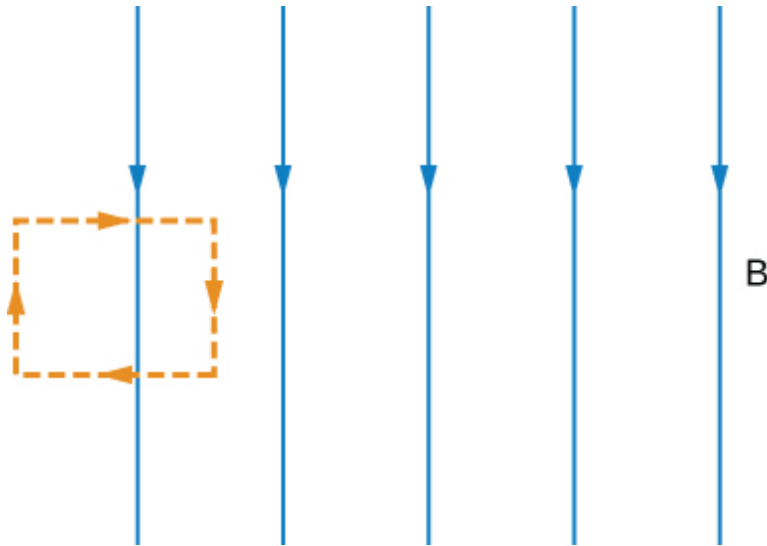


Solution:

a. above and below $B = \mu_0 j$, in the middle $B = 0$; b. above and below $B = 0$, in the middle $B = \mu_0 j$

Exercise:**Problem:**

We often assume that the magnetic field is uniform in a region and zero everywhere else. Show that in reality it is impossible for a magnetic field to drop abruptly to zero, as illustrated in the accompanying figure. (*Hint: Apply Ampère's law over the path shown.*)



Exercise:

Problem:

How is the fractional change in the strength of the magnetic field across the face of the toroid related to the fractional change in the radial distance from the axis of the toroid?

Solution:

$$\frac{dB}{B} = - \frac{dr}{r}$$

Exercise:

Problem:

Show that the expression for the magnetic field of a toroid reduces to that for the field of an infinite solenoid in the limit that the central radius goes to infinity.

Exercise:

Problem:

A toroid with an inner radius of 20 cm and an outer radius of 22 cm is tightly wound with one layer of wire that has a diameter of 0.25 mm.

(a) How many turns are there on the toroid? (b) If the current through the toroid windings is 2.0 A, what is the strength of the magnetic field at the center of the toroid?

Solution:

a. 5026 turns; b. 0.00957 T

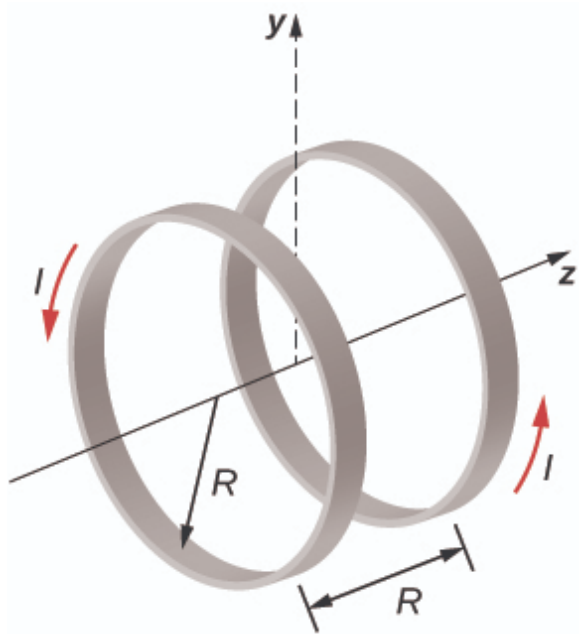
Exercise:**Problem:**

A wire element has $d\vec{l}$, $I d\vec{l} = \mathbf{J} A dl = \mathbf{J} dv$, where A and dv are the cross-sectional area and volume of the element, respectively. Use this, the Biot-Savart law, and $\mathbf{J} = ne\mathbf{v}$ to show that the magnetic field of a moving point charge q is given by:

$$\vec{B} = \frac{\mu_0}{4\pi} \frac{q\mathbf{v} \times \hat{r}}{r^2}$$

Exercise:**Problem:**

A reasonably uniform magnetic field over a limited region of space can be produced with the Helmholtz coil, which consists of two parallel coils centered on the same axis. The coils are connected so that they carry the same current I . Each coil has N turns and radius R , which is also the distance between the coils. (a) Find the magnetic field at any point on the z -axis shown in the accompanying figure. (b) Show that dB/dz and d^2B/dz^2 are both zero at $z = 0$. (These vanishing derivatives demonstrate that the magnetic field varies only slightly near $z = 0$.)



Solution:

$$B_1(x) = \frac{\mu_0 I R^2}{2(R^2 + z^2)^{3/2}}$$

Exercise:

Problem:

A charge of $4.0 \mu\text{C}$ is distributed uniformly around a thin ring of insulating material. The ring has a radius of 0.20 m and rotates at $2.0 \times 10^4 \text{ rev/min}$ around the axis that passes through its center and is perpendicular to the plane of the ring. What is the magnetic field at the center of the ring?

Exercise:

Problem:

A thin, nonconducting disk of radius R is free to rotate around the axis that passes through its center and is perpendicular to the face of the disk. The disk is charged uniformly with a total charge q . If the disk rotates at a constant angular velocity ω , what is the magnetic field at its center?

Solution:

$$B = \frac{\mu_0 \sigma \omega}{2} R$$

Exercise:

Problem:

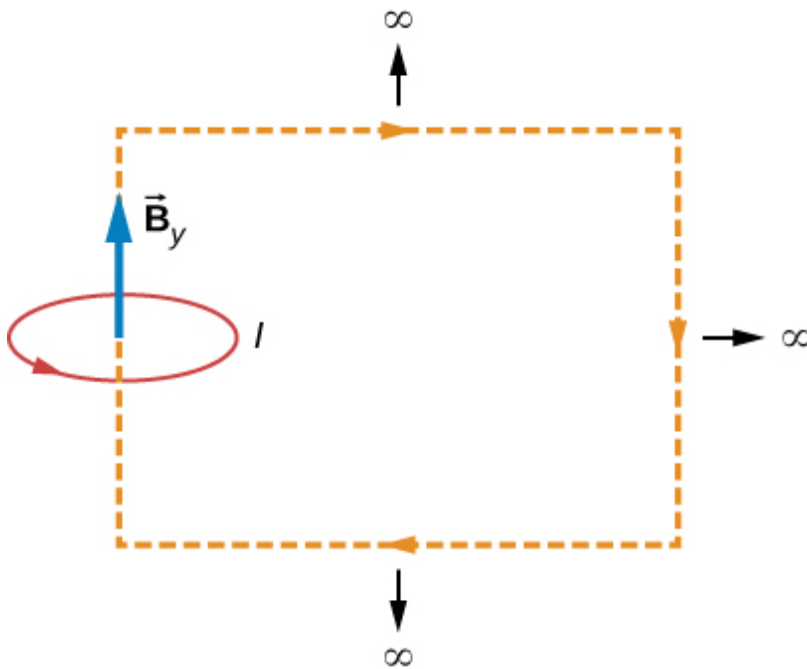
Consider the disk in the previous problem. Calculate the magnetic field at a point on its central axis that is a distance y above the disk.

Exercise:

Problem:

Consider the axial magnetic field $B_y = \mu_0 I R^2 / 2(y^2 + R^2)^{3/2}$ of the circular current loop shown below. (a) Evaluate $\int_{-a}^a B_y dy$. Also show

that $\lim_{a \rightarrow \infty} \int_{-a}^a B_y dy = \mu_0 I$. (b) Can you deduce this limit without evaluating the integral? (*Hint: See the accompanying figure.*)

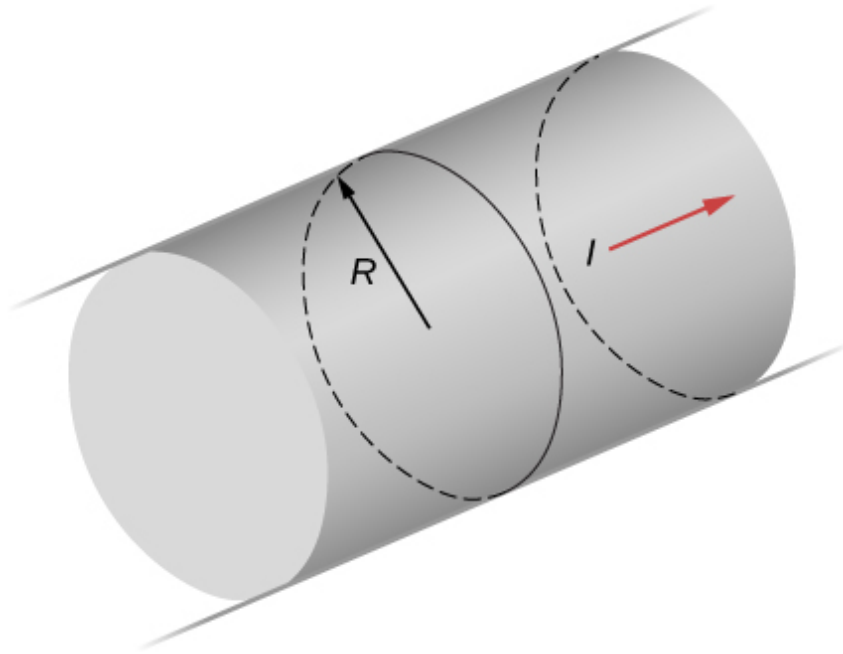


Solution:

derivation

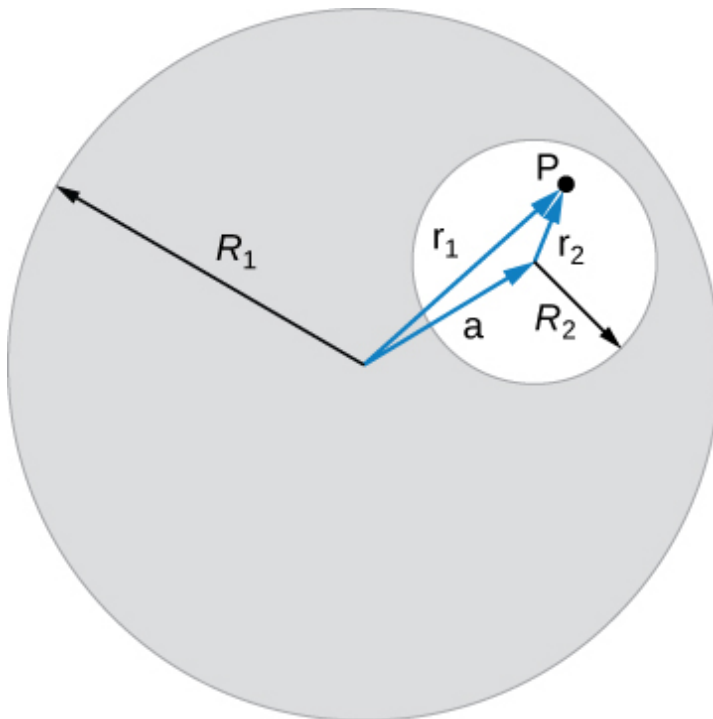
Exercise:**Problem:**

The current density in the long, cylindrical wire shown in the accompanying figure varies with distance r from the center of the wire according to $J = cr$, where c is a constant. (a) What is the current through the wire? (b) What is the magnetic field produced by this current for $r \leq R$? For $r \geq R$?

**Exercise:**

Problem:

A long, straight, cylindrical conductor contains a cylindrical cavity whose axis is displaced by a from the axis of the conductor, as shown in the accompanying figure. The current density in the conductor is given by $\vec{\mathbf{J}} = J_0 \hat{\mathbf{k}}$, where J_0 is a constant and $\hat{\mathbf{k}}$ is along the axis of the conductor. Calculate the magnetic field at an arbitrary point P in the cavity by superimposing the field of a solid cylindrical conductor with radius R_1 and current density $\vec{\mathbf{J}}$ onto the field of a solid cylindrical conductor with radius R_2 and current density $-\vec{\mathbf{J}}$. Then use the fact that the appropriate azimuthal unit vectors can be expressed as $\hat{\theta}_1 = \hat{\mathbf{k}} \times \hat{\mathbf{r}}_1$ and $\hat{\theta}_2 = \hat{\mathbf{k}} \times \hat{\mathbf{r}}_2$ to show that everywhere inside the cavity the magnetic field is given by the constant $\vec{\mathbf{B}} = \frac{1}{2} \mu_0 J_0 \hat{\mathbf{k}} \times \mathbf{a}$, where $\mathbf{a} = \mathbf{r}_1 - \mathbf{r}_2$ and $\mathbf{r}_1 = r_1 \hat{\mathbf{r}}_1$ is the position of P relative to the center of the conductor and $\mathbf{r}_2 = r_2 \hat{\mathbf{r}}_2$ is the position of P relative to the center of the cavity.



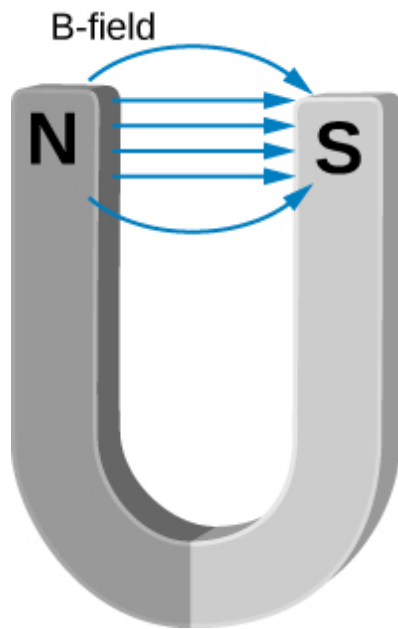
Solution:

derivation

Exercise:

Problem:

Between the two ends of a horseshoe magnet the field is uniform as shown in the diagram. As you move out to outside edges, the field bends. Show by Ampère's law that the field must bend and thereby the field weakens due to these bends.



Exercise:

Problem:

Show that the magnetic field of a thin wire and that of a current loop are zero if you are infinitely far away.

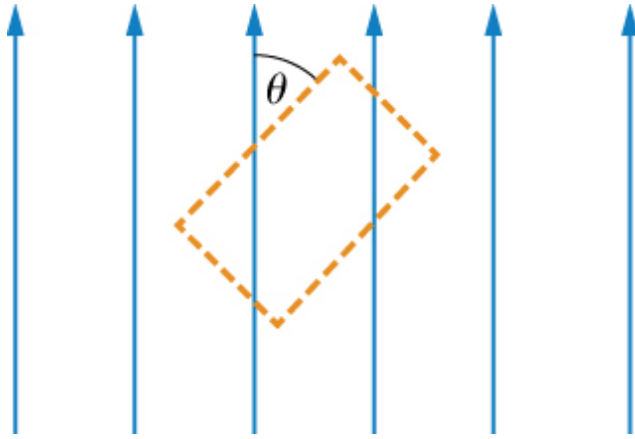
Solution:

As the radial distance goes to infinity, the magnetic fields of each of these formulae go to zero.

Exercise:

Problem:

An Ampère loop is chosen as shown by dashed lines for a parallel constant magnetic field as shown by solid arrows. Calculate $\vec{\mathbf{B}} \cdot d\vec{\mathbf{l}}$ for each side of the loop then find the entire $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{l}}$. Can you think of an Ampère loop that would make the problem easier? Do those results match these?

**Exercise:****Problem:**

A very long, thick cylindrical wire of radius R carries a current density J that varies across its cross-section. The magnitude of the current density at a point a distance r from the center of the wire is given by $J = J_0 \frac{r}{R}$, where J_0 is a constant. Find the magnetic field (a) at a point outside the wire and (b) at a point inside the wire. Write your answer in terms of the net current I through the wire.

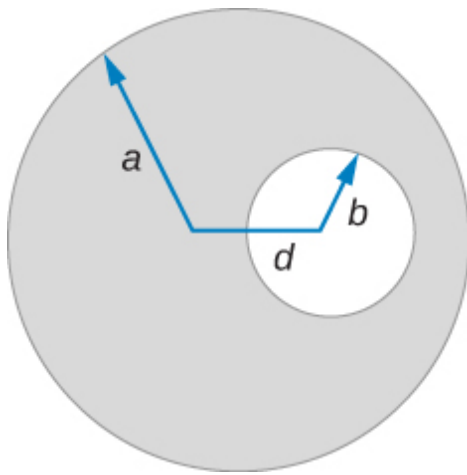
Solution:

$$\text{a. } B = \frac{\mu_0 I}{2\pi r}; \text{ b. } B = \frac{\mu_0 J_0 r^2}{3R}$$

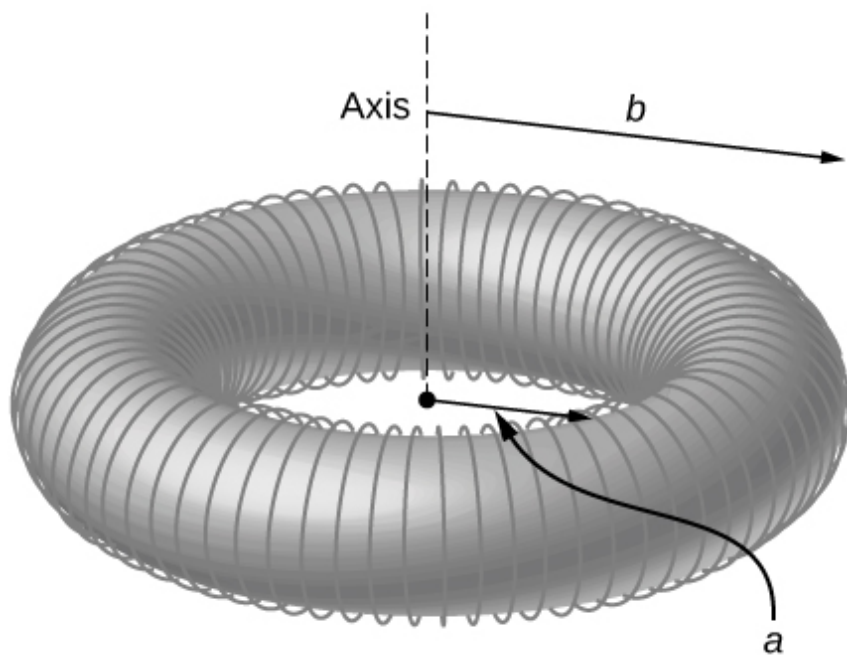
Exercise:

Problem:

A very long, cylindrical wire of radius a has a circular hole of radius b in it at a distance d from the center. The wire carries a uniform current of magnitude I through it. The direction of the current in the figure is out of the paper. Find the magnetic field (a) at a point at the edge of the hole closest to the center of the thick wire, (b) at an arbitrary point inside the hole, and (c) at an arbitrary point outside the wire. (*Hint:* Think of the hole as a sum of two wires carrying current in the opposite directions.)

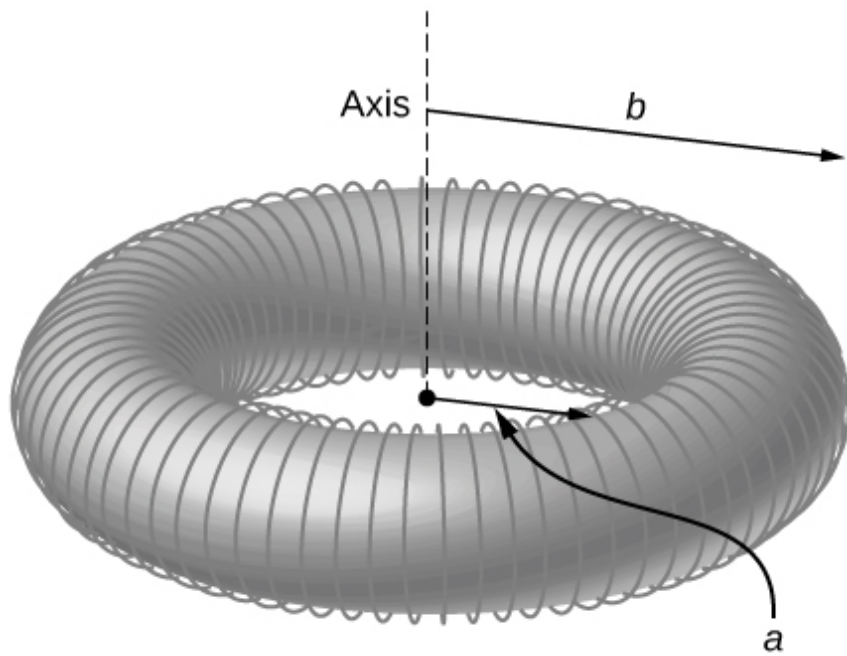
**Exercise:****Problem:**

Magnetic field inside a torus. Consider a torus of rectangular cross-section with inner radius a and outer radius b . N turns of an insulated thin wire are wound evenly on the torus tightly all around the torus and connected to a battery producing a steady current I in the wire. Assume that the current on the top and bottom surfaces in the figure is radial, and the current on the inner and outer radii surfaces is vertical. Find the magnetic field inside the torus as a function of radial distance r from the axis.



Solution:

$$B(r) = \mu_0 NI / 2\pi r$$

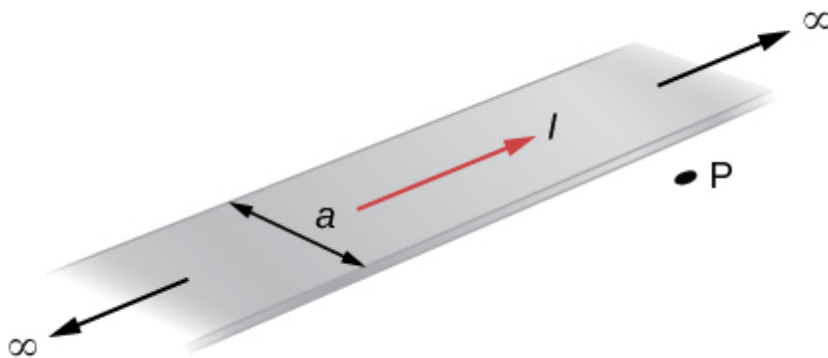


Exercise:**Problem:**

Two long coaxial copper tubes, each of length L , are connected to a battery of voltage V . The inner tube has inner radius a and outer radius b , and the outer tube has inner radius c and outer radius d . The tubes are then disconnected from the battery and rotated in the same direction at angular speed of ω radians per second about their common axis. Find the magnetic field (a) at a point inside the space enclosed by the inner tube $r < a$, and (b) at a point between the tubes $b < r < c$, and (c) at a point outside the tubes $r > d$. (*Hint: Think of copper tubes as a capacitor and find the charge density based on the voltage applied, $Q = VC$, $C = \frac{2\pi\epsilon_0 L}{\ln(c/b)}$.*)

Challenge Problems**Exercise:****Problem:**

The accompanying figure shows a flat, infinitely long sheet of width a that carries a current I uniformly distributed across it. Find the magnetic field at the point P, which is in the plane of the sheet and at a distance x from one edge. Test your result for the limit $a \rightarrow 0$.



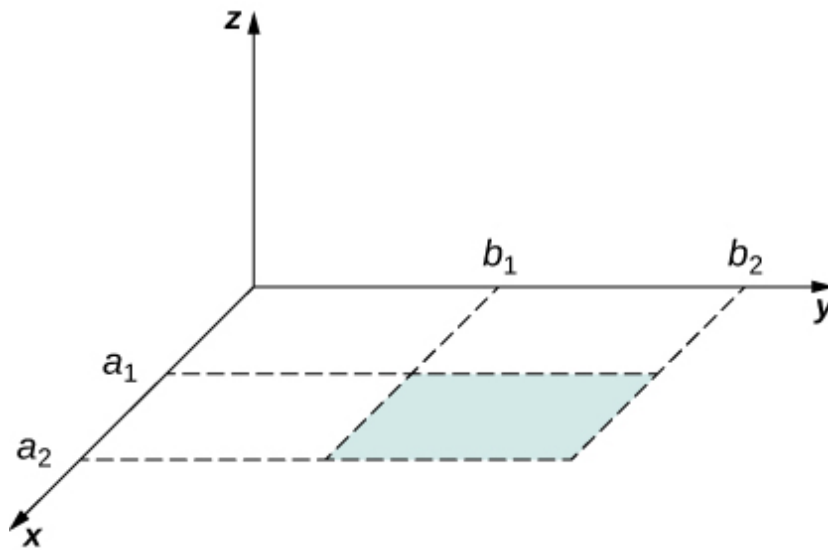
Solution:

$$B = \frac{\mu_0 I}{2\pi a} \ln \frac{x+a}{x}.$$

Exercise:

Problem:

A hypothetical current flowing in the z -direction creates the field $\vec{B} = C \left[(x/y^2)\hat{i} + (1/y)\hat{j} \right]$ in the rectangular region of the xy -plane shown in the accompanying figure. Use Ampère's law to find the current through the rectangle.



Exercise:

Problem:

A nonconducting hard rubber circular disk of radius R is painted with a uniform surface charge density σ . It is rotated about its axis with angular speed ω . (a) Find the magnetic field produced at a point on the axis a distance h meters from the center of the disk. (b) Find the numerical value of magnitude of the magnetic field when $\sigma = 1\text{C/m}^2$, $R = 20\text{ cm}$, $h = 2\text{ cm}$, and $\omega = 400\text{ rad/sec}$, and compare it with the magnitude of magnetic field of Earth, which is about $1/2$ Gauss.

Solution:

a. $B = \frac{\mu_0 \sigma \omega}{2} \left[\frac{2h^2 + R^2}{\sqrt{R^2 + h^2}} - 2h \right]$; b. $B = 4.09 \times 10^{-5} \text{T}$, 82% of Earth's magnetic field

Glossary

diamagnetic materials

their magnetic dipoles align oppositely to an applied magnetic field; when the field is removed, the material is unmagnetized

ferromagnetic materials

contain groups of dipoles, called domains, that align with the applied magnetic field; when this field is removed, the material is still magnetized

hysteresis

property of ferromagnets that is seen when a material's magnetic field is examined versus the applied magnetic field; a loop is created resulting from sweeping the applied field forward and reverse

magnetic susceptibility

ratio of the magnetic field in the material over the applied field at that time; positive susceptibilities are either paramagnetic or ferromagnetic (aligned with the field) and negative susceptibilities are diamagnetic (aligned oppositely with the field)

magnetic domains

groups of magnetic dipoles that are all aligned in the same direction and are coupled together quantum mechanically

paramagnetic materials

their magnetic dipoles align partially in the same direction as the applied magnetic field; when this field is removed, the material is unmagnetized

Introduction

class="introduction"

The black strip found on the back of credit cards and driver's licenses is a very thin layer of magnetic material with information stored on it. Reading and writing the information on the credit card is done with a swiping motion. The physical reason why this is necessary is called electromagnetic induction and is discussed in this chapter.
(credit: modification of work by Jane Whitney)



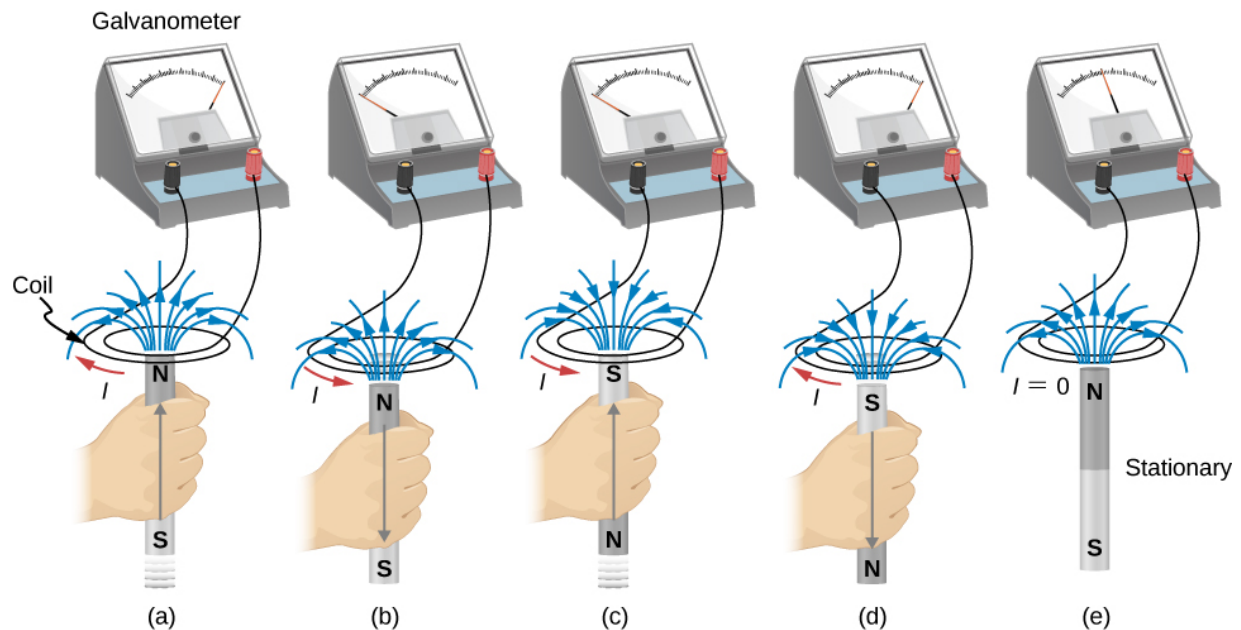
We have been considering electric fields created by fixed charge distributions and magnetic fields produced by constant currents, but electromagnetic phenomena are not restricted to these stationary situations. Most of the interesting applications of electromagnetism are, in fact, time-dependent. To investigate some of these applications, we now remove the time-independent assumption that we have been making and allow the fields to vary with time. In this and the next several chapters, you will see a wonderful symmetry in the behavior exhibited by time-varying electric and magnetic fields. Mathematically, this symmetry is expressed by an additional term in Ampère's law and by another key equation of electromagnetism called Faraday's law. We also discuss how moving a wire through a magnetic field produces an emf or voltage. Lastly, we describe applications of these principles, such as the card reader shown above.

Faraday's Law

By the end of this section, you will be able to:

- Determine the magnetic flux through a surface, knowing the strength of the magnetic field, the surface area, and the angle between the normal to the surface and the magnetic field
- Use Faraday's law to determine the magnitude of induced emf in a closed loop due to changing magnetic flux through the loop

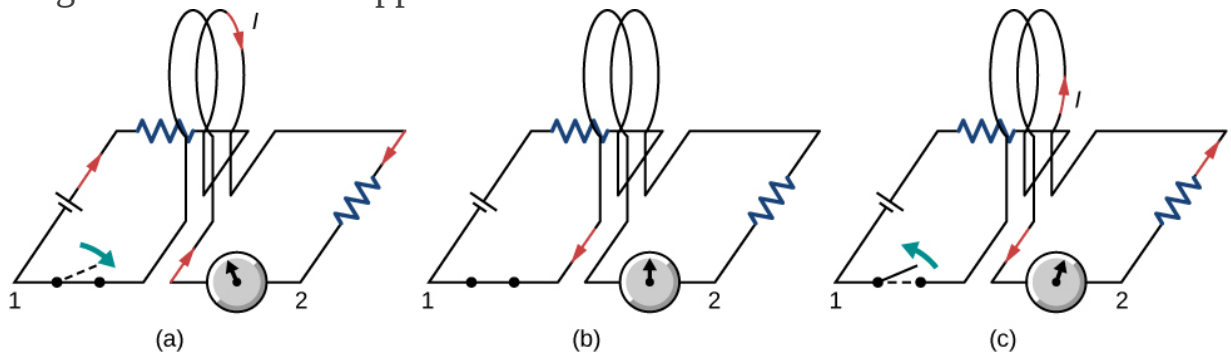
The first productive experiments concerning the effects of time-varying magnetic fields were performed by Michael Faraday in 1831. One of his early experiments is represented in [\[link\]](#). An emf is induced when the magnetic field in the coil is changed by pushing a bar magnet into or out of the coil. Emfs of opposite signs are produced by motion in opposite directions, and the directions of emfs are also reversed by reversing poles. The same results are produced if the coil is moved rather than the magnet—it is the relative motion that is important. The faster the motion, the greater the emf, and there is no emf when the magnet is stationary relative to the coil.



Movement of a magnet relative to a coil produces emfs as shown (a–d). The same emfs are produced if the coil is moved relative to the magnet. This short-lived emf is only present during the motion. The

greater the speed, the greater the magnitude of the emf, and the emf is zero when there is no motion, as shown in (e).

Faraday also discovered that a similar effect can be produced using two circuits—a changing current in one circuit induces a current in a second, nearby circuit. For example, when the switch is closed in circuit 1 of [\[link\]](#) (a), the ammeter needle of circuit 2 momentarily deflects, indicating that a short-lived current surge has been induced in that circuit. The ammeter needle quickly returns to its original position, where it remains. However, if the switch of circuit 1 is now suddenly opened, another short-lived current surge in the direction opposite from before is observed in circuit 2.



(a) Closing the switch of circuit 1 produces a short-lived current surge in circuit 2. (b) If the switch remains closed, no current is observed in circuit 2. (c) Opening the switch again produces a short-lived current in circuit 2 but in the opposite direction from before.

Faraday realized that in both experiments, a current flowed in the circuit containing the ammeter only when the magnetic field in the region occupied by that circuit was *changing*. As the magnet of the figure was moved, the strength of its magnetic field at the loop changed; and when the current in circuit 1 was turned on or off, the strength of its magnetic field at circuit 2 changed. Faraday was eventually able to interpret these and all other experiments involving magnetic fields that vary with time in terms of the following law:

Note:**Faraday's Law**

The emf ε induced is the negative change in the magnetic flux Φ_m per unit time. Any change in the magnetic field or change in orientation of the area of the coil with respect to the magnetic field induces a voltage (emf).

The **magnetic flux** is a measurement of the amount of magnetic field lines through a given surface area, as seen in [\[link\]](#). This definition is similar to the electric flux studied earlier. This means that if we have

Note:**Equation:**

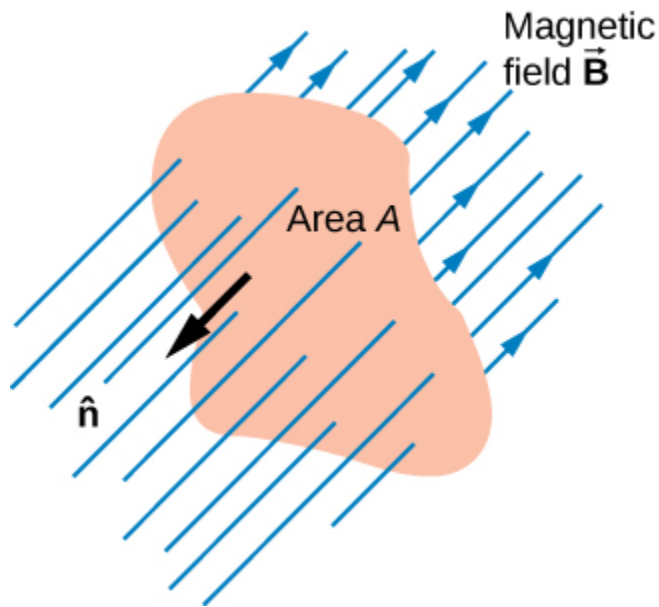
$$\Phi_m = \int_S \vec{\mathbf{B}} \cdot \hat{\mathbf{n}} dA,$$

then the **induced emf** or the voltage generated by a conductor or coil moving in a magnetic field is

Equation:

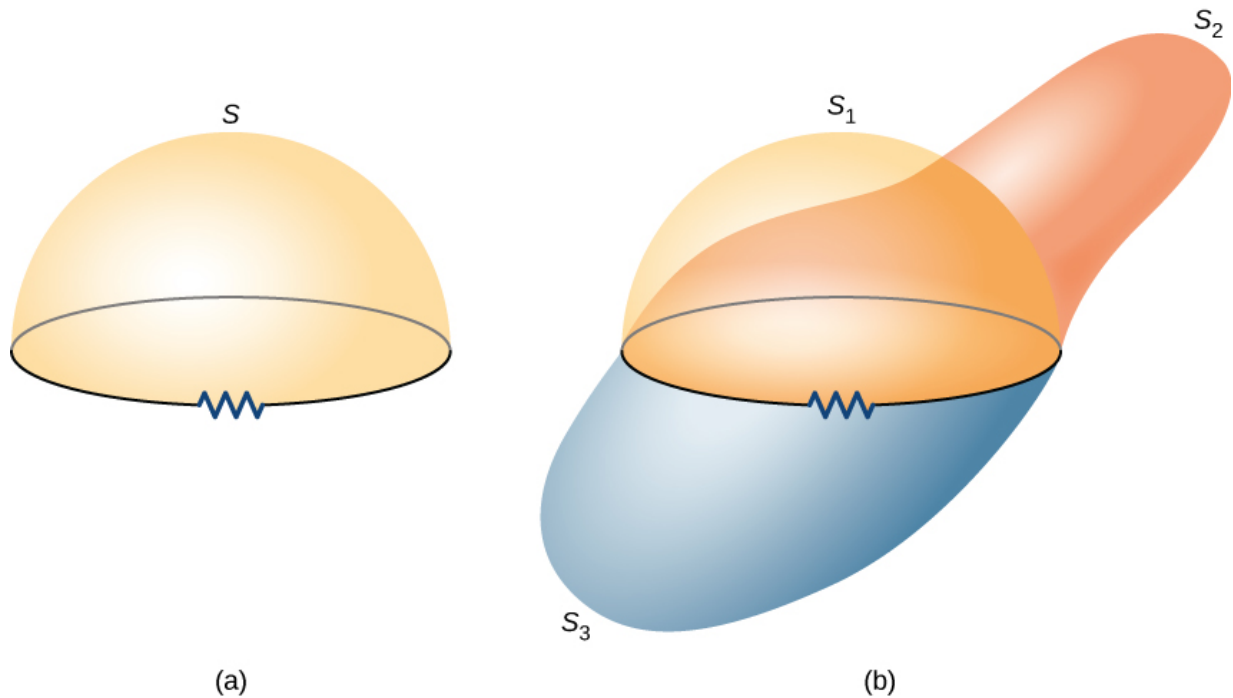
$$\varepsilon = -\frac{d}{dt} \int_S \vec{\mathbf{B}} \cdot \hat{\mathbf{n}} dA = -\frac{d\Phi_m}{dt}.$$

The negative sign describes the direction in which the induced emf drives current around a circuit. However, that direction is most easily determined with a rule known as Lenz's law, which we will discuss shortly.



The magnetic flux is the amount of magnetic field lines cutting through a surface area A defined by the unit area vector \hat{n} . If the angle between the unit area \hat{n} and magnetic field vector \vec{B} are parallel or antiparallel, as shown in the diagram, the magnetic flux is the highest possible value given the values of area and magnetic field.

Part (a) of [\[link\]](#) depicts a circuit and an arbitrary surface S that it bounds. Notice that S is an *open surface*. It can be shown that *any* open surface bounded by the circuit in question can be used to evaluate Φ_m . For example, Φ_m is the same for the various surfaces S_1, S_2, \dots of part (b) of the figure.



(a) A circuit bounding an arbitrary open surface S . The planar area bounded by the circuit is not part of S . (b) Three arbitrary open surfaces bounded by the same circuit. The value of Φ_m is the same for all these surfaces.

The SI unit for magnetic flux is the weber (Wb),

Equation:

$$1 \text{ Wb} = 1 \text{ T} \cdot \text{m}^2.$$

Occasionally, the magnetic field unit is expressed as webers per square meter (Wb/m^2) instead of teslas, based on this definition. In many practical applications, the circuit of interest consists of a number N of tightly wound turns (see [\[link\]](#)). Each turn experiences the same magnetic flux. Therefore, the net magnetic flux through the circuits is N times the flux through one turn, and Faraday's law is written as

Note:

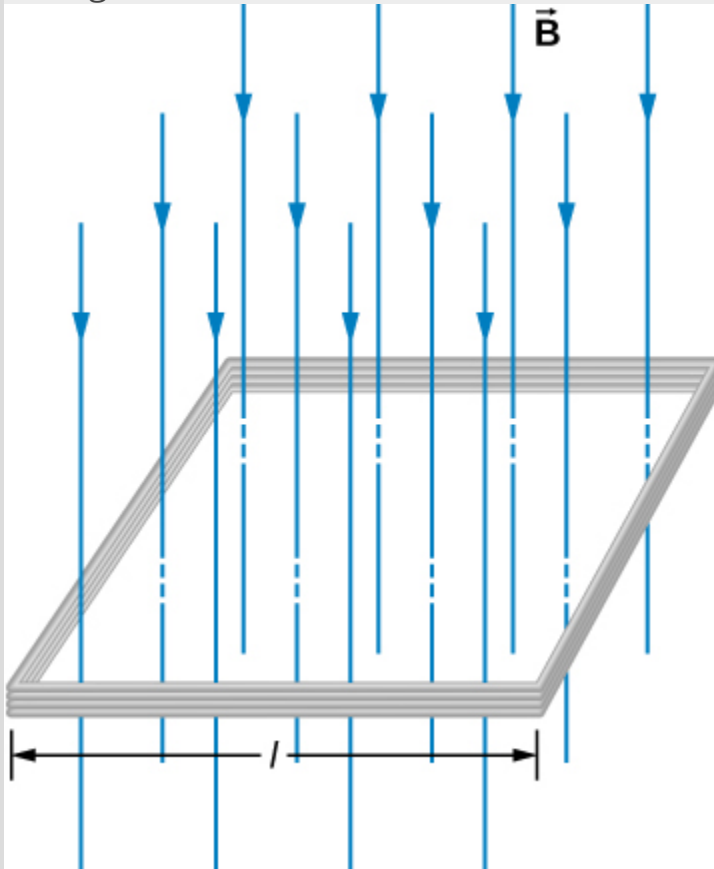
Equation:

$$\varepsilon = -\frac{d}{dt}(N\Phi_m) = -N\frac{d\Phi_m}{dt}.$$

Example:

A Square Coil in a Changing Magnetic Field

The square coil of [link](#) has sides $l = 0.25$ m long and is tightly wound with $N = 200$ turns of wire. The resistance of the coil is $R = 5.0\ \Omega$. The coil is placed in a spatially uniform magnetic field that is directed perpendicular to the face of the coil and whose magnitude is decreasing at a rate $dB/dt = -0.040$ T/s. (a) What is the magnitude of the emf induced in the coil? (b) What is the magnitude of the current circulating through the coil?



A square coil with N turns of wire with uniform magnetic field \vec{B} directed in the downward direction, perpendicular to the coil.

Strategy

The area vector, or \hat{n} direction, is perpendicular to area covering the loop. We will choose this to be pointing downward so that \vec{B} is parallel to \hat{n} and that the flux turns into multiplication of magnetic field times area. The area of the loop is not changing in time, so it can be factored out of the time derivative, leaving the magnetic field as the only quantity varying in time. Lastly, we can apply Ohm's law once we know the induced emf to find the current in the loop.

Solution

- a. The flux through one turn is

Equation:

$$\Phi_m = BA = Bl^2,$$

so we can calculate the magnitude of the emf from Faraday's law. The sign of the emf will be discussed in the next section, on Lenz's law:

Equation:

$$\begin{aligned} |\varepsilon| &= \left| -N \frac{d\Phi_m}{dt} \right| = Nl^2 \frac{dB}{dt} \\ &= (200)(0.25 \text{ m})^2 (0.040 \text{ T/s}) = 0.50 \text{ V}. \end{aligned}$$

- b. The magnitude of the current induced in the coil is

Equation:

$$I = \frac{\varepsilon}{R} = \frac{0.50 \text{ V}}{5.0 \Omega} = 0.10 \text{ A}.$$

Significance

If the area of the loop were changing in time, we would not be able to pull it out of the time derivative. Since the loop is a closed path, the result of this current would be a small amount of heating of the wires until the magnetic field stops changing. This may increase the area of the loop slightly as the wires are heated.

Note:

Exercise:

Problem:

Check Your Understanding A closely wound coil has a radius of 4.0 cm, 50 turns, and a total resistance of $40\ \Omega$. At what rate must a magnetic field perpendicular to the face of the coil change in order to produce Joule heating in the coil at a rate of 2.0 mW?

Solution:

1.1 T/s

Summary

- The magnetic flux through an enclosed area is defined as the amount of field lines cutting through a surface area A defined by the unit area vector.
- The units for magnetic flux are webers, where $1\ \text{Wb} = 1\ \text{T} \cdot \text{m}^2$.
- The induced emf in a closed loop due to a change in magnetic flux through the loop is known as Faraday's law. If there is no change in magnetic flux, no induced emf is created.

Conceptual Questions

Exercise:

Problem:

A stationary coil is in a magnetic field that is changing with time. Does the emf induced in the coil depend on the actual values of the magnetic field?

Solution:

The emf depends on the rate of change of the magnetic field.

Exercise:**Problem:**

In Faraday's experiments, what would be the advantage of using coils with many turns?

Exercise:**Problem:**

A copper ring and a wooden ring of the same dimensions are placed in magnetic fields so that there is the same change in magnetic flux through them. Compare the induced electric fields and currents in the rings.

Solution:

Both have the same induced electric fields; however, the copper ring has a much higher induced emf because it conducts electricity better than the wooden ring.

Exercise:**Problem:**

Discuss the factors determining the induced emf in a closed loop of wire.

Exercise:

Problem:

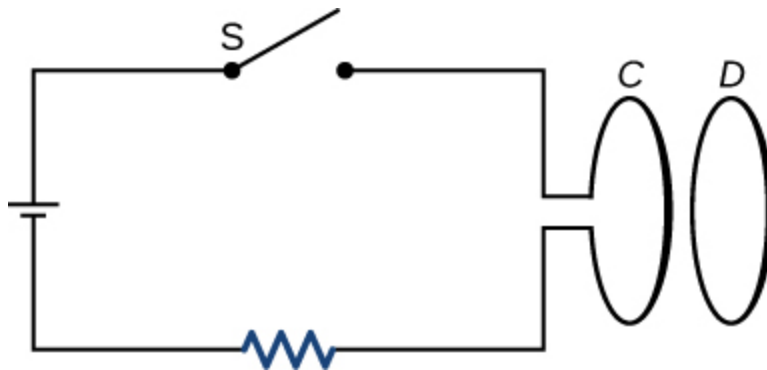
(a) Does the induced emf in a circuit depend on the resistance of the circuit? (b) Does the induced current depend on the resistance of the circuit?

Solution:

a. no; b. yes

Exercise:**Problem:**

How would changing the radius of loop D shown below affect its emf, assuming C and D are much closer together compared to their radii?

**Exercise:****Problem:**

Can there be an induced emf in a circuit at an instant when the magnetic flux through the circuit is zero?

Solution:

As long as the magnetic flux is changing from positive to negative or negative to positive, there could be an induced emf.

Exercise:

Problem:

Does the induced emf always act to decrease the magnetic flux through a circuit?

Exercise:**Problem:**

How would you position a flat loop of wire in a changing magnetic field so that there is no induced emf in the loop?

Solution:

Position the loop so that the field lines run perpendicular to the area vector or parallel to the surface.

Exercise:**Problem:**

The normal to the plane of a single-turn conducting loop is directed at an angle θ to a spatially uniform magnetic field \vec{B} . It has a fixed area and orientation relative to the magnetic field. Show that the emf induced in the loop is given by $\varepsilon = (dB/dt)(A \cos \theta)$, where A is the area of the loop.

Problems**Exercise:****Problem:**

A 50-turn coil has a diameter of 15 cm. The coil is placed in a spatially uniform magnetic field of magnitude 0.50 T so that the face of the coil and the magnetic field are perpendicular. Find the magnitude of the emf induced in the coil if the magnetic field is reduced to zero uniformly in (a) 0.10 s, (b) 1.0 s, and (c) 60 s.

Exercise:

Problem:

Repeat your calculations of the preceding problem's time of 0.1 s with the plane of the coil making an angle of (a) 30° , (b) 60° , and (c) 90° with the magnetic field.

Solution:

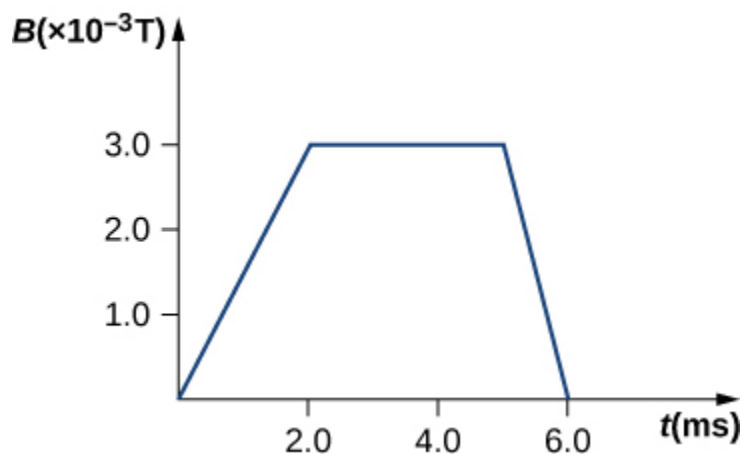
a. 3.8 V; b. 2.2 V; c. 0 V

Exercise:**Problem:**

A square loop whose sides are 6.0-cm long is made with copper wire of radius 1.0 mm. If a magnetic field perpendicular to the loop is changing at a rate of 5.0 mT/s, what is the current in the loop?

Exercise:**Problem:**

The magnetic field through a circular loop of radius 10.0 cm varies with time as shown below. The field is perpendicular to the loop. Plot the magnitude of the induced emf in the loop as a function of time.

**Solution:**

$$B = 1.5t, 0 \leq t < 2.0 \text{ ms}, B = 3.0 \text{ mT}, 2.0 \text{ ms} \leq t \leq 5.0 \text{ ms},$$

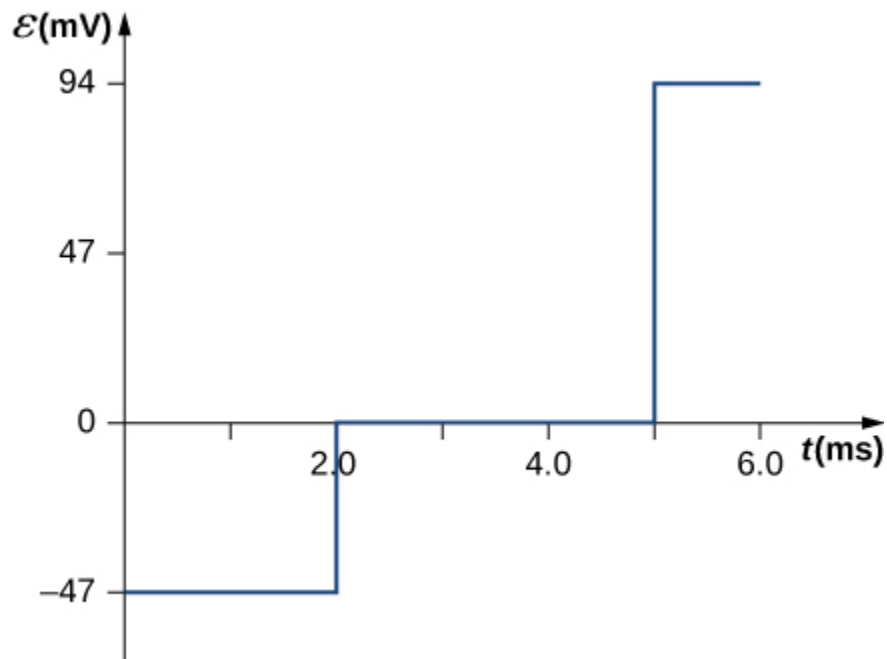
$$B = -3.0t + 18 \text{ mT}, 5.0 \text{ ms} < t \leq 6.0 \text{ ms},$$

$$\varepsilon = -\frac{d\Phi_m}{dt} = -\frac{d(BA)}{dt} = -A\frac{dB}{dt},$$

$$\begin{aligned}\varepsilon &= -\pi(0.100 \text{ m})^2 (1.5 \text{ T/s}) \\ &= -47 \text{ mV} \quad (0 \leq t < 2.0 \text{ ms}),\end{aligned}$$

$$\varepsilon = \pi(0.100 \text{ m})^2 (0) = 0 \quad (2.0 \text{ ms} \leq t \leq 5.0 \text{ ms}),$$

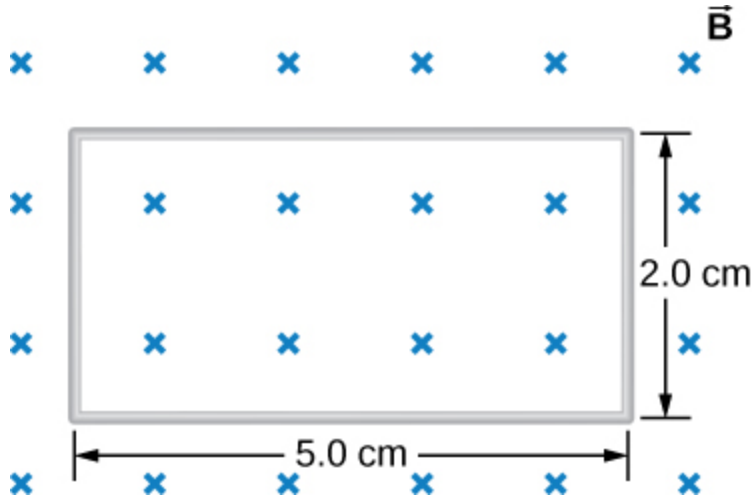
$$\varepsilon = -\pi(0.100 \text{ m})^2 (-3.0 \text{ T/s}) = 94 \text{ mV} \quad (5.0 \text{ ms} < t < 6.0 \text{ ms}).$$



Exercise:

Problem:

The accompanying figure shows a single-turn rectangular coil that has a resistance of 2.0Ω . The magnetic field at all points inside the coil varies according to $B = B_0 e^{-\alpha t}$, where $B_0 = 0.25 \text{ T}$ and $\alpha = 200 \text{ Hz}$. What is the current induced in the coil at (a) $t = 0.001 \text{ s}$, (b) 0.002 s , (c) 2.0 s ?



Exercise:

Problem:

How would the answers to the preceding problem change if the coil consisted of 20 closely spaced turns?

Solution:

Each answer is 20 times the previously given answers.

Exercise:

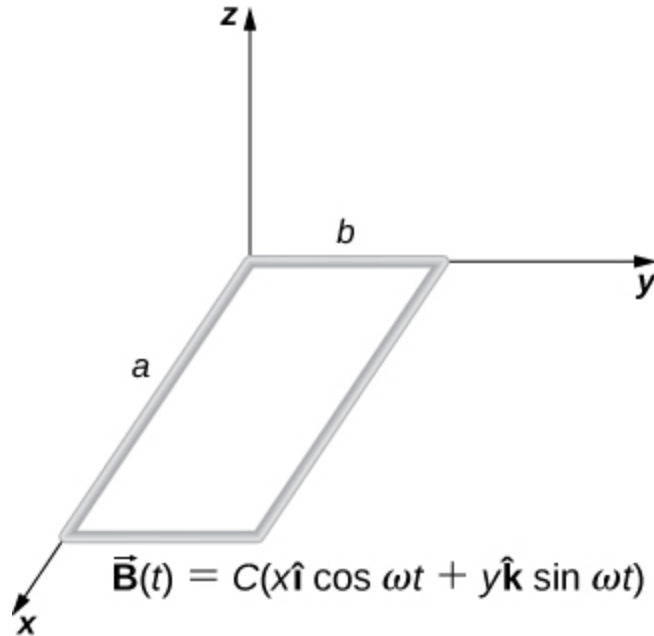
Problem:

A long solenoid with $n = 10$ turns per centimeter has a cross-sectional area of 5.0 cm^2 and carries a current of 0.25 A . A coil with five turns encircles the solenoid. When the current through the solenoid is turned off, it decreases to zero in 0.050 s . What is the average emf induced in the coil?

Exercise:

Problem:

A rectangular wire loop with length a and width b lies in the xy -plane, as shown below. Within the loop there is a time-dependent magnetic field given by $\vec{\mathbf{B}}(t) = C \left((x \cos \omega t) \hat{\mathbf{i}} + (y \sin \omega t) \hat{\mathbf{k}} \right)$, with $\vec{\mathbf{B}}(t)$ in tesla. Determine the emf induced in the loop as a function of time.



Solution:

$$\begin{aligned}\hat{\mathbf{n}} &= \hat{\mathbf{k}}, \quad d\Phi_{\text{m}} = Cy \sin(\omega t) dx dy, \\ \Phi_{\text{m}} &= \frac{Cab^2 \sin(\omega t)}{2}, \\ \varepsilon &= -\frac{Cab^2 \omega \cos(\omega t)}{2}.\end{aligned}$$

Exercise:

Problem:

The magnetic field perpendicular to a single wire loop of diameter 10.0 cm decreases from 0.50 T to zero. The wire is made of copper and has a diameter of 2.0 mm and length 1.0 cm. How much charge moves through the wire while the field is changing?

Glossary

Faraday's law

induced emf is created in a closed loop due to a change in magnetic flux through the loop

induced emf

short-lived voltage generated by a conductor or coil moving in a magnetic field

magnetic flux

measurement of the amount of magnetic field lines through a given area

Lenz's Law

By the end of this section, you will be able to:

- Use Lenz's law to determine the direction of induced emf whenever a magnetic flux changes
- Use Faraday's law with Lenz's law to determine the induced emf in a coil and in a solenoid

The direction in which the induced emf drives current around a wire loop can be found through the negative sign. However, it is usually easier to determine this direction with **Lenz's law**, named in honor of its discoverer, Heinrich Lenz (1804–1865). (Faraday also discovered this law, independently of Lenz.) We state Lenz's law as follows:

Note:

Lenz's Law

The direction of the induced emf drives current around a wire loop to always *oppose* the change in magnetic flux that causes the emf.

Lenz's law can also be considered in terms of conservation of energy. If pushing a magnet into a coil causes current, the energy in that current must have come from somewhere. If the induced current causes a magnetic field opposing the increase in field of the magnet we pushed in, then the situation is clear. We pushed a magnet against a field and did work on the system, and that showed up as current. If it were not the case that the induced field opposes the change in the flux, the magnet would be pulled in produce a current without anything having done work. Electric potential energy would have been created, violating the conservation of energy.

To determine an induced emf ε , you first calculate the magnetic flux Φ_m and then obtain $d\Phi_m/dt$. The magnitude of ε is given by $\varepsilon = |d\Phi_m/dt|$. Finally, you can apply Lenz's law to determine the sense of ε . This will be

developed through examples that illustrate the following problem-solving strategy.

Note:

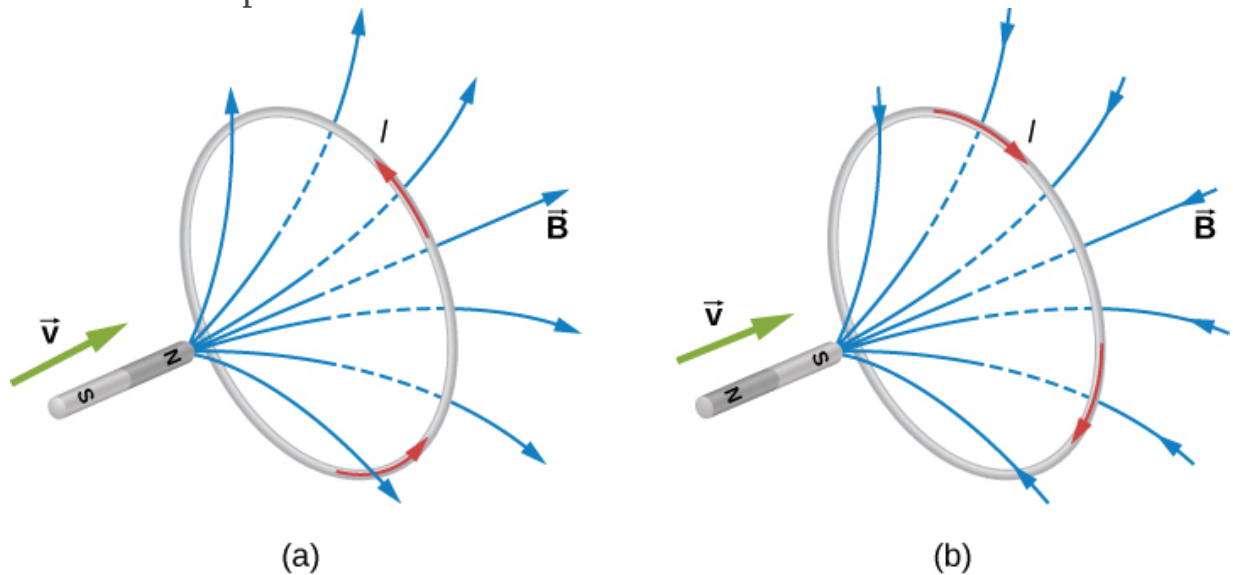
Lenz's Law

To use Lenz's law to determine the directions of induced magnetic fields, currents, and emfs:

1. Make a sketch of the situation for use in visualizing and recording directions.
2. Determine the direction of the applied magnetic field \vec{B} .
3. Determine whether its magnetic flux is increasing or decreasing.
4. Now determine the direction of the induced magnetic field \vec{B} . The induced magnetic field tries to reinforce a magnetic flux that is decreasing or opposes a magnetic flux that is increasing. Therefore, the induced magnetic field adds or subtracts to the applied magnetic field, depending on the change in magnetic flux.
5. Use right-hand rule 2 (RHR-2; see [Magnetic Forces and Fields](#)) to determine the direction of the induced current I that is responsible for the induced magnetic field \vec{B} .
6. The direction (or polarity) of the induced emf can now drive a conventional current in this direction.

Let's apply Lenz's law to the system of [\[link\]](#)(a). We designate the "front" of the closed conducting loop as the region containing the approaching bar magnet, and the "back" of the loop as the other region. As the north pole of the magnet moves toward the loop, the flux through the loop due to the field of the magnet increases because the strength of field lines directed from the front to the back of the loop is increasing. A current is therefore induced in the loop. By Lenz's law, the direction of the induced current must be such that its own magnetic field is directed in a way to *oppose* the changing flux caused by the field of the approaching magnet. Hence, the induced current

circulates so that its magnetic field lines through the loop are directed from the back to the front of the loop. By RHR-2, place your thumb pointing against the magnetic field lines, which is toward the bar magnet. Your fingers wrap in a counterclockwise direction as viewed from the bar magnet. Alternatively, we can determine the direction of the induced current by treating the current loop as an electromagnet that *opposes* the approach of the north pole of the bar magnet. This occurs when the induced current flows as shown, for then the face of the loop nearer the approaching magnet is also a north pole.

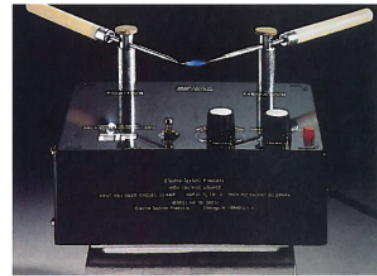
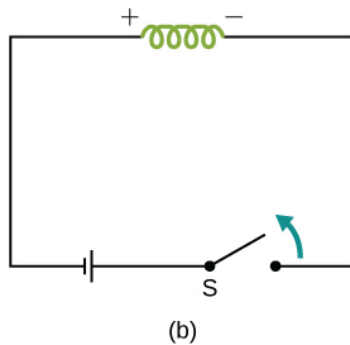
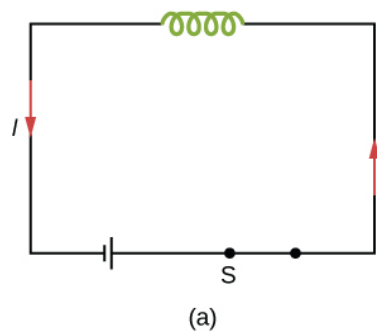


The change in magnetic flux caused by the approaching magnet induces a current in the loop. (a) An approaching north pole induces a counterclockwise current with respect to the bar magnet. (b) An approaching south pole induces a clockwise current with respect to the bar magnet.

Part (b) of the figure shows the south pole of a magnet moving toward a conducting loop. In this case, the flux through the loop due to the field of the magnet increases because the number of field lines directed from the back to the front of the loop is increasing. To oppose this change, a current is induced in the loop whose field lines through the loop are directed from the front to the back. Equivalently, we can say that the current flows in a direction so that the face of the loop nearer the approaching magnet is a

south pole, which then repels the approaching south pole of the magnet. By RHR-2, your thumb points away from the bar magnet. Your fingers wrap in a clockwise fashion, which is the direction of the induced current.

Another example illustrating the use of Lenz's law is shown in [\[link\]](#). When the switch is opened, the decrease in current through the solenoid causes a decrease in magnetic flux through its coils, which induces an emf in the solenoid. This emf must oppose the change (the termination of the current) causing it. Consequently, the induced emf has the polarity shown and drives in the direction of the original current. This may generate an arc across the terminals of the switch as it is opened.



- (a) A solenoid connected to a source of emf. (b) Opening switch S terminates the current, which in turn induces an emf in the solenoid. (c) A potential difference between the ends of the sharply pointed rods is produced by inducing an emf in a coil. This potential difference is large enough to produce an arc between the sharp points.

Note:

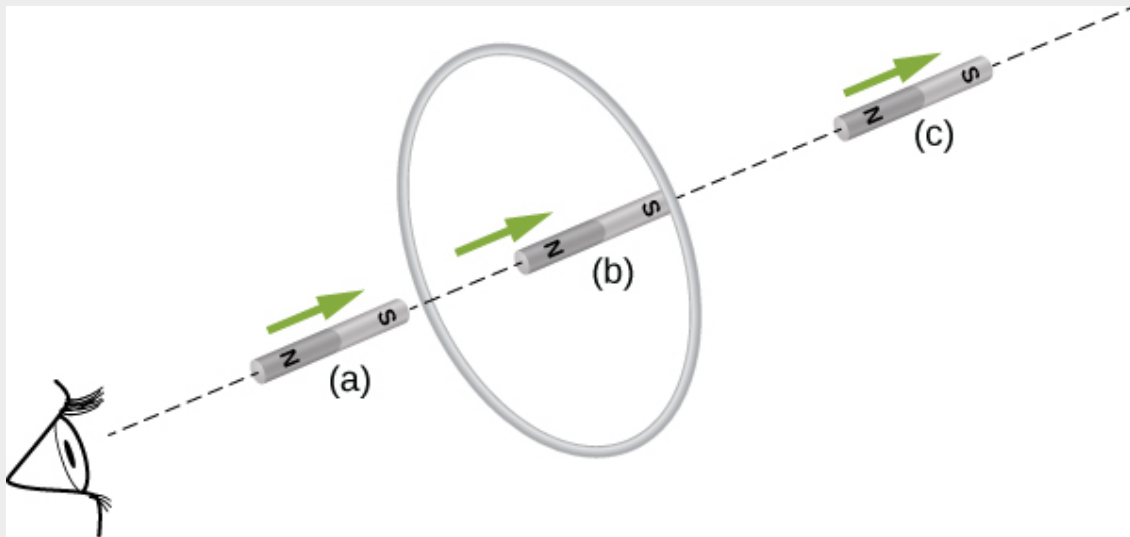
Exercise:

Problem:

Check Your Understanding Find the direction of the induced current in the wire loop shown below as the magnet enters, passes through, and leaves the loop.

Solution:

To the observer shown, the current flows clockwise as the magnet approaches, decreases to zero when the magnet is centered in the plane of the coil, and then flows counterclockwise as the magnet leaves the coil.

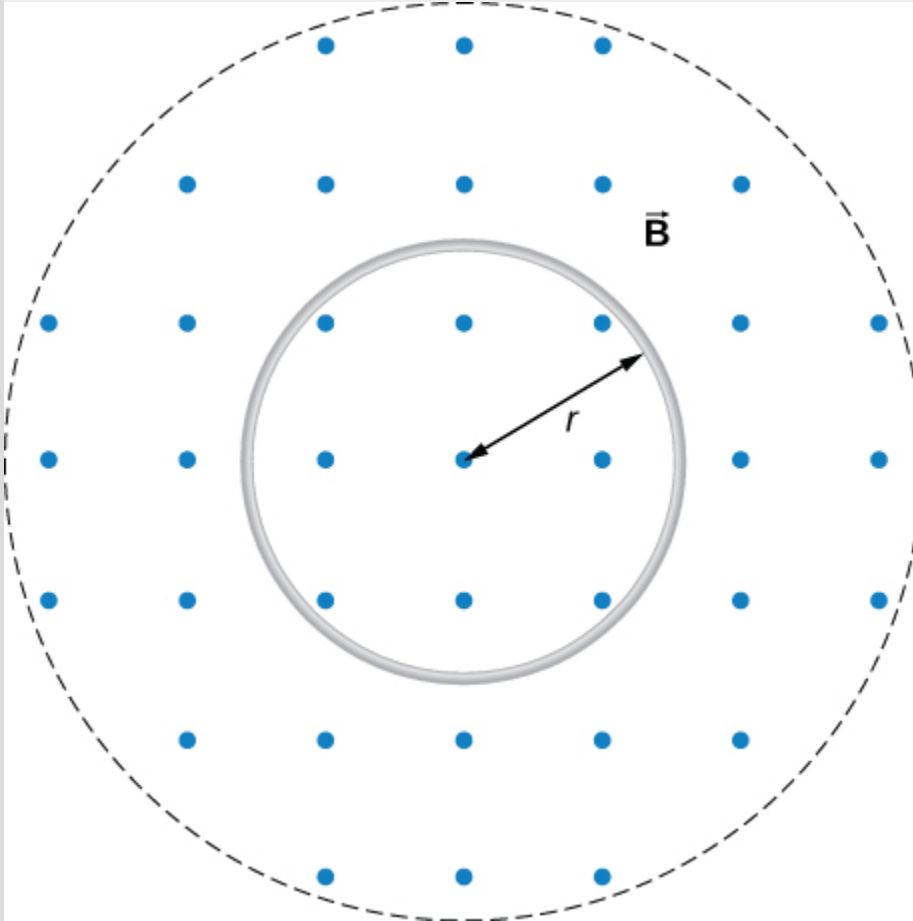
**Note:****Exercise:****Problem:**

Check Your Understanding Verify the directions of the induced currents in [\[link\]](#).

Example:**A Circular Coil in a Changing Magnetic Field**

A magnetic field \vec{B} is directed outward perpendicular to the plane of a circular coil of radius $r = 0.50 \text{ m}$ ([\[link\]](#)). The field is cylindrically

symmetrical with respect to the center of the coil, and its magnitude decays exponentially according to $B = (1.5\text{T})e^{-(5.0\text{s}^{-1})t}$, where B is in teslas and t is in seconds. (a) Calculate the emf induced in the coil at the times $t_1 = 0$, $t_2 = 5.0 \times 10^{-2}\text{s}$, and $t_3 = 1.0\text{s}$. (b) Determine the current in the coil at these three times if its resistance is $10\ \Omega$.



A circular coil in a decreasing magnetic field.

Strategy

Since the magnetic field is perpendicular to the plane of the coil and constant over each spot in the coil, the dot product of the magnetic field \vec{B} and normal to the area unit vector \hat{n} turns into a multiplication. The magnetic field can be pulled out of the integration, leaving the flux as the product of the magnetic field times area. We need to take the time

derivative of the exponential function to calculate the emf using Faraday's law. Then we use Ohm's law to calculate the current.

Solution

- a. Since \vec{B} is perpendicular to the plane of the coil, the magnetic flux is given by

Equation:

$$\begin{aligned}\Phi_m &= B\pi r^2 = (1.5e^{-5.0t} \text{ T})\pi(0.50 \text{ m})^2 \\ &= 1.2e^{-(5.0\text{s}^{-1})t} \text{ Wb}.\end{aligned}$$

From Faraday's law, the magnitude of the induced emf is

Equation:

$$\varepsilon = \left| \frac{d\Phi_m}{dt} \right| = \left| \frac{d}{dt}(1.2e^{-(5.0\text{s}^{-1})t} \text{ Wb}) \right| = 6.0 e^{-(5.0\text{s}^{-1})t} \text{ V}.$$

Since \vec{B} is directed out of the page and is decreasing, the induced current must flow counterclockwise when viewed from above so that the magnetic field it produces through the coil also points out of the page. For all three times, the sense of ε is counterclockwise; its magnitudes are

Equation:

$$\varepsilon(t_1) = 6.0 \text{ V}; \quad \varepsilon(t_2) = 4.7 \text{ V}; \quad \varepsilon(t_3) = 0.040 \text{ V}.$$

- b. From Ohm's law, the respective currents are

Equation:

$$\begin{aligned}I(t_1) &= \frac{\varepsilon(t_1)}{R} = \frac{6.0 \text{ V}}{10 \Omega} = 0.60 \text{ A}; \\ I(t_2) &= \frac{4.7 \text{ V}}{10 \Omega} = 0.47 \text{ A};\end{aligned}$$

and

Equation:

$$I(t_3) = \frac{0.040 \text{ V}}{10 \, \Omega} = 4.0 \times 10^{-3} \text{ A}.$$

Significance

An emf voltage is created by a changing magnetic flux over time. If we know how the magnetic field varies with time over a constant area, we can take its time derivative to calculate the induced emf.

Example:

Changing Magnetic Field Inside a Solenoid

The current through the windings of a solenoid with $n = 2000$ turns per meter is changing at a rate $dI/dt = 3.0 \text{ A/s}$. (See [Sources of Magnetic Fields](#) for a discussion of solenoids.) The solenoid is 50-cm long and has a cross-sectional diameter of 3.0 cm. A small coil consisting of $N = 20$ closely wound turns wrapped in a circle of diameter 1.0 cm is placed in the middle of the solenoid such that the plane of the coil is perpendicular to the central axis of the solenoid. Assuming that the infinite-solenoid approximation is valid at the location of the small coil, determine the magnitude of the emf induced in the coil.

Strategy

The magnetic field in the middle of the solenoid is a uniform value of $\mu_0 nI$. This field is producing a maximum magnetic flux through the coil as it is directed along the length of the solenoid. Therefore, the magnetic flux through the coil is the product of the solenoid's magnetic field times the area of the coil. Faraday's law involves a time derivative of the magnetic flux. The only quantity varying in time is the current, the rest can be pulled out of the time derivative. Lastly, we include the number of turns in the coil to determine the induced emf in the coil.

Solution

Since the field of the solenoid is given by $B = \mu_0 nI$, the flux through each turn of the small coil is

Equation:

$$\Phi_m = \mu_0 n I \left(\frac{\pi d^2}{4} \right),$$

where d is the diameter of the coil. Now from Faraday's law, the magnitude of the emf induced in the coil is

Equation:

$$\begin{aligned} \varepsilon &= \left| N \frac{d\Phi_m}{dt} \right| = \left| N \mu_0 n \frac{\pi d^2}{4} \frac{dI}{dt} \right| \\ &= 20 \left(4\pi \times 10^{-7} \text{ T} \cdot \text{m/s} \right) (2000 \text{ m}^{-1}) \frac{\pi (0.010 \text{ m})^2}{4} (3.0 \text{ A/s}) \\ &= 1.2 \times 10^{-5} \text{ V}. \end{aligned}$$

Significance

When the current is turned on in a vertical solenoid, as shown in [\[link\]](#), the ring has an induced emf from the solenoid's changing magnetic flux that opposes the change. The result is that the ring is fired vertically into the air.



The jumping ring. When a current is turned on in the vertical solenoid, a current is induced in the metal ring.

The stray field produced by the solenoid causes the ring to jump off the solenoid.

Note:

Visit this [website](#) for a demonstration of the jumping ring from MIT.

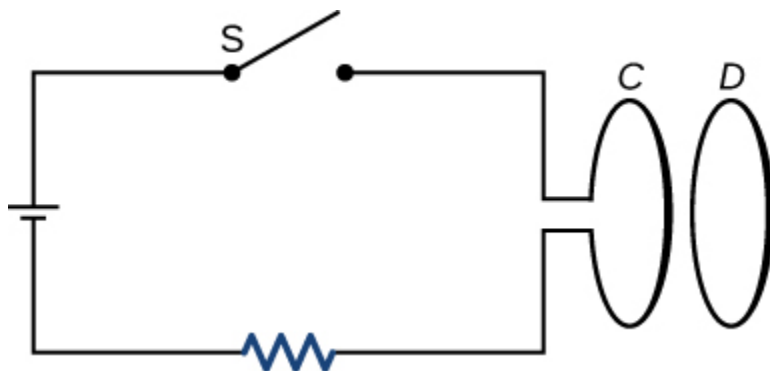
Summary

- We can use Lenz's law to determine the directions of induced magnetic fields, currents, and emfs.
- The direction of an induced emf always opposes the change in magnetic flux that causes the emf, a result known as Lenz's law.

Conceptual Questions

Exercise:**Problem:**

The circular conducting loops shown in the accompanying figure are parallel, perpendicular to the plane of the page, and coaxial. (a) When the switch S is closed, what is the direction of the current induced in D ? (b) When the switch is opened, what is the direction of the current induced in loop D ?



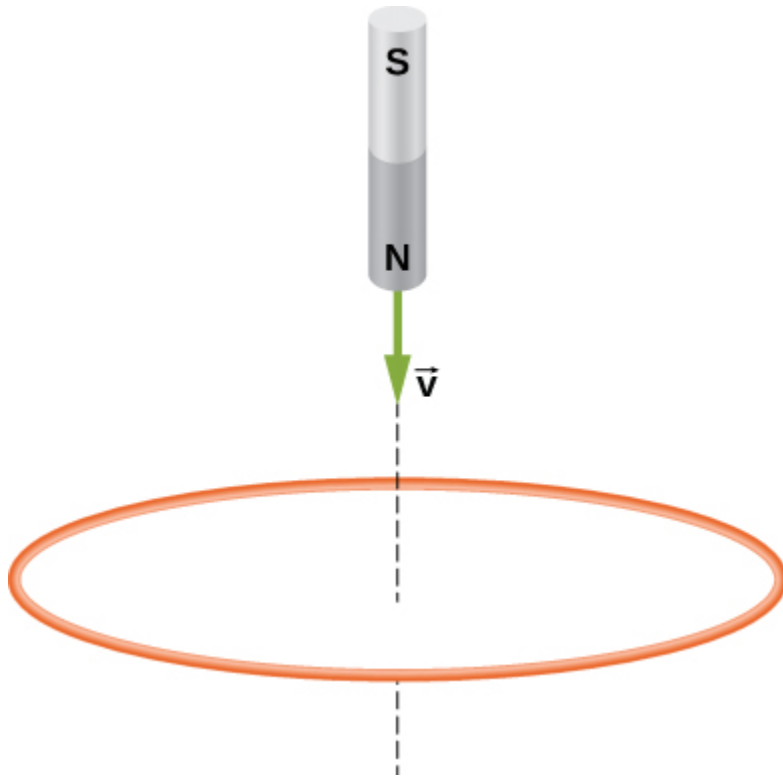
Solution:

a. CW as viewed from the circuit; b. CCW as viewed from the circuit

Exercise:

Problem:

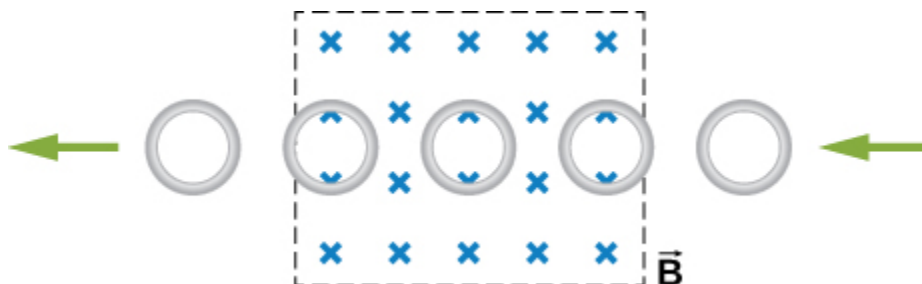
The north pole of a magnet is moved toward a copper loop, as shown below. If you are looking at the loop from above the magnet, will you say the induced current is circulating clockwise or counterclockwise?



Exercise:

Problem:

The accompanying figure shows a conducting ring at various positions as it moves through a magnetic field. What is the sense of the induced emf for each of those positions?



Solution:

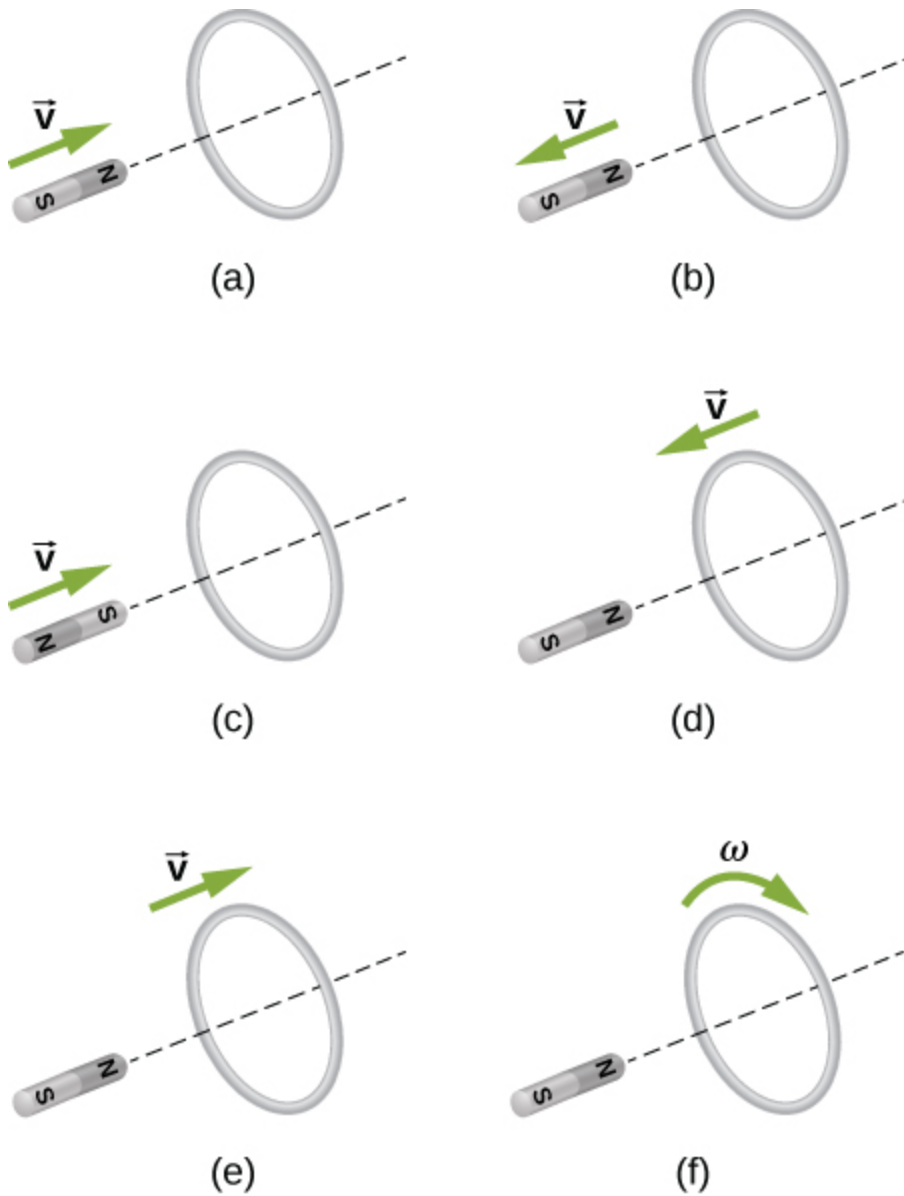
As the loop enters, the induced emf creates a CCW current while as the loop leaves the induced emf creates a CW current. While the loop is fully inside the magnetic field, there is no flux change and therefore no induced current.

Exercise:

Problem: Show that ε and $d\Phi_m/dt$ have the same units.

Exercise:**Problem:**

State the direction of the induced current for each case shown below, observing from the side of the magnet.



Solution:

a. CCW viewed from the magnet; b. CW viewed from the magnet; c. CW viewed from the magnet; d. CCW viewed from the magnet; e. CW viewed from the magnet; f. no current

Problems

Exercise:

Problem:

A single-turn circular loop of wire of radius 50 mm lies in a plane perpendicular to a spatially uniform magnetic field. During a 0.10-s time interval, the magnitude of the field increases uniformly from 200 to 300 mT. (a) Determine the emf induced in the loop. (b) If the magnetic field is directed out of the page, what is the direction of the current induced in the loop?

Solution:

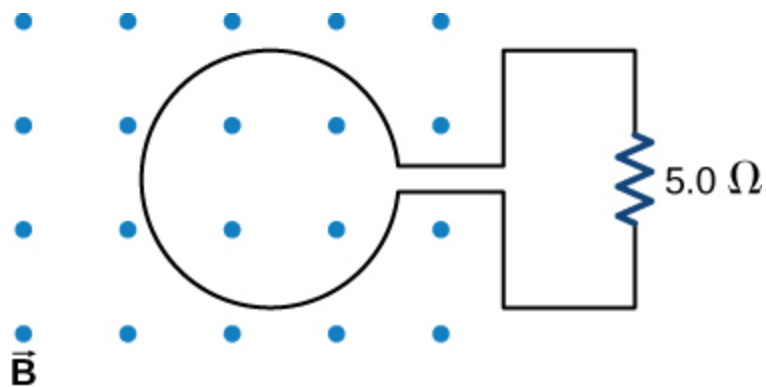
a. $7.8 \times 10^{-3} \text{ V}$; b. CCW from the same view as the magnetic field

Exercise:**Problem:**

When a magnetic field is first turned on, the flux through a 20-turn loop varies with time according to $\Phi_m = 5.0t^2 - 2.0t$, where Φ_m is in milliwebers, t is in seconds, and the loop is in the plane of the page with the unit normal pointing outward. (a) What is the emf induced in the loop as a function of time? What is the direction of the induced current at (b) $t = 0$, (c) 0.10, (d) 1.0, and (e) 2.0 s?

Exercise:**Problem:**

The magnetic flux through the loop shown in the accompanying figure varies with time according to $\Phi_m = 2.00e^{-3t}\sin(120\pi t)$, where Φ_m is in webers. What are the direction and magnitude of the current through the $5.00\text{-}\Omega$ resistor at (a) $t = 0$; (b) $t = 2.17 \times 10^{-2} \text{ s}$, and (c) $t = 3.00 \text{ s}$?



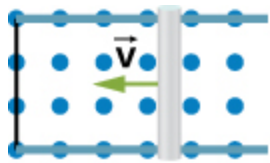
Solution:

a. 150 A downward through the resistor; b. 46 A upward through the resistor; c. 0.019 A downward through the resistor

Exercise:

Problem:

Use Lenz's law to determine the direction of induced current in each case.



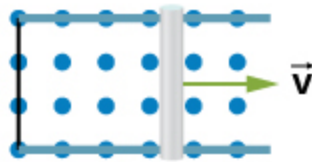
(a)



(b)



(c)

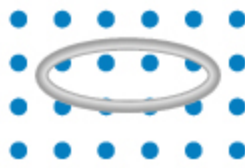


(d)



B increasing

(e)



B decreasing

(f)

Glossary

Lenz's law

direction of an induced emf opposes the change in magnetic flux that produced it; this is the negative sign in Faraday's law

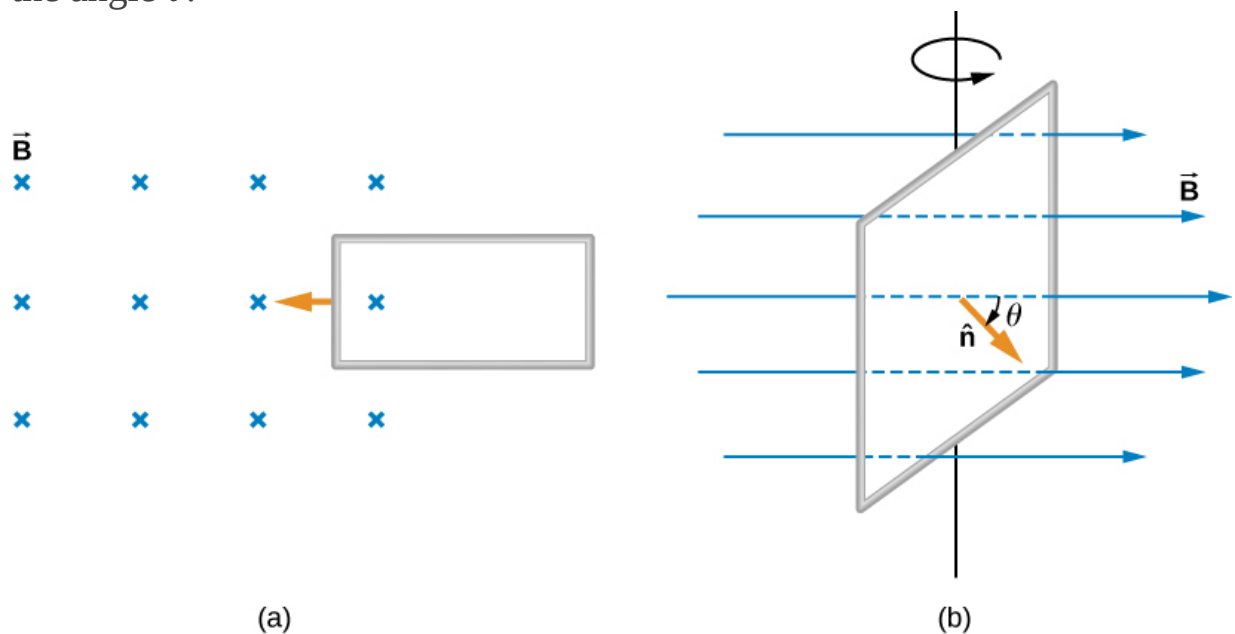
Motional Emf

By the end of this section, you will be able to:

- Determine the magnitude of an induced emf in a wire moving at a constant speed through a magnetic field
- Discuss examples that use motional emf, such as a rail gun and a tethered satellite

Magnetic flux depends on three factors: the strength of the magnetic field, the area through which the field lines pass, and the orientation of the field with the surface area. If any of these quantities varies, a corresponding variation in magnetic flux occurs. So far, we've only considered flux changes due to a changing field. Now we look at another possibility: a changing area through which the field lines pass including a change in the orientation of the area.

Two examples of this type of flux change are represented in [\[link\]](#). In part (a), the flux through the rectangular loop increases as it moves into the magnetic field, and in part (b), the flux through the rotating coil varies with the angle θ .



(a) Magnetic flux changes as a loop moves into a magnetic field; (b) magnetic flux changes as a loop rotates in a magnetic field.

It's interesting to note that what we perceive as the cause of a particular flux change actually depends on the frame of reference we choose. For example, if you are at rest relative to the moving coils of [\[link\]](#), you would see the flux vary because of a changing magnetic field—in part (a), the field moves from left to right in your reference frame, and in part (b), the field is rotating. It is often possible to describe a flux change through a coil that is moving in one particular reference frame in terms of a changing magnetic field in a second frame, where the coil is stationary. However, reference-frame questions related to magnetic flux are beyond the level of this textbook. We'll avoid such complexities by always working in a frame at rest relative to the laboratory and explain flux variations as due to either a changing field or a changing area.

Now let's look at a conducting rod pulled in a circuit, changing magnetic flux. The area enclosed by the circuit 'MNOP' of [\[link\]](#) is lx and is perpendicular to the magnetic field, so we can simplify the integration of [\[link\]](#) into a multiplication of magnetic field and area. The magnetic flux through the open surface is therefore

Equation:

$$\Phi_m = Blx.$$

Since B and l are constant and the velocity of the rod is $v = dx/dt$, we can now restate Faraday's law, [\[link\]](#), for the magnitude of the emf in terms of the moving conducting rod as

Note:

Equation:

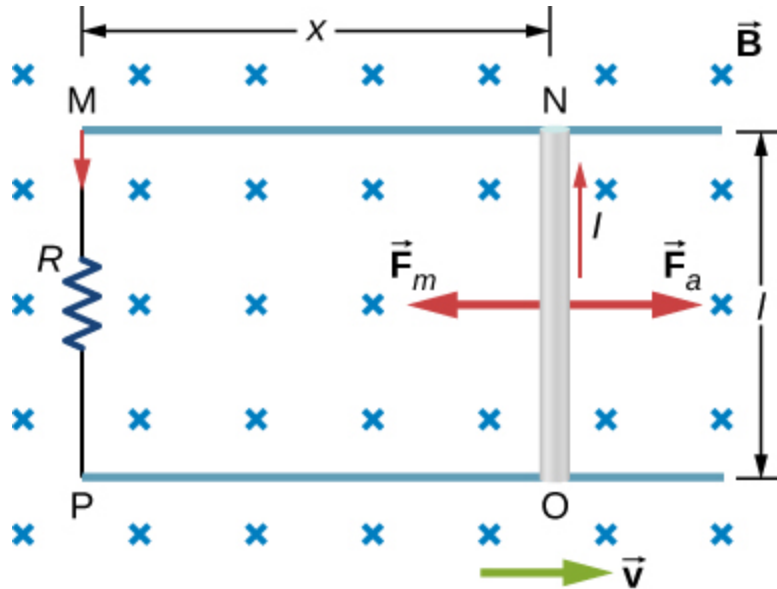
$$\varepsilon = \frac{d\Phi_m}{dt} = Bl \frac{dx}{dt} = Blv.$$

The current induced in the circuit is the emf divided by the resistance or
Equation:

$$I = \frac{Blv}{R}.$$

Furthermore, the direction of the induced emf satisfies Lenz's law, as you can verify by inspection of the figure.

This calculation of motionally induced emf is not restricted to a rod moving on conducting rails. With $\vec{F} = q\vec{v} \times \vec{B}$ as the starting point, it can be shown that $\varepsilon = -d\Phi_m/dt$ holds for any change in flux caused by the motion of a conductor. We saw in [Faraday's Law](#) that the emf induced by a time-varying magnetic field obeys this same relationship, which is Faraday's law. Thus Faraday's law *holds for all flux changes*, whether they are produced by a changing magnetic field, by motion, or by a combination of the two.



A conducting rod is pushed to the right at constant velocity. The resulting change in the magnetic flux induces a current in the circuit.

From an energy perspective, \vec{F}_a produces power $F_a v$, and the resistor dissipates power $I^2 R$. Since the rod is moving at constant velocity, the applied force F_a must balance the magnetic force $F_m = IlB$ on the rod when it is carrying the induced current I . Thus the power produced is

Equation:

$$F_a v = IlBv = \frac{Blv}{R} \cdot lBv = \frac{l^2 B^2 v^2}{R}.$$

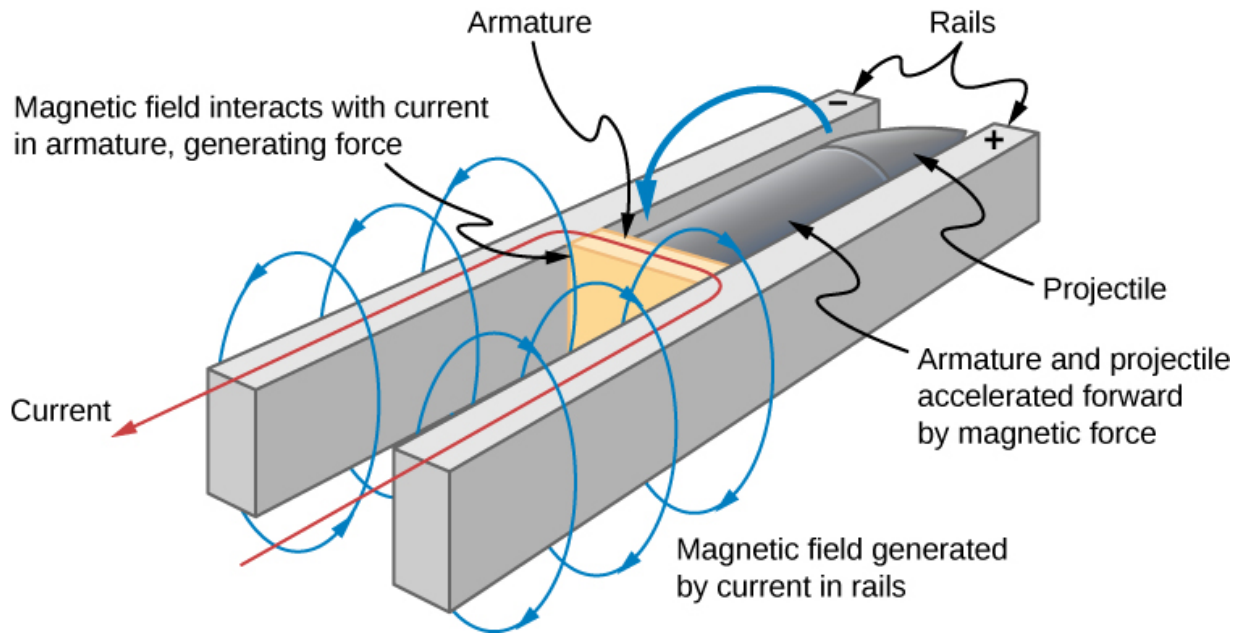
The power dissipated is

Equation:

$$P = I^2 R = \left(\frac{Blv}{R} \right)^2 R = \frac{l^2 B^2 v^2}{R}.$$

In satisfying the principle of energy conservation, the produced and dissipated powers are equal.

This principle can be seen in the operation of a rail gun. A rail gun is an electromagnetic projectile launcher that uses an apparatus similar to [\[link\]](#) and is shown in schematic form in [\[link\]](#). The conducting rod is replaced with a projectile or weapon to be fired. So far, we've only heard about how motion causes an emf. In a rail gun, the optimal shutting off/ramping down of a magnetic field decreases the flux in between the rails, causing a current to flow in the rod (armature) that holds the projectile. This current through the armature experiences a magnetic force and is propelled forward. Rail guns, however, are not used widely in the military due to the high cost of production and high currents: Nearly one million amps is required to produce enough energy for a rail gun to be an effective weapon.



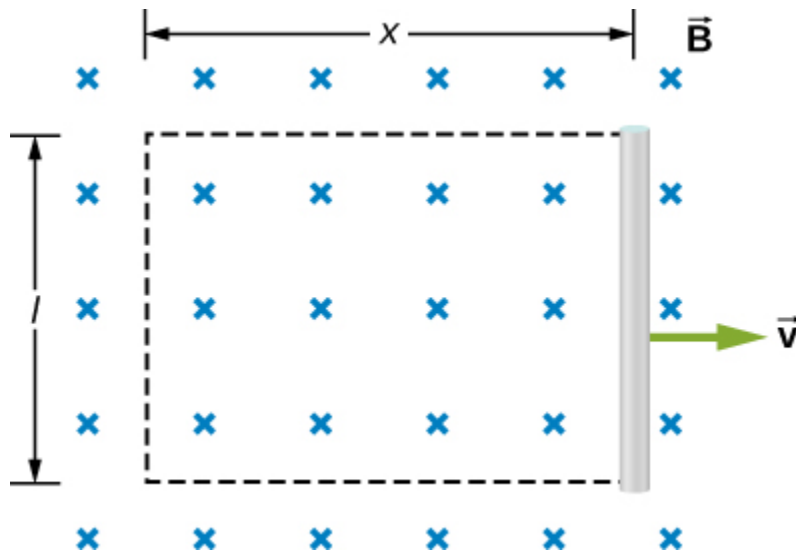
Current through two rails drives a conductive projectile forward by the magnetic force created.

We can calculate a **motionally induced emf** with Faraday's law *even when an actual closed circuit is not present*. We simply imagine an enclosed area whose boundary includes the moving conductor, calculate Φ_m , and then find the emf from Faraday's law. For example, we can let the moving rod of [\[link\]](#) be one side of the imaginary rectangular area represented by the dashed lines. The area of the rectangle is lx , so the magnetic flux through it is $\Phi_m = Blx$. Differentiating this equation, we obtain

Equation:

$$\frac{d\Phi_m}{dt} = Bl \frac{dx}{dt} = Blv,$$

which is identical to the potential difference between the ends of the rod that we determined earlier.



With the imaginary rectangle shown, we can use Faraday's law to calculate the induced emf in the moving rod.

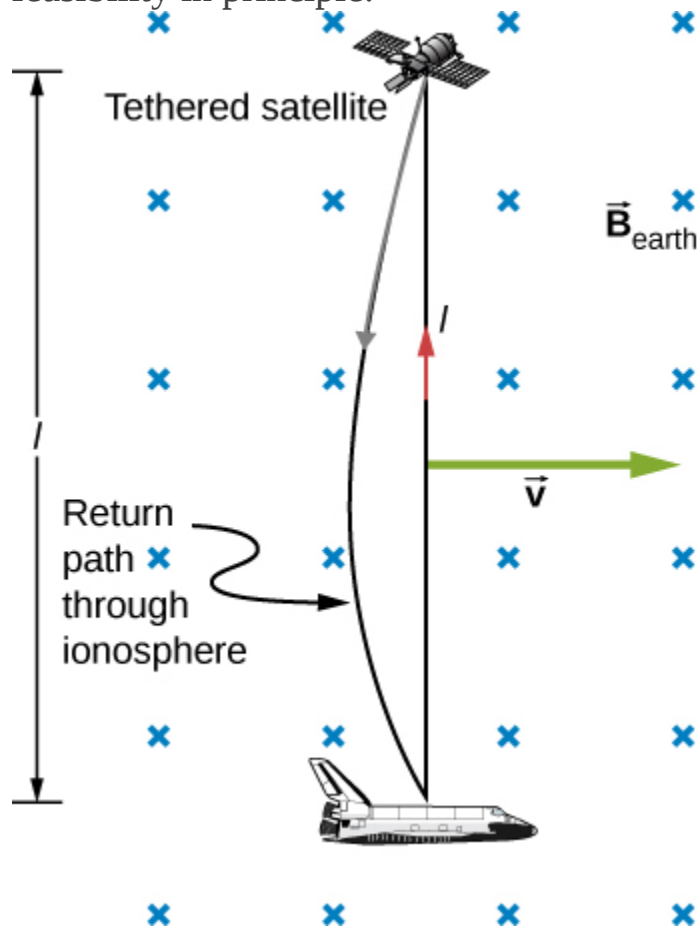
Motional emfs in Earth's weak magnetic field are not ordinarily very large, or we would notice voltage along metal rods, such as a screwdriver, during ordinary motions. For example, a simple calculation of the motional emf of a 1.0-m rod moving at 3.0 m/s perpendicular to the Earth's field gives

Equation:

$$\text{emf} = B\ell v = (5.0 \times 10^{-5} \text{ T})(1.0 \text{ m})(3.0 \text{ m/s}) = 150 \mu\text{V}.$$

This small value is consistent with experience. There is a spectacular exception, however. In 1992 and 1996, attempts were made with the space shuttle to create large motional emfs. The tethered satellite was to be let out on a 20-km length of wire, as shown in [\[link\]](#), to create a 5-kV emf by moving at orbital speed through Earth's field. This emf could be used to convert some of the shuttle's kinetic and potential energy into electrical energy if a complete circuit could be made. To complete the circuit, the stationary ionosphere was to supply a return path through which current could flow. (The ionosphere is the rarefied and partially ionized atmosphere

at orbital altitudes. It conducts because of the ionization. The ionosphere serves the same function as the stationary rails and connecting resistor in [\[link\]](#), without which there would not be a complete circuit.) Drag on the current in the cable due to the magnetic force $F = I\ell B \sin \theta$ does the work that reduces the shuttle's kinetic and potential energy, and allows it to be converted into electrical energy. Both tests were unsuccessful. In the first, the cable hung up and could only be extended a couple of hundred meters; in the second, the cable broke when almost fully extended. [\[link\]](#) indicates feasibility in principle.



Motional emf as electrical power conversion for the space shuttle was the motivation for the tethered satellite experiment. A 5-kV emf was predicted to be induced in the 20-km tether while moving at orbital speed in Earth's magnetic field. The circuit

is completed by a return path through the stationary ionosphere.

Example:

Calculating the Large Motional Emf of an Object in Orbit

Calculate the motional emf induced along a 20.0-km conductor moving at an orbital speed of 7.80 km/s perpendicular to Earth's $5.00 \times 10^{-5} \text{ T}$ magnetic field.

Strategy

This is a great example of using the equation motional $\varepsilon = B\ell v$.

Solution

Entering the given values into $\varepsilon = B\ell v$ gives

Equation:

$$\begin{aligned}\varepsilon &= B\ell v \\ &= (5.00 \times 10^{-5} \text{ T})(2.00 \times 10^4 \text{ m})(7.80 \times 10^3 \text{ m/s}) \\ &= 7.80 \times 10^3 \text{ V}.\end{aligned}$$

Significance

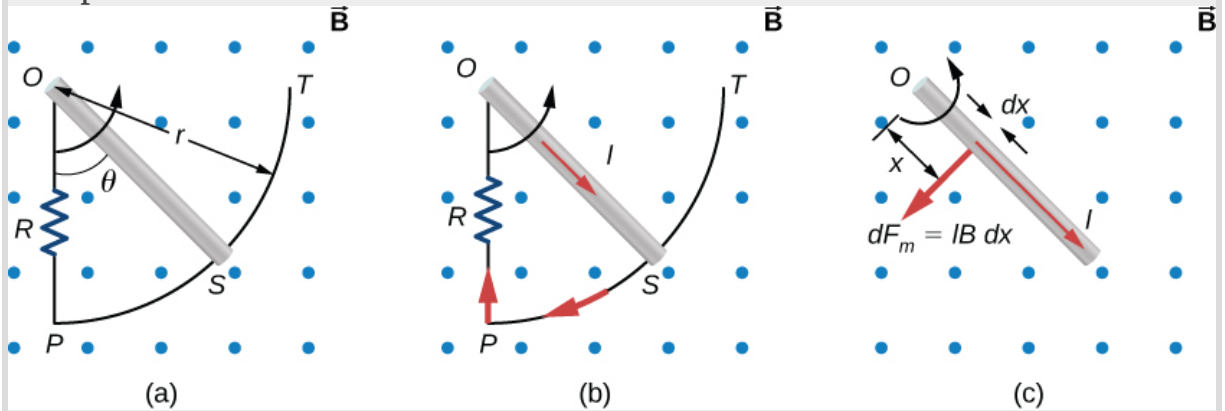
The value obtained is greater than the 5-kV measured voltage for the shuttle experiment, since the actual orbital motion of the tether is not perpendicular to Earth's field. The 7.80-kV value is the maximum emf obtained when $\theta = 90^\circ$ and so $\sin \theta = 1$.

Example:

A Metal Rod Rotating in a Magnetic Field

Part (a) of [\[link\]](#) shows a metal rod OS that is rotating in a horizontal plane around point O . The rod slides along a wire that forms a circular arc PST of radius r . The system is in a constant magnetic field \vec{B} that is directed out of the page. (a) If you rotate the rod at a constant angular velocity ω , what is the current I in the closed loop $OPSO$? Assume that the resistor R furnishes

all of the resistance in the closed loop. (b) Calculate the work per unit time that you do while rotating the rod and show that it is equal to the power dissipated in the resistor.



(a) The end of a rotating metal rod slides along a circular wire in a horizontal plane. (b) The induced current in the rod. (c) The magnetic force on an infinitesimal current segment.

Strategy

The magnetic flux is the magnetic field times the area of the quarter circle or $A = r^2\theta/2$. When finding the emf through Faraday's law, all variables are constant in time but θ , with $\omega = d\theta/dt$. To calculate the work per unit time, we know this is related to the torque times the angular velocity. The torque is calculated by knowing the force on a rod and integrating it over the length of the rod.

Solution

- a. From geometry, the area of the loop $OPSO$ is $A = \frac{r^2\theta}{2}$. Hence, the magnetic flux through the loop is

Equation:

$$\Phi_m = BA = B \frac{r^2\theta}{2}.$$

Differentiating with respect to time and using $\omega = d\theta/dt$, we have

Equation:

$$\varepsilon = \left| \frac{d\Phi_m}{dt} \right| = \frac{Br^2\omega}{2}.$$

When divided by the resistance R of the loop, this yields for the magnitude of the induced current

Equation:

$$I = \frac{\varepsilon}{R} = \frac{Br^2\omega}{2R}.$$

As θ increases, so does the flux through the loop due to \vec{B} . To counteract this increase, the magnetic field due to the induced current must be directed into the page in the region enclosed by the loop. Therefore, as part (b) of [\[link\]](#) illustrates, the current circulates clockwise.

- b. You rotate the rod by exerting a torque on it. Since the rod rotates at constant angular velocity, this torque is equal and opposite to the torque exerted on the current in the rod by the original magnetic field. The magnetic force on the infinitesimal segment of length dx shown in part (c) of [\[link\]](#) is $dF_m = IBdx$, so the magnetic torque on this segment is

Equation:

$$d\tau_m = x \cdot dF_m = IBx dx.$$

The net magnetic torque on the rod is then

Equation:

$$\tau_m = \int_0^r d\tau_m = IB \int_0^r x dx = \frac{1}{2} IB r^2.$$

The torque τ that you exert on the rod is equal and opposite to τ_m , and the work that you do when the rod rotates through an angle $d\theta$ is $dW = \tau d\theta$. Hence, the work per unit time that you do on the rod is

Equation:

$$\frac{dW}{dt} = \tau \frac{d\theta}{dt} = \frac{1}{2} I B r^2 \frac{d\theta}{dt} = \frac{1}{2} \left(\frac{B r^2 \omega}{2R} \right) B r^2 \omega = \frac{B^2 r^4 \omega^2}{4R},$$

where we have substituted for I . The power dissipated in the resistor is $P = I^2 R$, which can be written as

Equation:

$$P = \left(\frac{B r^2 \omega}{2R} \right)^2 R = \frac{B^2 r^4 \omega^2}{4R}.$$

Therefore, we see that

Equation:

$$P = \frac{dW}{dt}.$$

Hence, the power dissipated in the resistor is equal to the work per unit time done in rotating the rod.

Significance

An alternative way of looking at the induced emf from Faraday's law is to integrate in space instead of time. The solution, however, would be the same. The motional emf is

Equation:

$$|\varepsilon| = \int B v dl.$$

The velocity can be written as the angular velocity times the radius and the differential length written as dr . Therefore,

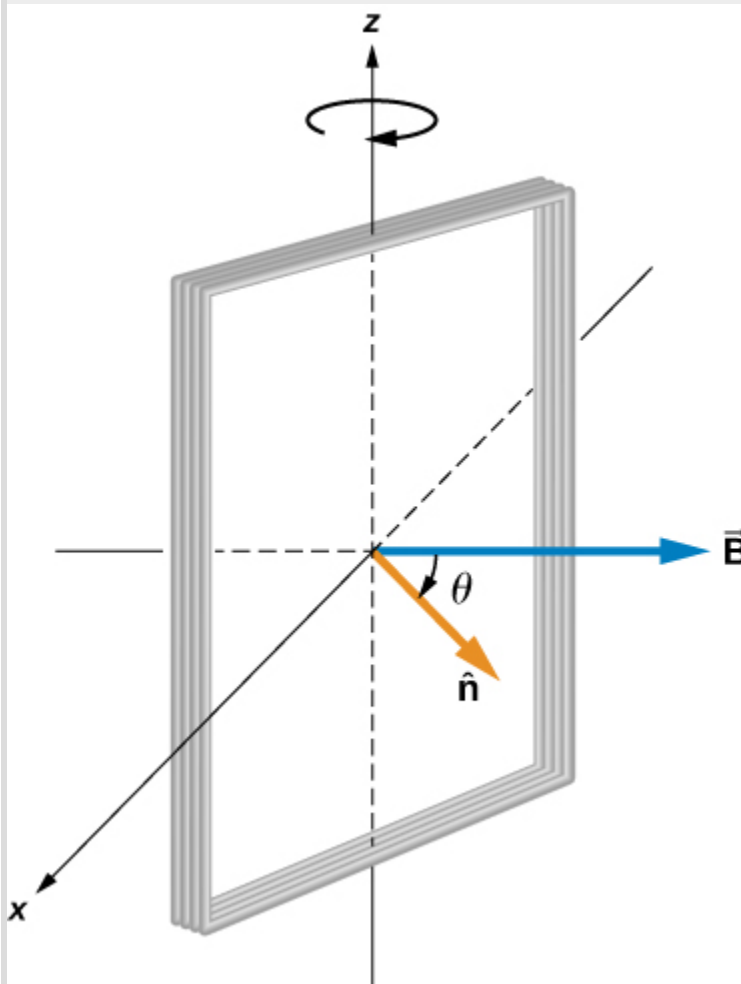
Equation:

$$|\varepsilon| = B \int v dr = B \omega \int_0^l r dr = \frac{1}{2} B \omega l^2,$$

which is the same solution as before.

Example:**A Rectangular Coil Rotating in a Magnetic Field**

A rectangular coil of area A and N turns is placed in a uniform magnetic field $\vec{B} = B\hat{j}$, as shown in [\[link\]](#). The coil is rotated about the z -axis through its center at a constant angular velocity ω . Obtain an expression for the induced emf in the coil.



A rectangular coil rotating in a uniform magnetic field.

Strategy

According to the diagram, the angle between the perpendicular to the surface (\hat{n}) and the magnetic field (\vec{B}) is θ . The dot product of $\vec{B} \cdot \hat{n}$ simplifies to only the $\cos \theta$ component of the magnetic field, namely where

the magnetic field projects onto the unit area vector \hat{n} . The magnitude of the magnetic field and the area of the loop are fixed over time, which makes the integration simplify quickly. The induced emf is written out using Faraday's law.

Solution

When the coil is in a position such that its normal vector \hat{n} makes an angle θ with the magnetic field \vec{B} , the magnetic flux through a single turn of the coil is

Equation:

$$\Phi_m = \int_S \vec{B} \cdot \hat{n} dA = BA \cos \theta.$$

From Faraday's law, the emf induced in the coil is

Equation:

$$\varepsilon = -N \frac{d\Phi_m}{dt} = NBA \sin \theta \frac{d\theta}{dt}.$$

The constant angular velocity is $\omega = d\theta/dt$. The angle θ represents the time evolution of the angular velocity or ωt . This changes the function to time space rather than θ . The induced emf therefore varies sinusoidally with time according to

Equation:

$$\varepsilon = \varepsilon_0 \sin \omega t,$$

where $\varepsilon_0 = NBA\omega$.

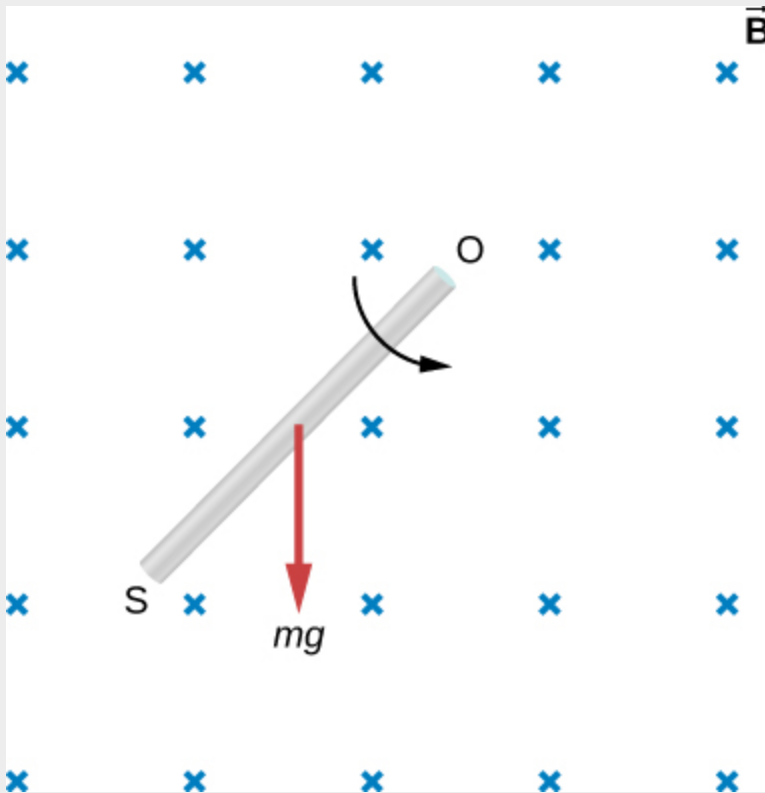
Significance

If the magnetic field strength or area of the loop were also changing over time, these variables wouldn't be able to be pulled out of the time derivative to simplify the solution as shown. This example is the basis for an electric generator, as we will give a full discussion in [Applications of Newton's Law](#).

Note:

Exercise:**Problem:**

Check Your Understanding Shown below is a rod of length l that is rotated counterclockwise around the axis through O by the torque due to $m\vec{g}$. Assuming that the rod is in a uniform magnetic field \vec{B} , what is the emf induced between the ends of the rod when its angular velocity is ω ? Which end of the rod is at a higher potential?

**Solution:**

$$\varepsilon = Bl^2\omega/2, \text{ with } O \text{ at a higher potential than } S$$

Note:**Exercise:**

Problem:

Check Your Understanding A rod of length 10 cm moves at a speed of 10 m/s perpendicularly through a 1.5-T magnetic field. What is the potential difference between the ends of the rod?

Solution:

1.5 V

Summary

- The relationship between an induced emf ε in a wire moving at a constant speed v through a magnetic field B is given by $\varepsilon = Blv$.
- An induced emf from Faraday's law is created from a motional emf that opposes the change in flux.

Conceptual Questions

Exercise:**Problem:**

A bar magnet falls under the influence of gravity along the axis of a long copper tube. If air resistance is negligible, will there be a force to oppose the descent of the magnet? If so, will the magnet reach a terminal velocity?

Exercise:**Problem:**

Around the geographic North Pole (or magnetic South Pole), Earth's magnetic field is almost vertical. If an airplane is flying northward in this region, which side of the wing is positively charged and which is negatively charged?

Solution:

Positive charges on the wings would be to the west, or to the left of the pilot while negative charges would be pulled east or to the right of the pilot. Thus, the left hand tips of the wings would be positive and the right hand tips would be negative.

Exercise:**Problem:**

A wire loop moves translationally (no rotation) in a uniform magnetic field. Is there an emf induced in the loop?

Problems**Exercise:****Problem:**

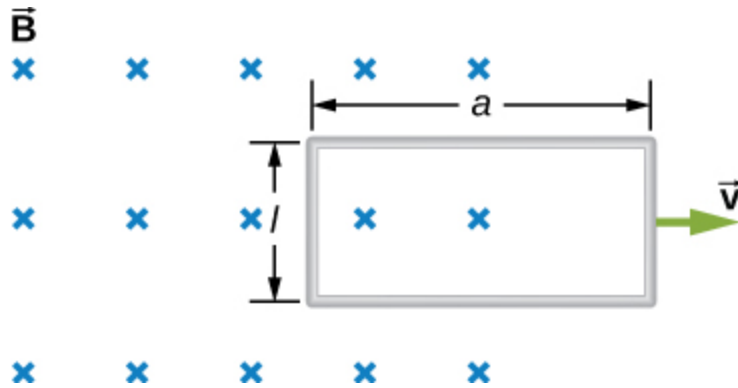
An automobile with a radio antenna 1.0 m long travels at 100.0 km/h in a location where the Earth's horizontal magnetic field is $5.5 \times 10^{-5} \text{ T}$. What is the maximum possible emf induced in the antenna due to this motion?

Solution:

0.0015 V

Exercise:**Problem:**

The rectangular loop of N turns shown below moves to the right with a constant velocity \vec{v} while leaving the poles of a large electromagnet. (a) Assuming that the magnetic field is uniform between the pole faces and negligible elsewhere, determine the induced emf in the loop. (b) What is the source of work that produces this emf?



Exercise:

Problem:

Suppose the magnetic field of the preceding problem oscillates with time according to $B = B_0 \sin \omega t$. What then is the emf induced in the loop when its trailing side is a distance d from the right edge of the magnetic field region?

Solution:

$$\varepsilon = -B_0 l d \omega \cos(\Omega t) l d + B_0 \sin(\Omega t) l v$$

Exercise:

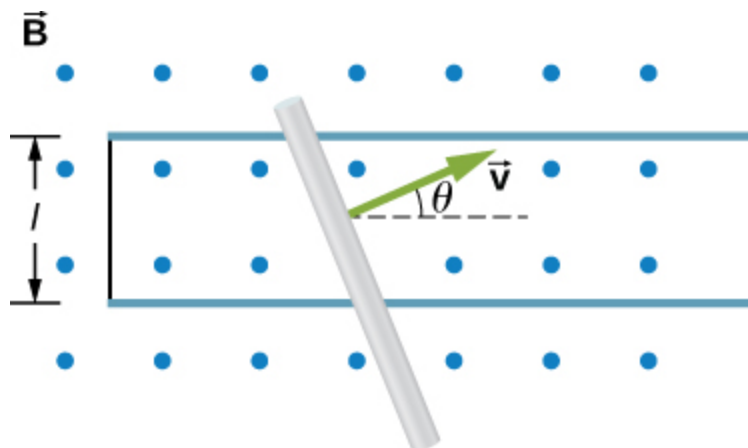
Problem:

A coil of 1000 turns encloses an area of 25 cm^2 . It is rotated in 0.010 s from a position where its plane is perpendicular to Earth's magnetic field to one where its plane is parallel to the field. If the strength of the field is $6.0 \times 10^{-5} \text{ T}$, what is the average emf induced in the coil?

Exercise:

Problem:

In the circuit shown in the accompanying figure, the rod slides along the conducting rails at a constant velocity \vec{v} . The velocity is in the same plane as the rails and directed at an angle θ to them. A uniform magnetic field \vec{B} is directed out of the page. What is the emf induced in the rod?



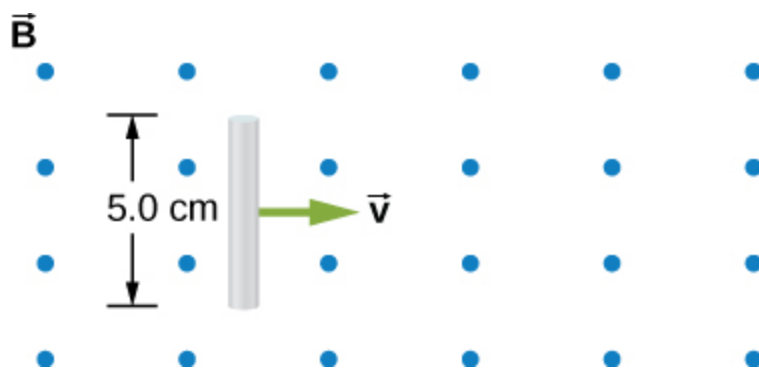
Solution:

$$\varepsilon = Blv \cos \theta$$

Exercise:

Problem:

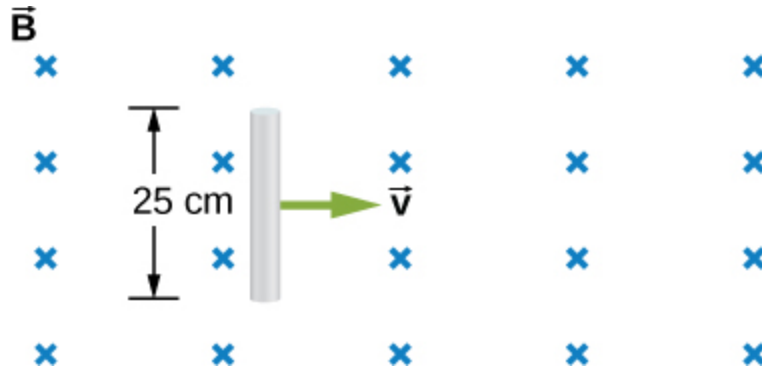
The rod shown in the accompanying figure is moving through a uniform magnetic field of strength $B = 0.50 \text{ T}$ with a constant velocity of magnitude $v = 8.0 \text{ m/s}$. What is the potential difference between the ends of the rod? Which end of the rod is at a higher potential?



Exercise:

Problem:

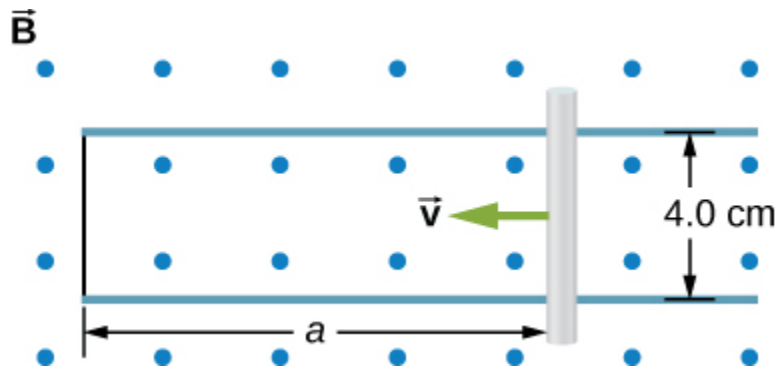
A 25-cm rod moves at 5.0 m/s in a plane perpendicular to a magnetic field of strength 0.25 T. The rod, velocity vector, and magnetic field vector are mutually perpendicular, as indicated in the accompanying figure. Calculate (a) the magnetic force on an electron in the rod, (b) the electric field in the rod, and (c) the potential difference between the ends of the rod. (d) What is the speed of the rod if the potential difference is 1.0 V?

**Solution:**

a. $2 \times 10^{-19} \text{ T}$; b. 1.25 V/m; c. 0.3125 V; d. 16 m/s

Exercise:**Problem:**

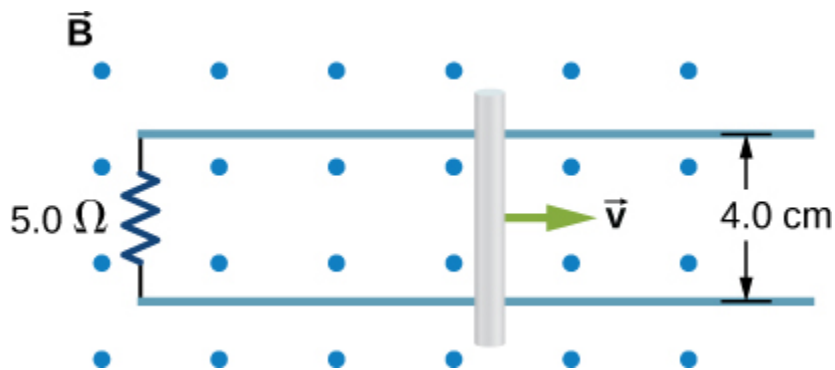
In the accompanying figure, the rails, connecting end piece, and rod all have a resistance per unit length of $2.0 \, \Omega/\text{cm}$. The rod moves to the left at $v = 3.0 \text{ m/s}$. If $B = 0.75 \text{ T}$ everywhere in the region, what is the current in the circuit (a) when $a = 8.0 \text{ cm}$? (b) when $a = 5.0 \text{ cm}$? Specify also the sense of the current flow.



Exercise:

Problem:

The rod shown below moves to the right on essentially zero-resistance rails at a speed of $v = 3.0\text{ m/s}$. If $B = 0.75\text{ T}$ everywhere in the region, what is the current through the $5.0\text{-}\Omega$ resistor? Does the current circulate clockwise or counterclockwise?



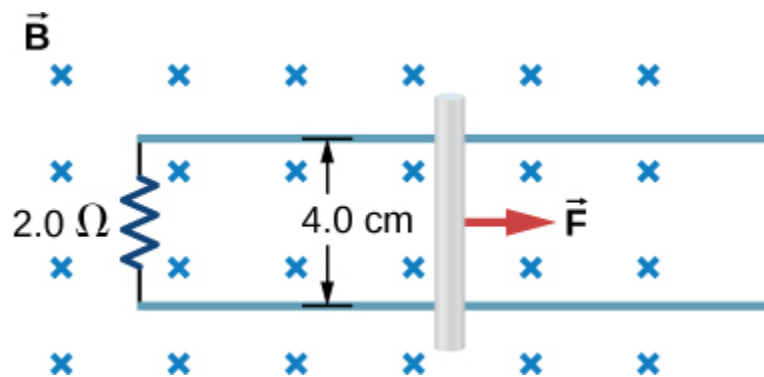
Solution:

0.018 A , CW as seen in the diagram

Exercise:

Problem:

Shown below is a conducting rod that slides along metal rails. The apparatus is in a uniform magnetic field of strength 0.25 T, which is directly into the page. The rod is pulled to the right at a constant speed of 5.0 m/s by a force \vec{F} . The only significant resistance in the circuit comes from the $2.0\text{-}\Omega$ resistor shown. (a) What is the emf induced in the circuit? (b) What is the induced current? Does it circulate clockwise or counter clockwise? (c) What is the magnitude of \vec{F} ? (d) What are the power output of \vec{F} and the power dissipated in the resistor?



Glossary

motionally induced emf

voltage produced by the movement of a conducting wire in a magnetic field

Induced Electric Fields

By the end of this section, you will be able to:

- Connect the relationship between an induced emf from Faraday's law to an electric field, thereby showing that a changing magnetic flux creates an electric field
- Solve for the electric field based on a changing magnetic flux in time

The fact that emfs are induced in circuits implies that work is being done on the conduction electrons in the wires. What can possibly be the source of this work? We know that it's neither a battery nor a magnetic field, for a battery does not have to be present in a circuit where current is induced, and magnetic fields never do work on moving charges. The answer is that the source of the work is an electric field $\vec{\mathbf{E}}$ that is induced in the wires. The work done by $\vec{\mathbf{E}}$ in moving a unit charge completely around a circuit is the induced emf ε ; that is,

Equation:

$$\varepsilon = \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}},$$

where \oint represents the line integral around the circuit. Faraday's law can be written in terms of the **induced electric field** as

Equation:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} = -\frac{d\Phi_{\text{m}}}{dt}.$$

There is an important distinction between the electric field induced by a changing magnetic field and the electrostatic field produced by a fixed charge distribution. Specifically, the induced electric field is nonconservative because it does net work in moving a charge over a closed path, whereas the electrostatic field is conservative and does no net work

over a closed path. Hence, electric potential can be associated with the electrostatic field, but not with the induced field. The following equations represent the distinction between the two types of electric field:

Equation:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} \neq 0 \text{ (induced);}$$
$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} = 0 \text{ (electrostatic).}$$

Our results can be summarized by combining these equations:

Note:

Equation:

$$\varepsilon = \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} = -\frac{d\Phi_{\text{m}}}{dt}.$$

Example:

Induced Electric Field in a Circular Coil

What is the induced electric field in the circular coil of [\[link\]](#) (and [\[link\]](#)) at the three times indicated?

Strategy

Using cylindrical symmetry, the electric field integral simplifies into the electric field times the circumference of a circle. Since we already know the induced emf, we can connect these two expressions by Faraday's law to solve for the induced electric field.

Solution

The induced electric field in the coil is constant in magnitude over the cylindrical surface, similar to how Ampere's law problems with cylinders are solved. Since $\vec{\mathbf{E}}$ is tangent to the coil,

Equation:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} = \oint E dl = 2\pi r E.$$

When combined with [\[link\]](#), this gives

Equation:

$$E = \frac{\varepsilon}{2\pi r}.$$

The direction of ε is counterclockwise, and $\vec{\mathbf{E}}$ circulates in the same direction around the coil. The values of E are

Equation:

$$\begin{aligned} E(t_1) &= \frac{6.0 \text{ V}}{2\pi (0.50 \text{ m})} = 1.9 \text{ V/m}; \\ E(t_2) &= \frac{4.7 \text{ V}}{2\pi (0.50 \text{ m})} = 1.5 \text{ V/m}; \\ E(t_3) &= \frac{0.040 \text{ V}}{2\pi (0.50 \text{ m})} = 0.013 \text{ V/m}. \end{aligned}$$

Significance

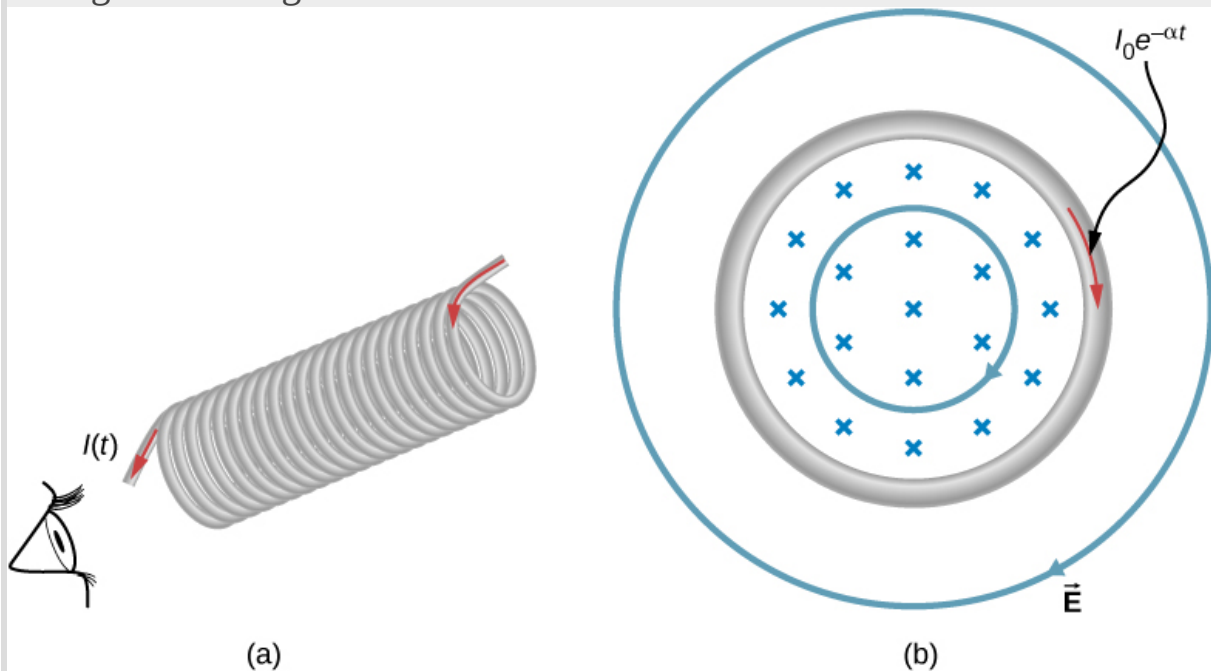
When the magnetic flux through a circuit changes, a nonconservative electric field is induced, which drives current through the circuit. But what happens if $dB/dt \neq 0$ in free space where there isn't a conducting path?

The answer is that this case can be treated *as if a conducting path were present*; that is, nonconservative electric fields are induced wherever $dB/dt \neq 0$, whether or not there is a conducting path present.

These nonconservative electric fields always satisfy [\[link\]](#). For example, if the circular coil of [\[link\]](#) were removed, an electric field *in free space* at $r = 0.50 \text{ m}$ would still be directed counterclockwise, and its magnitude would still be 1.9 V/m at $t = 0$, 1.5 V/m at $t = 5.0 \times 10^{-2} \text{ s}$, etc. The existence of induced electric fields is certainly *not* restricted to wires in circuits.

Example:**Electric Field Induced by the Changing Magnetic Field of a Solenoid**

Part (a) of [\[link\]](#) shows a long solenoid with radius R and n turns per unit length; its current decreases with time according to $I = I_0 e^{-\alpha t}$. What is the magnitude of the induced electric field at a point a distance r from the central axis of the solenoid (a) when $r > R$ and (b) when $r < R$ [see part (b) of [\[link\]](#)]. (c) What is the direction of the induced field at both locations? Assume that the infinite-solenoid approximation is valid throughout the regions of interest.



(a) The current in a long solenoid is decreasing exponentially. (b) A cross-sectional view of the solenoid from its left end. The cross-section shown is near the middle of the solenoid. An electric field is induced both inside and outside the solenoid.

Strategy

Using the formula for the magnetic field inside an infinite solenoid and Faraday's law, we calculate the induced emf. Since we have cylindrical symmetry, the electric field integral reduces to the electric field times the circumference of the integration path. Then we solve for the electric field.

Solution

- a. The magnetic field is confined to the interior of the solenoid where
Equation:

$$B = \mu_0 n I = \mu_0 n I_0 e^{-\alpha t}.$$

Thus, the magnetic flux through a circular path whose radius r is greater than R , the solenoid radius, is

Equation:

$$\Phi_m = BA = \mu_0 n I_0 \pi R^2 e^{-\alpha t}.$$

The induced field \vec{E} is tangent to this path, and because of the cylindrical symmetry of the system, its magnitude is constant on the path. Hence, we have

Equation:

$$\left| \oint \vec{E} \cdot d\vec{l} \right| = \left| \frac{d\Phi_m}{dt} \right|,$$

$$E(2\pi r) = \left| \frac{d}{dt} (\mu_0 n I_0 \pi R^2 e^{-\alpha t}) \right| = \alpha \mu_0 n I_0 \pi R^2 e^{-\alpha t},$$

$$E = \frac{\alpha \mu_0 n I_0 R^2}{2r} e^{-\alpha t} \quad (r > R).$$

- b. For a path of radius r inside the solenoid, $\Phi_m = B\pi r^2$, so
Equation:

$$E(2\pi r) = \left| \frac{d}{dt} (\mu_0 n I_0 \pi r^2 e^{-\alpha t}) \right| = \alpha \mu_0 n I_0 \pi r^2 e^{-\alpha t},$$

and the induced field is

Equation:

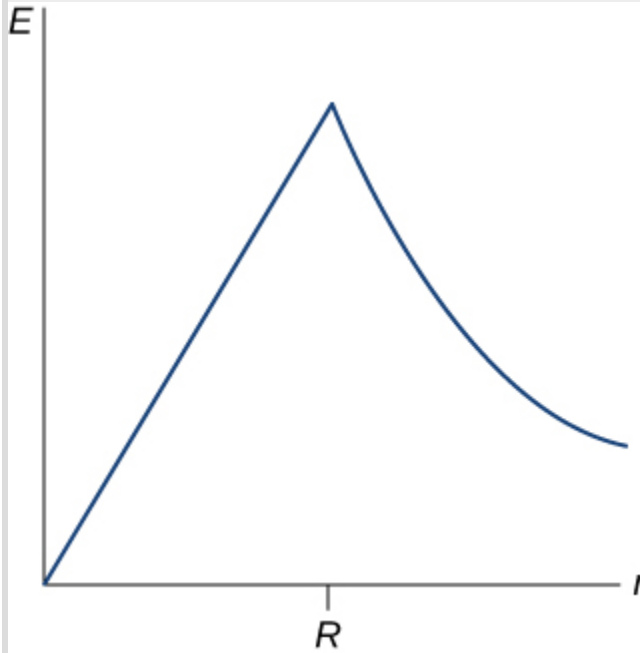
$$E = \frac{\alpha \mu_0 n I_0 r}{2} e^{-\alpha t} \quad (r < R).$$

- c. The magnetic field points into the page as shown in part (b) and is decreasing. If either of the circular paths were occupied by conducting

rings, the currents induced in them would circulate as shown, in conformity with Lenz's law. The induced electric field must be so directed as well.

Significance

In part (b), note that $|\vec{E}|$ increases with r inside and decreases as $1/r$ outside the solenoid, as shown in [\[link\]](#).



The electric field vs. distance r .
When $r < R$, the electric field rises linearly, whereas when $r > R$, the electric field falls off proportional to $1/r$.

Note:

Exercise:

Problem:

Check Your Understanding Suppose that the coil of [\[link\]](#) is a square rather than circular. Can [\[link\]](#) be used to calculate (a) the induced emf and (b) the induced electric field?

Solution:

a. yes; b. Yes; however there is a lack of symmetry between the electric field and coil, making $\oint \vec{E} \cdot d\vec{l}$ a more complicated relationship that can't be simplified as shown in the example.

Note:**Exercise:****Problem:**

Check Your Understanding What is the magnitude of the induced electric field in [\[link\]](#) at $t = 0$ if $r = 6.0$ cm, $R = 2.0$ cm, $n = 2000$ turns per meter, $I_0 = 2.0$ A, and $\alpha = 200$ s⁻¹?

Solution:

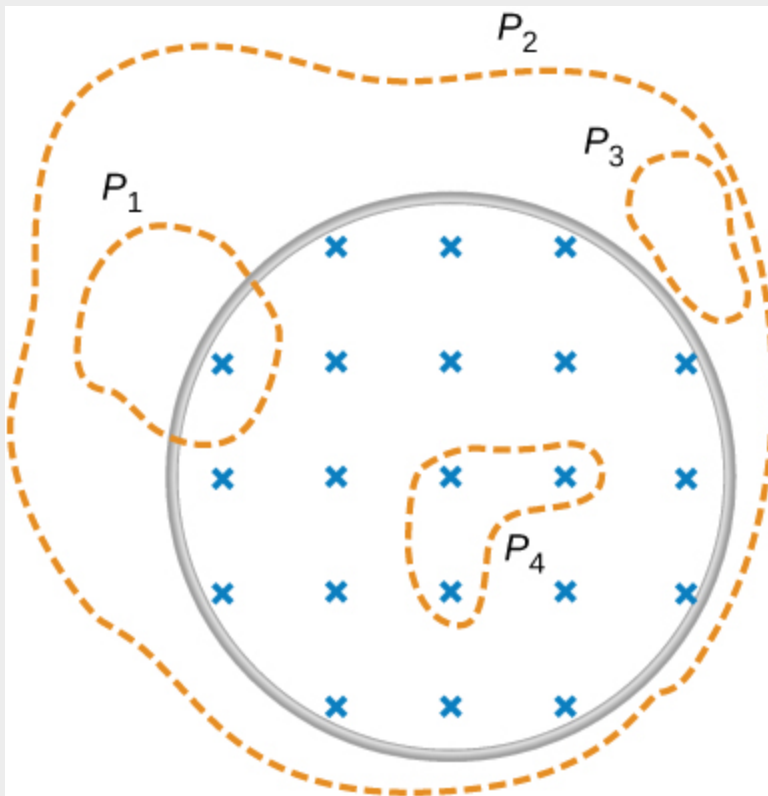
$$3.4 \times 10^{-3} \text{ V/m}$$

Note:**Exercise:**

Problem:

Check Your Understanding The magnetic field shown below is confined to the cylindrical region shown and is changing with time.

Identify those paths for which $\varepsilon = \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} \neq 0$.



Solution:

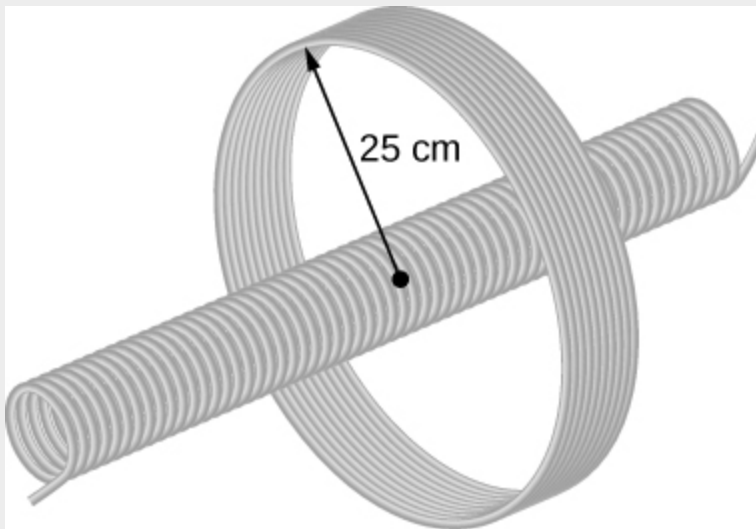
P_1, P_2, P_4

Note:

Exercise:

Problem:

Check Your Understanding A long solenoid of cross-sectional area 5.0 cm^2 is wound with 25 turns of wire per centimeter. It is placed in the middle of a closely wrapped coil of 10 turns and radius 25 cm, as shown below. (a) What is the emf induced in the coil when the current through the solenoid is decreasing at a rate $dI/dt = -0.20 \text{ A/s}$? (b) What is the electric field induced in the coil?

**Solution:**

a. $3.1 \times 10^{-6} \text{ V}$; b. $2.0 \times 10^{-7} \text{ V/m}$

Summary

- A changing magnetic flux induces an electric field.
- Both the changing magnetic flux and the induced electric field are related to the induced emf from Faraday's law.

Conceptual Questions

Exercise:

Problem:

Is the work required to accelerate a rod from rest to a speed v in a magnetic field greater than the final kinetic energy of the rod? Why?

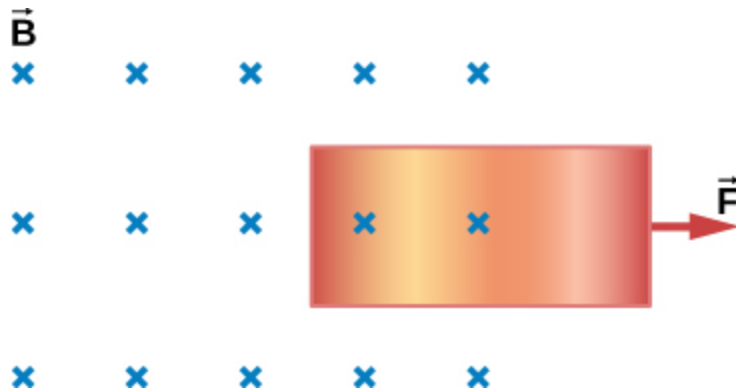
Solution:

The work is greater than the kinetic energy because it takes energy to counteract the induced emf.

Exercise:

Problem:

The copper sheet shown below is partially in a magnetic field. When it is pulled to the right, a resisting force pulls it to the left. Explain. What happens if the sheet is pushed to the left?



Problems

Exercise:

Problem:

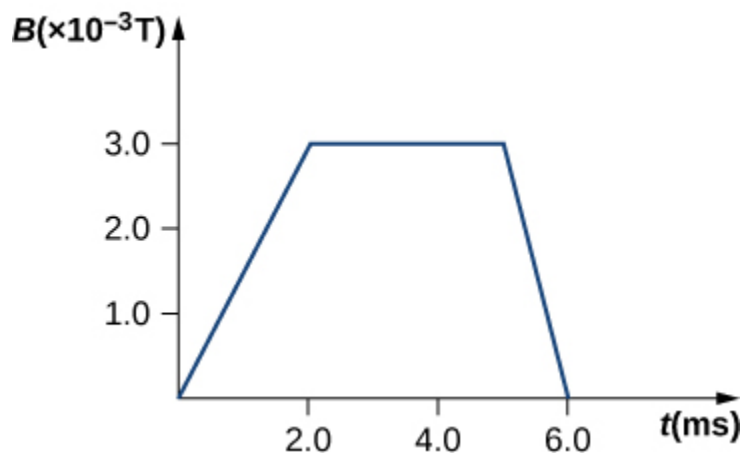
Calculate the induced electric field in a 50-turn coil with a diameter of 15 cm that is placed in a spatially uniform magnetic field of magnitude 0.50 T so that the face of the coil and the magnetic field are perpendicular. This magnetic field is reduced to zero in 0.10 seconds. Assume that the magnetic field is cylindrically symmetric with respect to the central axis of the coil.

Solution:

9.375 V/m

Exercise:**Problem:**

The magnetic field through a circular loop of radius 10.0 cm varies with time as shown in the accompanying figure. The field is perpendicular to the loop. Assuming cylindrical symmetry with respect to the central axis of the loop, plot the induced electric field in the loop as a function of time.

**Exercise:**

Problem:

The current I through a long solenoid with n turns per meter and radius R is changing with time as given by dI/dt . Calculate the induced electric field as a function of distance r from the central axis of the solenoid.

Solution:

Inside, $B = \mu_0 n I$, $\oint \vec{E} \cdot d\vec{l} = (\pi r^2) \mu_0 n \frac{dI}{dt}$, so, $E = \frac{\mu_0 n r}{2} \cdot \frac{dI}{dt}$

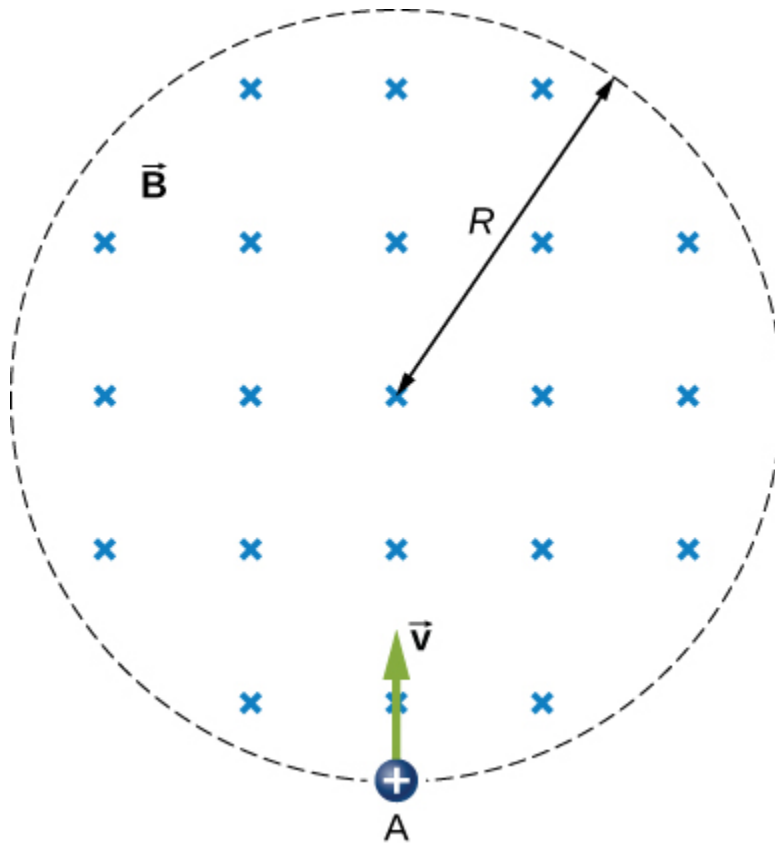
(inside). Outside, $E (2\pi r) = \pi R^2 \mu_0 n \frac{dI}{dt}$, so, $E = \frac{\mu_0 n R^2}{2r} \cdot \frac{dI}{dt}$
(outside)

Exercise:**Problem:**

Calculate the electric field induced both inside and outside the solenoid of the preceding problem if $I = I_0 \sin \omega t$.

Exercise:**Problem:**

Over a region of radius R , there is a spatially uniform magnetic field \vec{B} . (See below.) At $t = 0$, $B = 1.0$ T, after which it decreases at a constant rate to zero in 30 s. (a) What is the electric field in the regions where $r \leq R$ and $r \geq R$ during that 30-s interval? (b) Assume that $R = 10.0$ cm. How much work is done by the electric field on a proton that is carried once clock wise around a circular path of radius 5.0 cm? (c) How much work is done by the electric field on a proton that is carried once counterclockwise around a circular path of any radius $r \geq R$? (d) At the instant when $B = 0.50$ T, a proton enters the magnetic field at A, moving a velocity \vec{v} ($v = 5.0 \times 10^6$ m/s) as shown. What are the electric and magnetic forces on the proton at that instant?



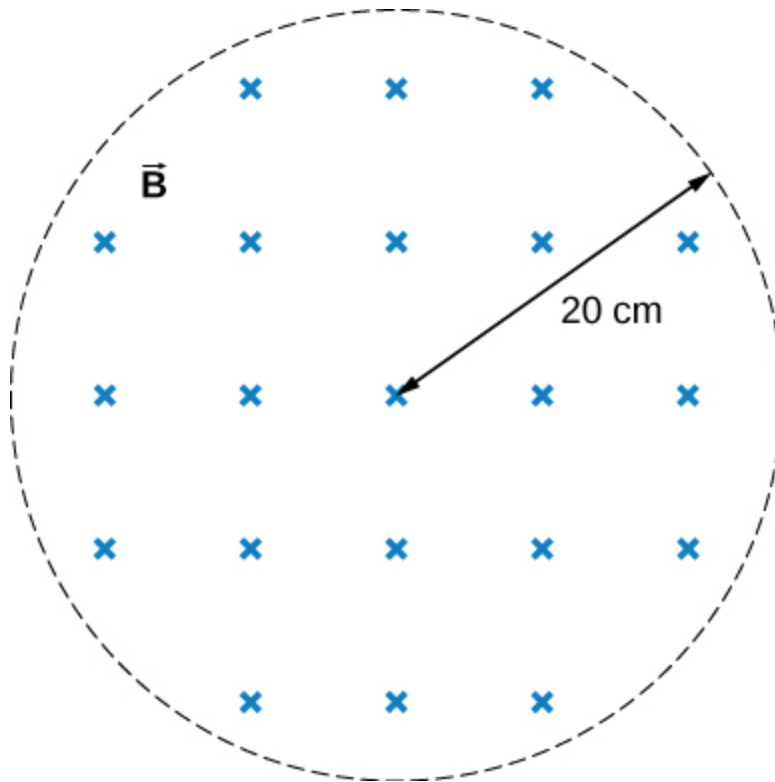
Solution:

a. $E_{\text{inside}} = -\frac{r}{2} \frac{dB}{dt}$, $E_{\text{outside}} = -\frac{dB}{dt} \frac{R^2}{2r}$; b. $W = 4.19 \times 10^{-23} \text{ J}$; c. 0 J ; d. $F_{\text{mag}} = 4 \times 10^{-13} \text{ N}$, $F_{\text{elec}} = 2.7 \times 10^{-22} \text{ N}$

Exercise:

Problem:

The magnetic field at all points within the cylindrical region whose cross-section is indicated in the accompanying figure starts at 1.0 T and decreases uniformly to zero in 20 s. What is the electric field (both magnitude and direction) as a function of r , the distance from the geometric center of the region?



Exercise:

Problem:

The current in a long solenoid with 20 turns per centimeter of radius 3 cm is varied with time at a rate of 2 A/s. A circular loop of wire of radius 5 cm and resistance $2\ \Omega$ surrounds the solenoid. Find the electrical current induced in the loop.

Solution:

$$7.1\ \mu\text{A}$$

Exercise:

Problem:

The current in a long solenoid of radius 3 cm and 20 turns/cm is varied with time at a rate of 2 A/s. Find the electric field at a distance of 4 cm from the center of the solenoid.

Glossary

induced electric field

created based on the changing magnetic flux with time

Eddy Currents

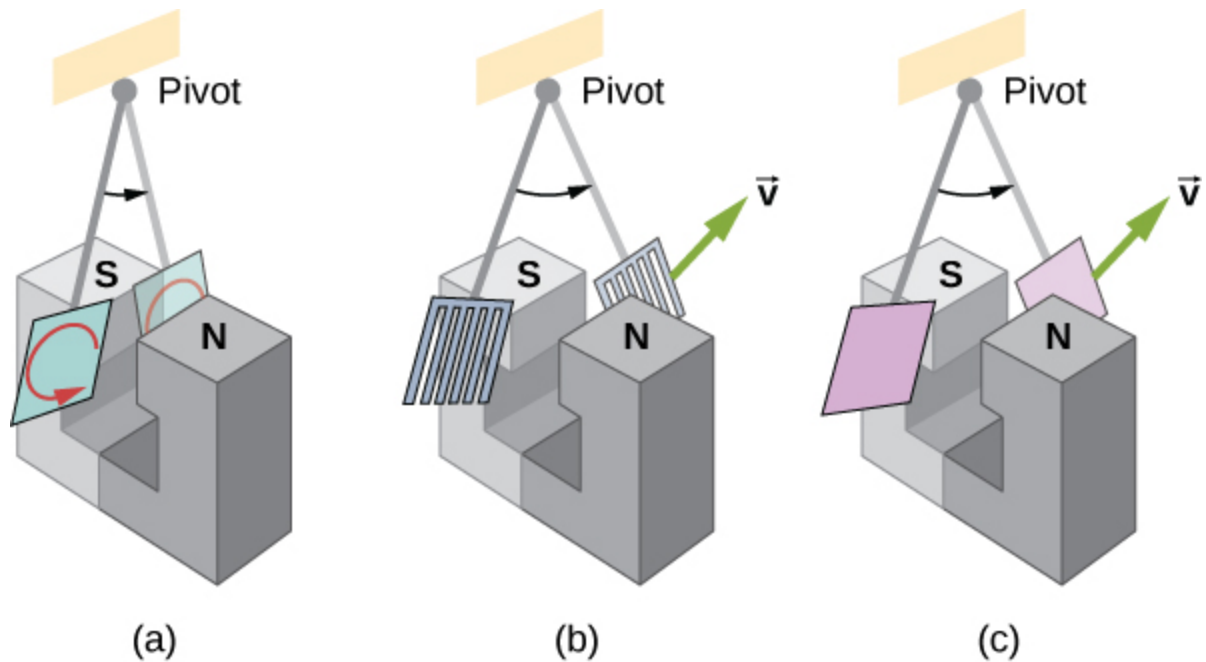
By the end of this section, you will be able to:

- Explain how eddy currents are created in metals
- Describe situations where eddy currents are beneficial and where they are not helpful

As discussed two sections earlier, a motional emf is induced when a conductor moves in a magnetic field or when a magnetic field moves relative to a conductor. If motional emf can cause a current in the conductor, we refer to that current as an **eddy current**.

Magnetic Damping

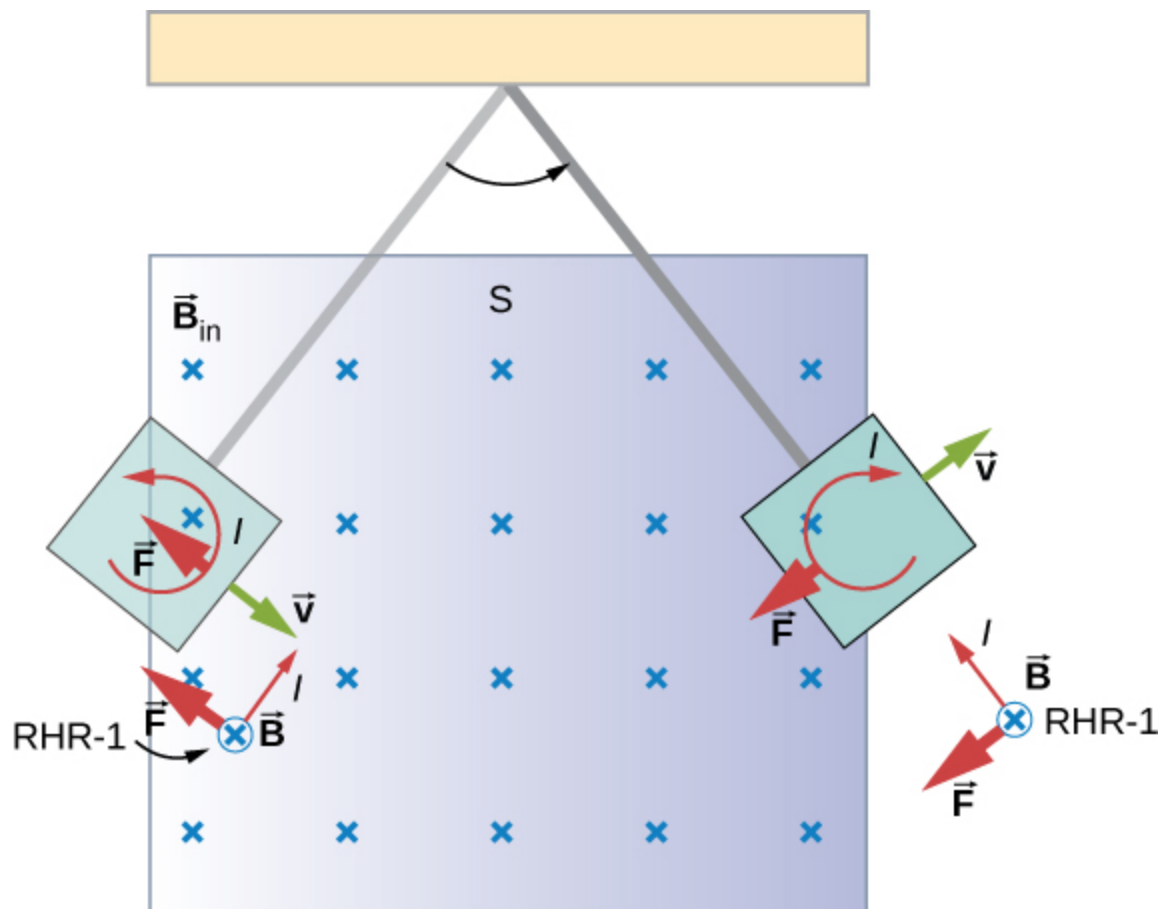
Eddy currents can produce significant drag, called **magnetic damping**, on the motion involved. Consider the apparatus shown in [\[link\]](#), which swings a pendulum bob between the poles of a strong magnet. (This is another favorite physics demonstration.) If the bob is metal, significant drag acts on the bob as it enters and leaves the field, quickly damping the motion. If, however, the bob is a slotted metal plate, as shown in part (b) of the figure, the magnet produces a much smaller effect. There is no discernible effect on a bob made of an insulator. Why does drag occur in both directions, and are there any uses for magnetic drag?



A common physics demonstration device for exploring eddy currents and magnetic damping. (a) The motion of a metal pendulum bob swinging between the poles of a magnet is quickly damped by the action of eddy currents. (b) There is little effect on the motion of a slotted metal bob, implying that eddy currents are made less effective. (c) There is also no magnetic damping on a nonconducting bob, since the eddy currents are extremely small.

[\[link\]](#) shows what happens to the metal plate as it enters and leaves the magnetic field. In both cases, it experiences a force opposing its motion. As it enters from the left, flux increases, setting up an eddy current (Faraday's law) in the counterclockwise direction (Lenz's law), as shown. Only the right-hand side of the current loop is in the field, so an unopposed force acts on it to the left (RHR-1). When the metal plate is completely inside the field, there is no eddy current if the field is uniform, since the flux remains constant in this region. But when the plate leaves the field on the right, flux decreases, causing an eddy current in the clockwise direction that, again, experiences a force to the left, further slowing the motion. A similar analysis of what happens when the plate swings from the right toward the

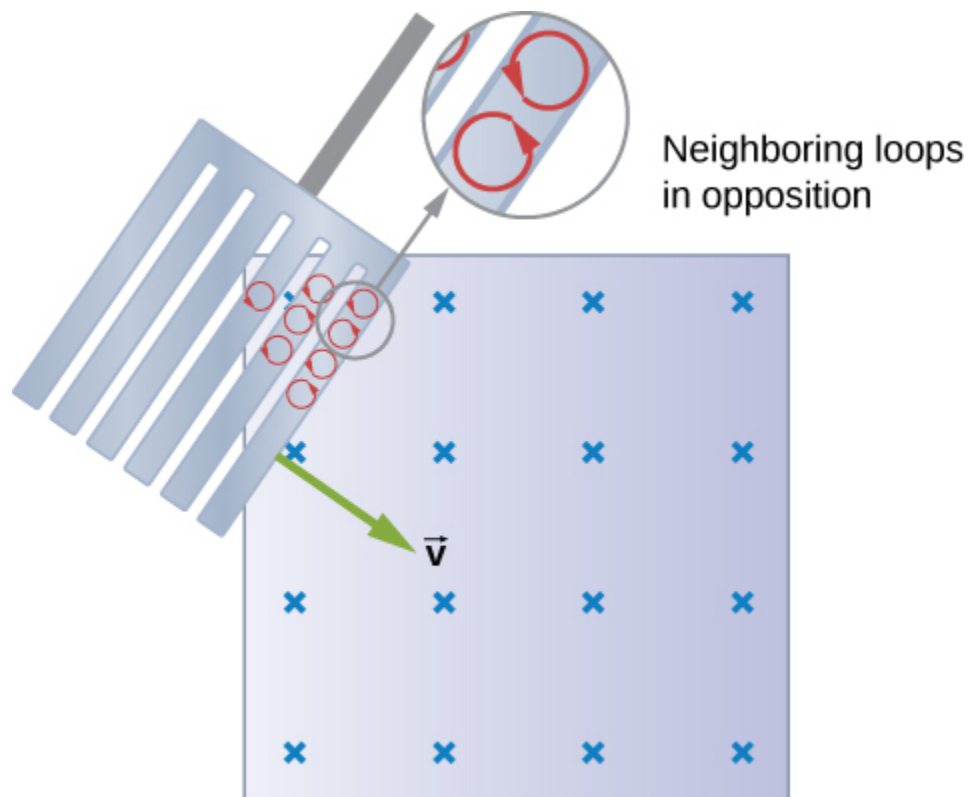
left shows that its motion is also damped when entering and leaving the field.



A more detailed look at the conducting plate passing between the poles of a magnet. As it enters and leaves the field, the change in flux produces an eddy current. Magnetic force on the current loop opposes the motion. There is no current and no magnetic drag when the plate is completely inside the uniform field.

When a slotted metal plate enters the field ([\[link\]](#)), an emf is induced by the change in flux, but it is less effective because the slots limit the size of the current loops. Moreover, adjacent loops have currents in opposite directions, and their effects cancel. When an insulating material is used, the

eddy current is extremely small, so magnetic damping on insulators is negligible. If eddy currents are to be avoided in conductors, then they must be slotted or constructed of thin layers of conducting material separated by insulating sheets.

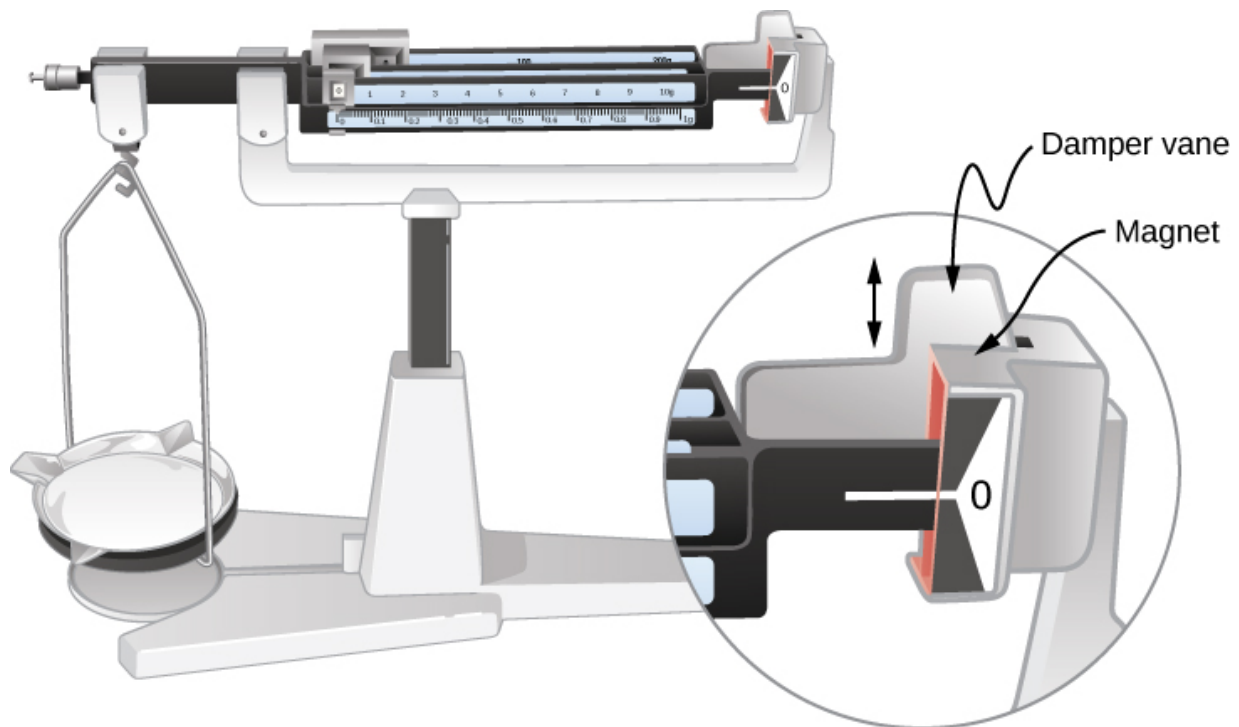


Eddy currents induced in a slotted metal plate entering a magnetic field form small loops, and the forces on them tend to cancel, thereby making magnetic drag almost zero.

Applications of Magnetic Damping

One use of magnetic damping is found in sensitive laboratory balances. To have maximum sensitivity and accuracy, the balance must be as friction-free as possible. But if it is friction-free, then it will oscillate for a very long

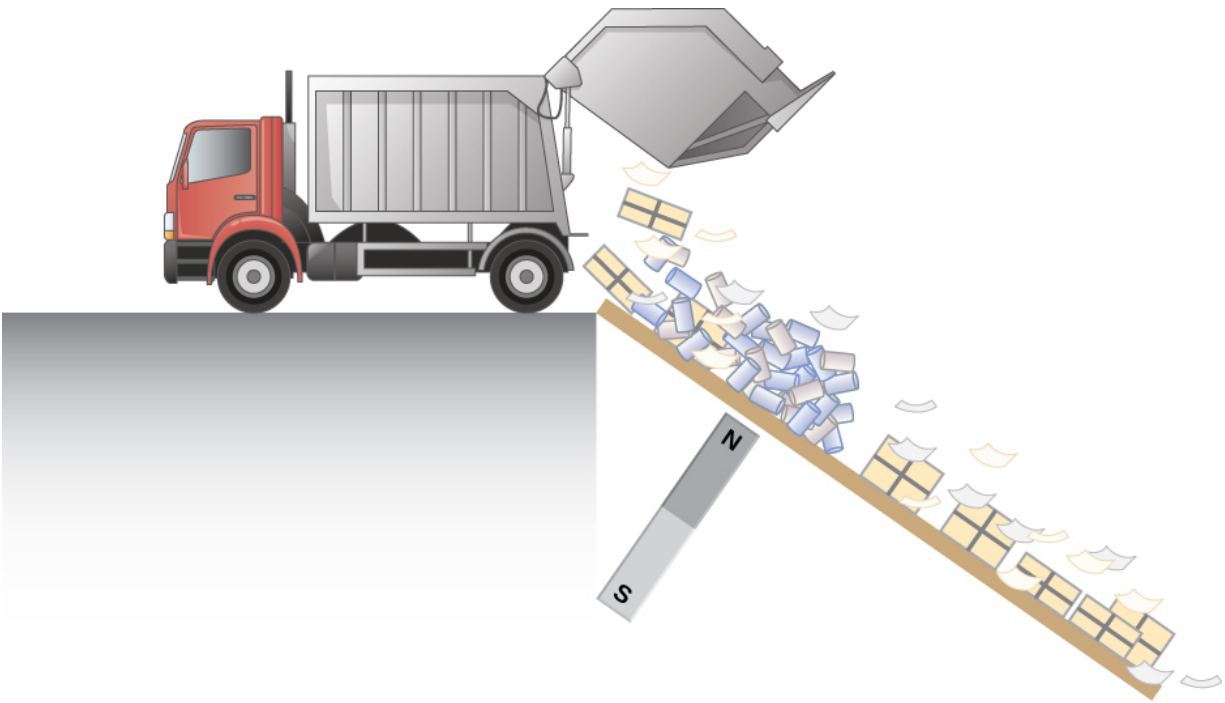
time. Magnetic damping is a simple and ideal solution. With magnetic damping, drag is proportional to speed and becomes zero at zero velocity. Thus, the oscillations are quickly damped, after which the damping force disappears, allowing the balance to be very sensitive ([\[link\]](#)). In most balances, magnetic damping is accomplished with a conducting disc that rotates in a fixed field.



Magnetic damping of this sensitive balance slows its oscillations. Since Faraday's law of induction gives the greatest effect for the most rapid change, damping is greatest for large oscillations and goes to zero as the motion stops.

Since eddy currents and magnetic damping occur only in conductors, recycling centers can use magnets to separate metals from other materials. Trash is dumped in batches down a ramp, beneath which lies a powerful magnet. Conductors in the trash are slowed by magnetic damping while nonmetals in the trash move on, separating from the metals ([\[link\]](#)). This

works for all metals, not just ferromagnetic ones. A magnet can separate out the ferromagnetic materials alone by acting on stationary trash.



Metals can be separated from other trash by magnetic drag. Eddy currents and magnetic drag are created in the metals sent down this ramp by the powerful magnet beneath it. Nonmetals move on.

Other major applications of eddy currents appear in metal detectors and braking systems in trains and roller coasters. Portable metal detectors ([link](#)) consist of a primary coil carrying an alternating current and a secondary coil in which a current is induced. An eddy current is induced in a piece of metal close to the detector, causing a change in the induced current within the secondary coil. This can trigger some sort of signal, such as a shrill noise.



A soldier in Iraq uses a metal detector to search for explosives and weapons. (credit: U.S. Army)

Braking using eddy currents is safer because factors such as rain do not affect the braking and the braking is smoother. However, eddy currents cannot bring the motion to a complete stop, since the braking force produced decreases as speed is reduced. Thus, speed can be reduced from say 20 m/s to 5 m/s, but another form of braking is needed to completely stop the vehicle. Generally, powerful rare-earth magnets such as neodymium magnets are used in roller coasters. [\[link\]](#) shows rows of magnets in such an application. The vehicle has metal fins (normally containing copper) that pass through the magnetic field, slowing the vehicle down in much the same way as with the pendulum bob shown in [\[link\]](#).



The rows of rare-earth magnets (protruding horizontally) are used for magnetic braking in roller coasters. (credit: Stefan Scheer)

Induction cooktops have electromagnets under their surface. The magnetic field is varied rapidly, producing eddy currents in the base of the pot, causing the pot and its contents to increase in temperature. Induction cooktops have high efficiencies and good response times when the base of the pot is a conductor, such as iron or steel.

Summary

- Current loops induced in moving conductors are called eddy currents. They can create significant drag, called magnetic damping.

- Manipulation of eddy currents has resulted in applications such as metal detectors, braking in trains or roller coasters, and induction cooktops.

Conceptual Questions

Exercise:

Problem:

A conducting sheet lies in a plane perpendicular to a magnetic field \vec{B} that is below the sheet. If \vec{B} oscillates at a high frequency and the conductor is made of a material of low resistivity, the region above the sheet is effectively shielded from \vec{B} . Explain why. Will the conductor shield this region from static magnetic fields?

Solution:

The conducting sheet is shielded from the changing magnetic fields by creating an induced emf. This induced emf creates an induced magnetic field that opposes any changes in magnetic fields from the field underneath. Therefore, there is no net magnetic field in the region above this sheet. If the field were due to a static magnetic field, no induced emf will be created since you need a changing magnetic flux to induce an emf. Therefore, this static magnetic field will not be shielded.

Exercise:

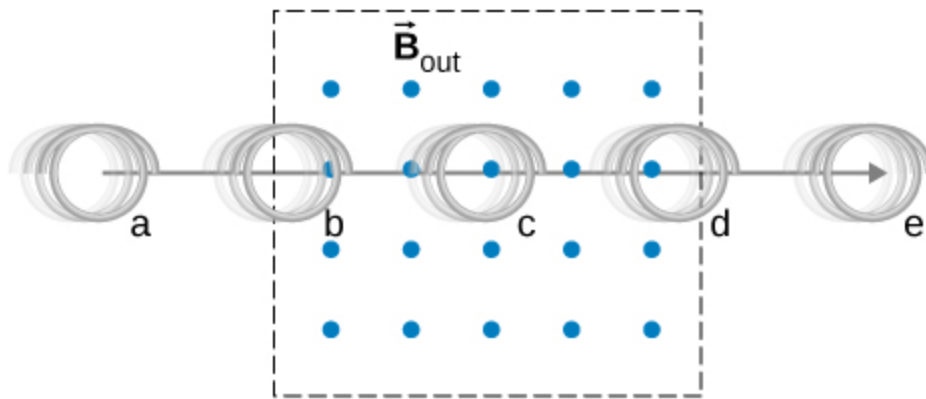
Problem:

Electromagnetic braking can be achieved by applying a strong magnetic field to a spinning metal disk attached to a shaft. (a) How can a magnetic field slow the spinning of a disk? (b) Would the brakes work if the disk was made of plastic instead of metal?

Exercise:

Problem:

A coil is moved through a magnetic field as shown below. The field is uniform inside the rectangle and zero outside. What is the direction of the induced current and what is the direction of the magnetic force on the coil at each position shown?



Solution:

a. zero induced current, zero force; b. clockwise induced current, force is to the left; c. zero induced current, zero force; d. counterclockwise induced current, force is to the left; e. zero induced current, zero force.

Glossary

magnetic damping

drag produced by eddy currents

eddy current

current loop in a conductor caused by motional emf

Electric Generators and Back Emf

By the end of this section, you will be able to:

- Explain how an electric generator works
- Determine the induced emf in a loop at any time interval, rotating at a constant rate in a magnetic field
- Show that rotating coils have an induced emf; in motors this is called back emf because it opposes the emf input to the motor

A variety of important phenomena and devices can be understood with Faraday's law. In this section, we examine two of these.

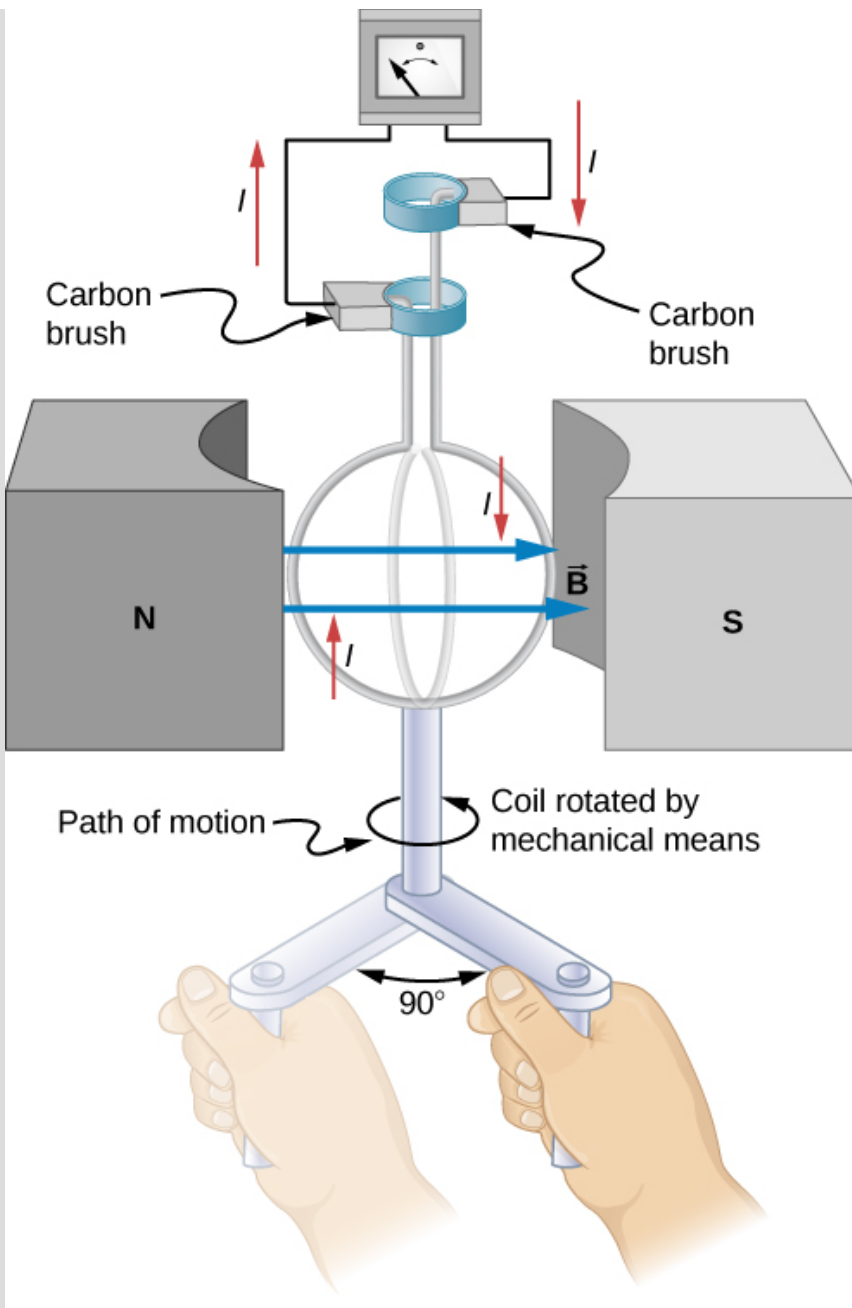
Electric Generators

Electric generators induce an emf by rotating a coil in a magnetic field, as briefly discussed in [Motional Emf](#). We now explore generators in more detail. Consider the following example.

Example:

Calculating the Emf Induced in a Generator Coil

The generator coil shown in [\[link\]](#) is rotated through one-fourth of a revolution (from $\theta = 0^\circ$ to $\theta = 90^\circ$) in 15.0 ms. The 200-turn circular coil has a 5.00-cm radius and is in a uniform 0.80-T magnetic field. What is the emf induced?



When this generator coil is rotated through one-fourth of a revolution, the magnetic flux Φ_m changes from its maximum to zero, inducing an emf.

Strategy

Faraday's law of induction is used to find the emf induced:

Equation:

$$\varepsilon = -N \frac{d\Phi_m}{dt}.$$

We recognize this situation as the same one in [\[link\]](#). According to the diagram, the projection of the surface normal vector \hat{n} to the magnetic field is initially $\cos \theta$, and this is inserted by the definition of the dot product. The magnitude of the magnetic field and area of the loop are fixed over time, which makes the integration simplify quickly. The induced emf is written out using Faraday's law:

Equation:

$$\varepsilon = NBA \sin \theta \frac{d\theta}{dt}.$$

Solution

We are given that $N = 200$, $B = 0.80 \text{ T}$, $\theta = 90^\circ$, $d\theta = 90^\circ = \pi/2$, and $dt = 15.0 \text{ ms}$. The area of the loop is

Equation:

$$A = \pi r^2 = (3.14)(0.0500 \text{ m})^2 = 7.85 \times 10^{-3} \text{ m}^2.$$

Entering this value gives

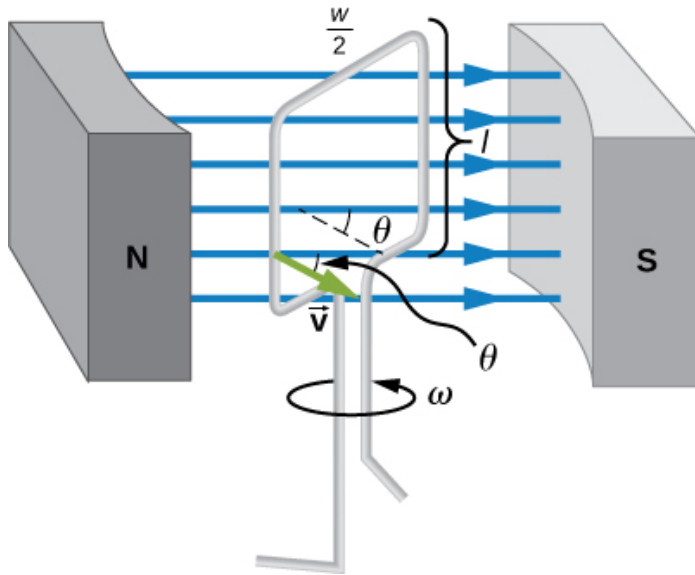
Equation:

$$\varepsilon = (200)(0.80 \text{ T})(7.85 \times 10^{-3} \text{ m}^2) \sin(90^\circ) \frac{\pi/2}{15.0 \times 10^{-3} \text{ s}} = 131 \text{ V}.$$

Significance

This is a practical average value, similar to the 120 V used in household power.

The emf calculated in [\[link\]](#) is the average over one-fourth of a revolution. What is the emf at any given instant? It varies with the angle between the magnetic field and a perpendicular to the coil. We can get an expression for emf as a function of time by considering the motional emf on a rotating rectangular coil of width w and height l in a uniform magnetic field, as illustrated in [\[link\]](#).



A generator with a single rectangular coil rotated at constant angular velocity in a uniform magnetic field produces an emf that varies sinusoidally in time. Note the generator is similar to a motor, except the shaft is rotated to produce a current rather than the other way around.

Charges in the wires of the loop experience the magnetic force, because they are moving in a magnetic field. Charges in the vertical wires experience forces parallel to the wire, causing currents. But those in the top and bottom segments feel a force perpendicular to the wire, which does not cause a current. We can thus find the induced emf by considering only the side wires. Motional emf is given to be $\varepsilon = Blv$, where the velocity v is perpendicular to the magnetic field B . Here the velocity is at an angle θ with B , so that its component perpendicular to B is $v \sin \theta$ (see [\[link\]](#)). Thus, in this case, the emf induced on each side is $\varepsilon = Blv \sin \theta$, and they are in the same direction. The total emf around the loop is then

Equation:

$$\varepsilon = 2Blv \sin \theta.$$

This expression is valid, but it does not give emf as a function of time. To find the time dependence of emf, we assume the coil rotates at a constant angular velocity ω . The angle θ is related to angular velocity by $\theta = \omega t$, so that

Equation:

$$\varepsilon = 2Blv \sin(\omega t).$$

Now, linear velocity v is related to angular velocity ω by $v = r\omega$. Here, $r = w/2$, so that $v = (w/2)\omega$, and

Equation:

$$\varepsilon = 2Bl \frac{w}{2} \omega \sin \omega t = (lw)B\omega \sin \omega t.$$

Noting that the area of the loop is $A = lw$, and allowing for N loops, we find that

Note:

Equation:

$$\varepsilon = NBA\omega \sin(\omega t).$$

This is the emf induced in a generator coil of N turns and area A rotating at a constant angular velocity ω in a uniform magnetic field B . This can also be expressed as

Equation:

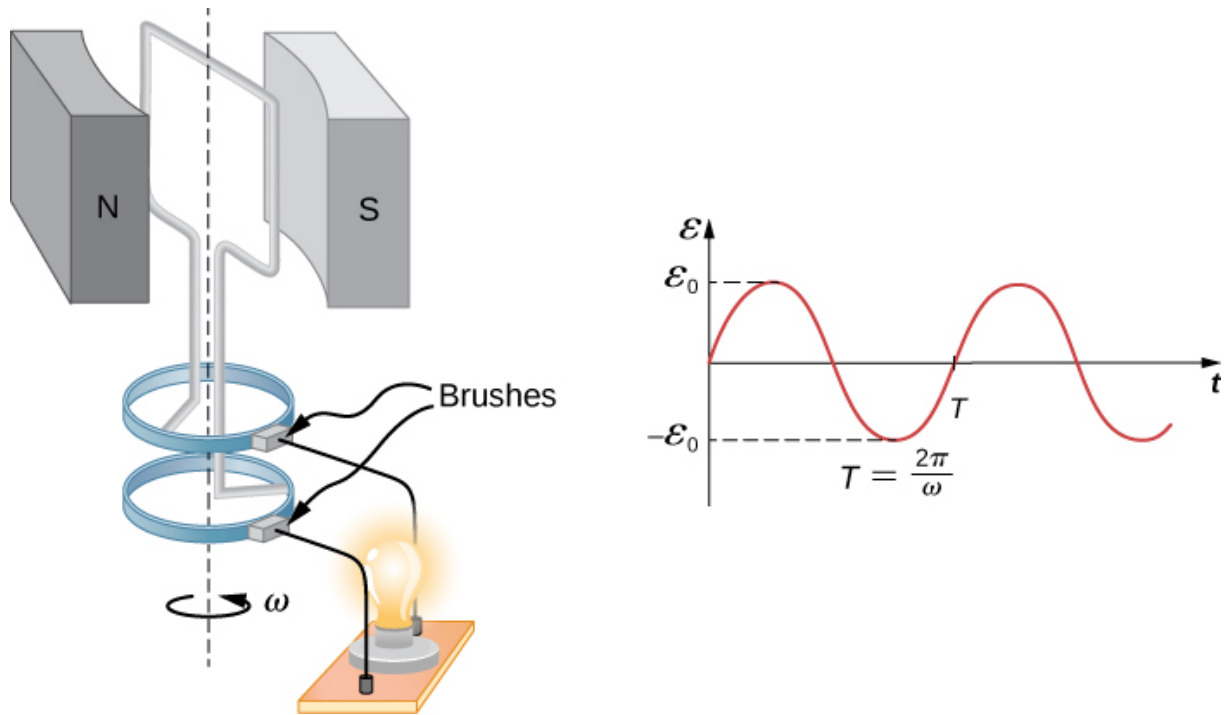
$$\varepsilon = \varepsilon_0 \sin \omega t,$$

where

Equation:

$$\varepsilon_0 = NAB\omega$$

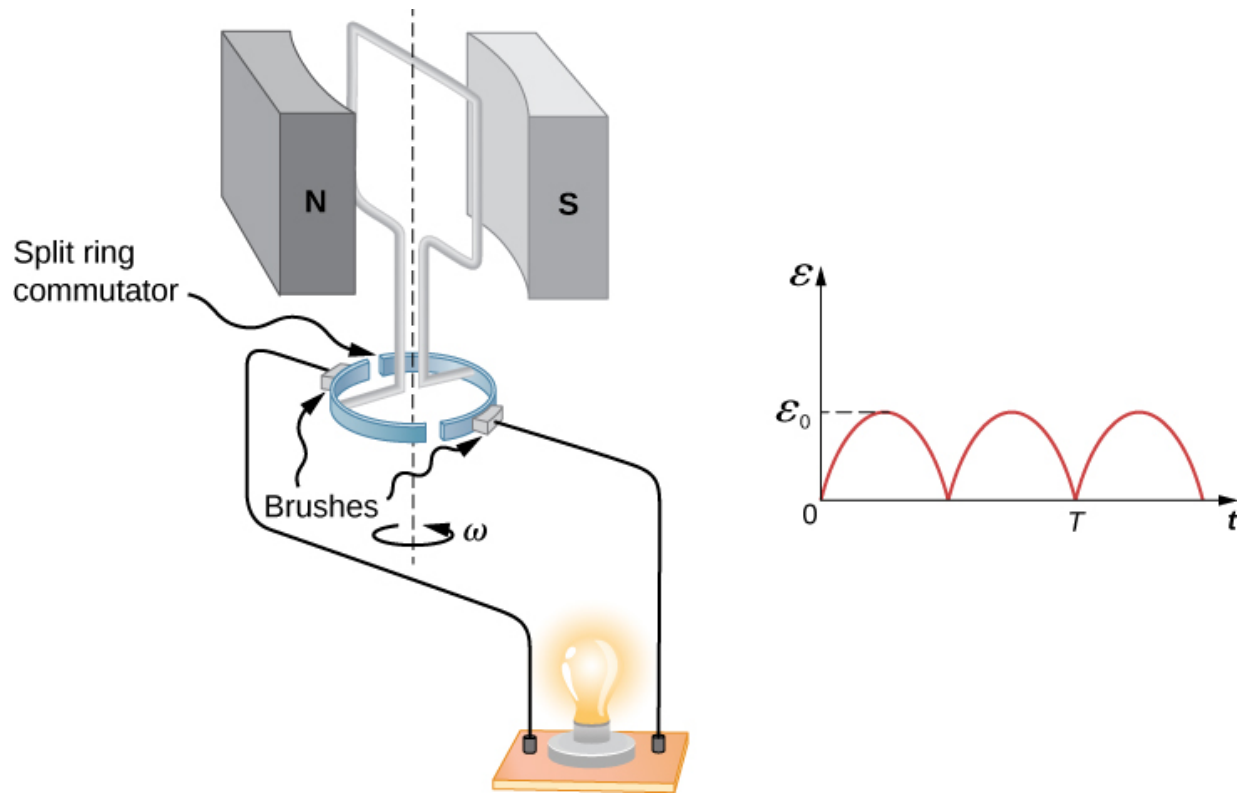
is the peak emf, since the maximum value of $\sin(\omega t) = 1$. Note that the frequency of the oscillation is $f = \omega/2\pi$ and the period is $T = 1/f = 2\pi/\omega$. [\[link\]](#) shows a graph of emf as a function of time, and it now seems reasonable that ac voltage is sinusoidal.



The emf of a generator is sent to a light bulb with the system of rings and brushes shown. The graph gives the emf of the generator as a function of time, where \mathcal{E}_0 is the peak emf. The period is $T = 1/f = 2\pi/\omega$, where f is the frequency.

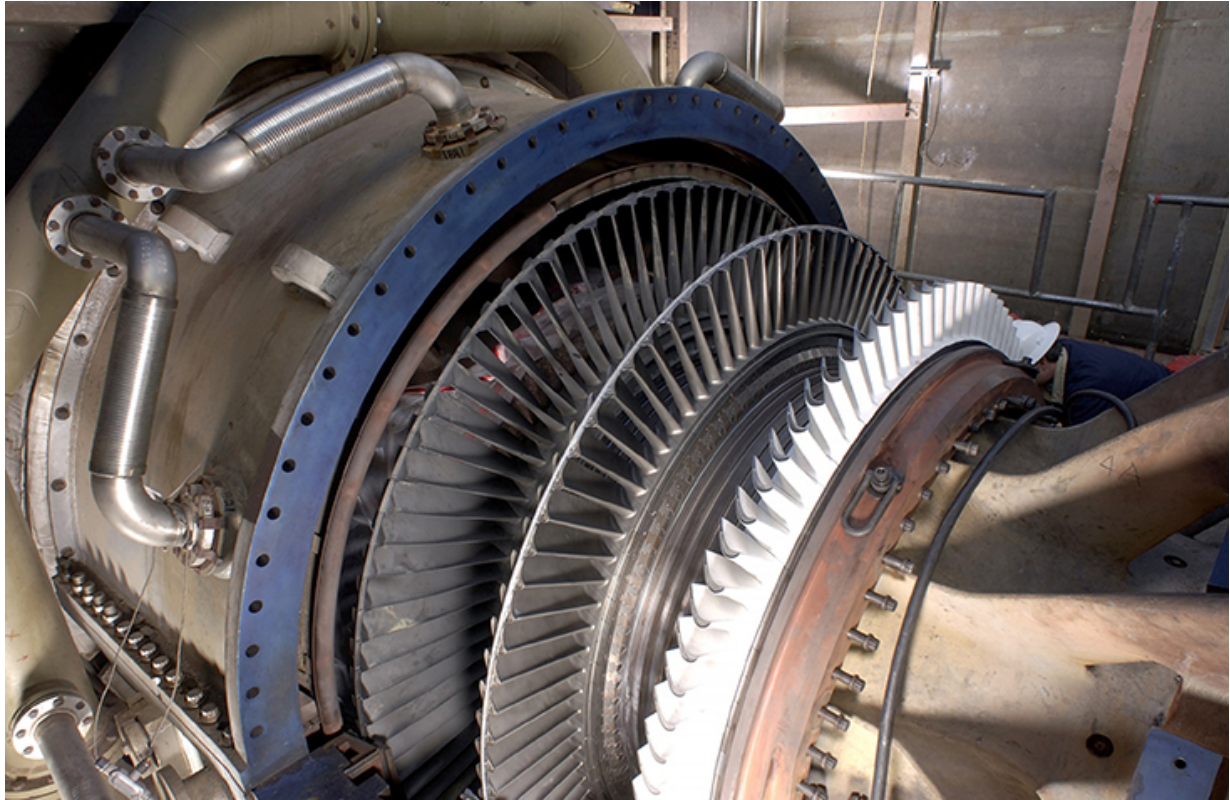
The fact that the peak emf is $\mathcal{E}_0 = NBA\omega$ makes good sense. The greater the number of coils, the larger their area, and the stronger the field, the greater the output voltage. It is interesting that the faster the generator is spun (greater ω), the greater the emf. This is noticeable on bicycle generators—at least the cheaper varieties.

[\[link\]](#) shows a scheme by which a generator can be made to produce pulsed dc. More elaborate arrangements of multiple coils and split rings can produce smoother dc, although electronic rather than mechanical means are usually used to make ripple-free dc.



Split rings, called commutators, produce a pulsed dc emf output in this configuration.

In real life, electric generators look a lot different from the figures in this section, but the principles are the same. The source of mechanical energy that turns the coil can be falling water (hydropower), steam produced by the burning of fossil fuels, or the kinetic energy of wind. [\[link\]](#) shows a cutaway view of a steam turbine; steam moves over the blades connected to the shaft, which rotates the coil within the generator. The generation of electrical energy from mechanical energy is the basic principle of all power that is sent through our electrical grids to our homes.



Steam turbine/generator. The steam produced by burning coal impacts the turbine blades, turning the shaft, which is connected to the generator.

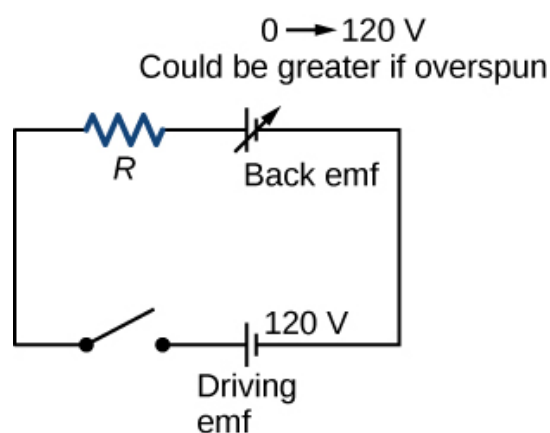
Generators illustrated in this section look very much like the motors illustrated previously. This is not coincidental. In fact, a motor becomes a generator when its shaft rotates. Certain early automobiles used their starter motor as a generator. In the next section, we further explore the action of a motor as a generator.

Back Emf

Generators convert mechanical energy into electrical energy, whereas motors convert electrical energy into mechanical energy. Thus, it is not surprising that motors and generators have the same general construction. A motor works by sending a current through a loop of wire located in a magnetic field. As a result, the magnetic field exerts torque on the loop. This rotates a shaft, thereby extracting mechanical work out of the electrical current sent in initially. (Refer to [Force and Torque on a Current Loop](#) for a discussion on motors that will help you understand more about them before proceeding.)

When the coil of a motor is turned, magnetic flux changes through the coil, and an emf (consistent with Faraday's law) is induced. The motor thus acts as a generator whenever

its coil rotates. This happens whether the shaft is turned by an external input, like a belt drive, or by the action of the motor itself. That is, when a motor is doing work and its shaft is turning, an emf is generated. Lenz's law tells us the emf opposes any change, so that the input emf that powers the motor is opposed by the motor's self-generated emf, called the **back emf** of the motor ([\[link\]](#)).



The coil of a dc motor is represented as a resistor in this schematic. The back emf is represented as a variable emf that opposes the emf driving the motor. Back emf is zero when the motor is not turning and increases proportionally to the motor's angular velocity.

The generator output of a motor is the difference between the supply voltage and the back emf. The back emf is zero when the motor is first turned on, meaning that the coil receives the full driving voltage and the motor draws maximum current when it is on but not turning. As the motor turns faster, the back emf grows, always opposing the driving emf, and reduces both the voltage across the coil and the amount of current it draws. This effect is noticeable in many common situations. When a vacuum cleaner, refrigerator, or washing machine is first turned on, lights in the same circuit dim briefly due to the IR drop produced in feeder lines by the large current drawn by the motor.

When a motor first comes on, it draws more current than when it runs at its normal operating speed. When a mechanical load is placed on the motor, like an electric wheelchair going up a hill, the motor slows, the back emf drops, more current flows,

and more work can be done. If the motor runs at too low a speed, the larger current can overheat it (via resistive power in the coil, $P = I^2 R$), perhaps even burning it out. On the other hand, if there is no mechanical load on the motor, it increases its angular velocity ω until the back emf is nearly equal to the driving emf. Then the motor uses only enough energy to overcome friction.

Eddy currents in iron cores of motors can cause troublesome energy losses. These are usually minimized by constructing the cores out of thin, electrically insulated sheets of iron. The magnetic properties of the core are hardly affected by the lamination of the insulating sheet, while the resistive heating is reduced considerably. Consider, for example, the motor coils represented in [\[link\]](#). The coils have an equivalent resistance of $0.400\ \Omega$ and are driven by an emf of $48.0\ \text{V}$. Shortly after being turned on, they draw a current

Equation:

$$I = V/R = (48.0\ \text{V}) / (0.400\ \Omega) = 120\ \text{A}$$

and thus dissipate $P = I^2 R = 5.76\ \text{kW}$ of energy as heat transfer. Under normal operating conditions for this motor, suppose the back emf is $40.0\ \text{V}$. Then at operating speed, the total voltage across the coils is $8.0\ \text{V}$ ($48.0\ \text{V}$ minus the $40.0\ \text{V}$ back emf), and the current drawn is

Equation:

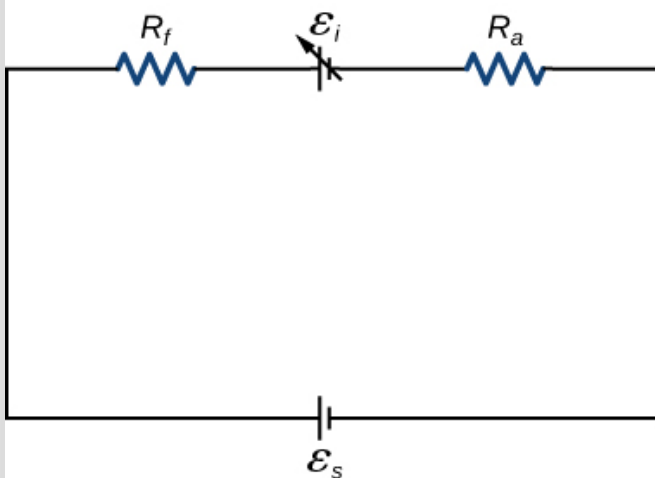
$$I = V/R = (8.0\ \text{V}) / (0.400\ \Omega) = 20\ \text{A} .$$

Under normal load, then, the power dissipated is $P = IV = (20\ \text{A})(8.0\ \text{V}) = 160\ \text{W}$. This does not cause a problem for this motor, whereas the former $5.76\ \text{kW}$ would burn out the coils if sustained.

Example:

A Series-Wound Motor in Operation

The total resistance ($R_f + R_a$) of a series-wound dc motor is $2.0\ \Omega$ ([\[link\]](#)). When connected to a 120-V source (\mathcal{E}_S), the motor draws $10\ \text{A}$ while running at constant angular velocity. (a) What is the back emf induced in the rotating coil, \mathcal{E}_i ? (b) What is the mechanical power output of the motor? (c) How much power is dissipated in the resistance of the coils? (d) What is the power output of the 120-V source? (e) Suppose the load on the motor increases, causing it to slow down to the point where it draws $20\ \text{A}$. Answer parts (a) through (d) for this situation.



Circuit representation of a series-wound direct current motor.

Strategy

The back emf is calculated based on the difference between the supplied voltage and the loss from the current through the resistance. The power from each device is calculated from one of the power formulas based on the given information.

Solution

- a. The back emf is

Equation:

$$\varepsilon_i = \varepsilon_s - I(R_f + R_a) = 120 \text{ V} - (10 \text{ A})(2.0 \Omega) = 100 \text{ V}.$$

- b. Since the potential across the armature is 100 V when the current through it is 10 A, the power output of the motor is

Equation:

$$P_m = \varepsilon_i I = (100 \text{ V})(10 \text{ A}) = 1.0 \times 10^3 \text{ W}.$$

- c. A 10-A current flows through coils whose combined resistance is 2.0Ω , so the power dissipated in the coils is

Equation:

$$P_R = I^2 R = (10 \text{ A})^2 (2.0 \Omega) = 2.0 \times 10^2 \text{ W}.$$

- d. Since 10 A is drawn from the 120-V source, its power output is

Equation:

$$P_s = \varepsilon_s I = (120 \text{ V})(10 \text{ A}) = 1.2 \times 10^3 \text{ W}.$$

e. Repeating the same calculations with $I = 20 \text{ A}$, we find

Equation:

$$\varepsilon_i = 80 \text{ V}, P_m = 1.6 \times 10^3 \text{ W}, P_R = 8.0 \times 10^2 \text{ W}, \text{ and } P_s = 2.4 \times 10^3 \text{ W}.$$

The motor is turning more slowly in this case, so its power output and the power of the source are larger.

Significance

Notice that we have an energy balance in part (d):

$$1.2 \times 10^3 \text{ W} = 1.0 \times 10^3 \text{ W} + 2.0 \times 10^2 \text{ W}.$$

Summary

- An electric generator rotates a coil in a magnetic field, inducing an emf given as a function of time by $\varepsilon = NBA\omega \sin(\omega t)$ where A is the area of an N -turn coil rotated at a constant angular velocity ω in a uniform magnetic field \vec{B} .
- The peak emf of a generator is $\varepsilon_0 = NBA\omega$.
- Any rotating coil produces an induced emf. In motors, this is called back emf because it opposes the emf input to the motor.

Problems

Exercise:

Problem:

Design a current loop that, when rotated in a uniform magnetic field of strength 0.10 T , will produce an emf $\varepsilon = \varepsilon_0 \sin \omega t$, where $\varepsilon_0 = 110 \text{ V}$ and $\omega = 120\pi \text{ rad/s}$.

Solution:

three turns with an area of 1 m^2

Exercise:

Problem:

A flat, square coil of 20 turns that has sides of length 15.0 cm is rotating in a magnetic field of strength 0.050 T. If the maximum emf produced in the coil is 30.0 mV, what is the angular velocity of the coil?

Exercise:**Problem:**

A 50-turn rectangular coil with dimensions $0.15\text{ m} \times 0.40\text{ m}$ rotates in a uniform magnetic field of magnitude 0.75 T at 3600 rev/min. (a) Determine the emf induced in the coil as a function of time. (b) If the coil is connected to a $1000\text{-}\Omega$ resistor, what is the power as a function of time required to keep the coil turning at 3600 rpm? (c) Answer part (b) if the coil is connected to a $2000\text{-}\Omega$ resistor.

Solution:

- a. $\omega = 120\pi\text{ rad/s},$
 $\varepsilon = 850 \sin 120 \pi t\text{ V};$
- b. $P = 720 \sin^2 120 \pi t\text{ W};$
- c. $P = 360 \sin^2 120 \pi t\text{ W}$

Exercise:**Problem:**

The square armature coil of an alternating current generator has 200 turns and is 20.0 cm on side. When it rotates at 3600 rpm, its peak output voltage is 120 V. (a) What is the frequency of the output voltage? (b) What is the strength of the magnetic field in which the coil is turning?

Exercise:**Problem:**

A flip coil is a relatively simple device used to measure a magnetic field. It consists of a circular coil of N turns wound with fine conducting wire. The coil is attached to a ballistic galvanometer, a device that measures the total charge that passes through it. The coil is placed in a magnetic field \vec{B} such that its face is perpendicular to the field. It is then flipped through 180° , and the total charge Q that flows through the galvanometer is measured. (a) If the total resistance of the coil and galvanometer is R , what is the relationship between B and Q ? Because the coil is very small, you can assume that \vec{B} is uniform over it. (b) How can you determine whether or not the magnetic field is perpendicular to the face of the coil?

Solution:

a. B is proportional to Q ; b. If the coin turns easily, the magnetic field is perpendicular. If the coin is at an equilibrium position, it is parallel.

Exercise:**Problem:**

The flip coil of the preceding problem has a radius of 3.0 cm and is wound with 40 turns of copper wire. The total resistance of the coil and ballistic galvanometer is $0.20\ \Omega$. When the coil is flipped through 180° in a magnetic field \vec{B} , a change of 0.090 C flows through the ballistic galvanometer. (a) Assuming that \vec{B} and the face of the coil are initially perpendicular, what is the magnetic field? (b) If the coil is flipped through 90° , what is the reading of the galvanometer?

Exercise:**Problem:**

A 120-V, series-wound motor has a field resistance of $80\ \Omega$ and an armature resistance of $10\ \Omega$. When it is operating at full speed, a back emf of 75 V is generated. (a) What is the initial current drawn by the motor? When the motor is operating at full speed, where are (b) the current drawn by the motor, (c) the power output of the source, (d) the power output of the motor, and (e) the power dissipated in the two resistances?

Solution:

a. 1.33 A; b. 0.50 A; c. 60 W; d. 37.5 W; e. 22.5W

Exercise:**Problem:**

A small series-wound dc motor is operated from a 12-V car battery. Under a normal load, the motor draws 4.0 A, and when the armature is clamped so that it cannot turn, the motor draws 24 A. What is the back emf when the motor is operating normally?

Glossary**back emf**

emf generated by a running motor, because it consists of a coil turning in a magnetic field; it opposes the voltage powering the motor

peak emf

maximum emf produced by a generator

electric generator

device for converting mechanical work into electric energy; it induces an emf by rotating a coil in a magnetic field

Applications of Electromagnetic Induction

By the end of this section, you will be able to:

- Explain how computer hard drives and graphic tablets operate using magnetic induction
- Explain how hybrid/electric vehicles and transcranial magnetic stimulation use magnetic induction to their advantage

Modern society has numerous applications of Faraday's law of induction, as we will explore in this chapter and others. At this juncture, let us mention several that involve recording information using magnetic fields.

Some computer hard drives apply the principle of magnetic induction. Recorded data are made on a coated, spinning disk. Historically, reading these data was made to work on the principle of induction. However, most input information today is carried in digital rather than analog form—a series of 0s or 1s are written upon the spinning hard drive. Therefore, most hard drive readout devices do not work on the principle of induction, but use a technique known as giant magnetoresistance. Giant magnetoresistance is the effect of a large change of electrical resistance induced by an applied magnetic field to thin films of alternating ferromagnetic and nonmagnetic layers. This is one of the first large successes of nanotechnology.

Graphics tablets, or tablet computers where a specially designed pen is used to draw digital images, also applies induction principles. The tablets discussed here are labeled as passive tablets, since there are other designs that use either a battery-operated pen or optical signals to write with. The passive tablets are different than the touch tablets and phones many of us use regularly, but may still be found when signing your signature at a cash register. Underneath the screen, shown in [\[link\]](#), are tiny wires running across the length and width of the screen. The pen has a tiny magnetic field coming from the tip. As the tip brushes across the screen, a changing magnetic field is felt in the wires which translates into an induced emf that is converted into the line you just drew.



A tablet with a specially designed pen to write with is another application of magnetic induction. (credit: Jane Whitney)

Another application of induction is the magnetic stripe on the back of your personal credit card as used at the grocery store or the ATM machine. This works on the same principle as the audio or video tape, in which a playback head reads personal information from your card.

Note:

Check out this [video](#) to see how flashlights can use magnetic induction. A magnet moves by your mechanical work through a wire. The induced

current charges a capacitor that stores the charge that will light the lightbulb even while you are not doing this mechanical work.

Electric and hybrid vehicles also take advantage of electromagnetic induction. One limiting factor that inhibits widespread acceptance of 100% electric vehicles is that the lifetime of the battery is not as long as the time you get to drive on a full tank of gas. To increase the amount of charge in the battery during driving, the motor can act as a generator whenever the car is braking, taking advantage of the back emf produced. This extra emf can be newly acquired stored energy in the car's battery, prolonging the life of the battery.

Another contemporary area of research in which electromagnetic induction is being successfully implemented is transcranial magnetic stimulation (TMS). A host of disorders, including depression and hallucinations, can be traced to irregular localized electrical activity in the brain. In transcranial magnetic stimulation, a rapidly varying and very localized magnetic field is placed close to certain sites identified in the brain. The usage of TMS as a diagnostic technique is well established.

Note:

Check out this [Youtube video](#) to see how rock-and-roll instruments like electric guitars use electromagnetic induction to get those strong beats.

Summary

- Hard drives utilize magnetic induction to read/write information.
- Other applications of magnetic induction can be found in graphics tablets, electric and hybrid vehicles, and in transcranial magnetic stimulation.

Key Equations

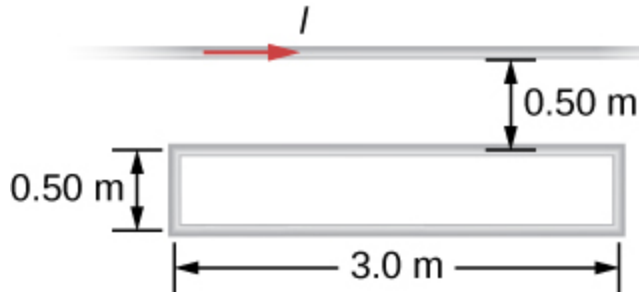
Magnetic flux	$\Phi_m = \int_S \vec{\mathbf{B}} \cdot \hat{\mathbf{n}} dA$
Faraday's law	$\varepsilon = -N \frac{d\Phi_m}{dt}$
Motionally induced emf	$\varepsilon = Blv$
Motional emf around a circuit	$\varepsilon = \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}} = -\frac{d\Phi_m}{dt}$
Emf produced by an electric generator	$\varepsilon = NBA \omega \sin(\omega t)$

Additional Problems

Exercise:

Problem:

Shown in the following figure is a long, straight wire and a single-turn rectangular loop, both of which lie in the plane of the page. The wire is parallel to the long sides of the loop and is 0.50 m away from the closer side. At an instant when the emf induced in the loop is 2.0 V, what is the time rate of change of the current in the wire?



Solution:

$$4.8 \times 10^6 \text{ A/s}$$

Exercise:

Problem:

A metal bar of mass 500 g slides outward at a constant speed of 1.5 cm/s over two parallel rails separated by a distance of 30 cm which are part of a U-shaped conductor. There is a uniform magnetic field of magnitude 2 T pointing out of the page over the entire area. The railings and metal bar have an equivalent resistance of 150Ω . (a) Determine the induced current, both magnitude and direction. (b) Find the direction of the induced current if the magnetic field is pointing into the page. (c) Find the direction of the induced current if the magnetic field is pointed into the page and the bar moves inwards.

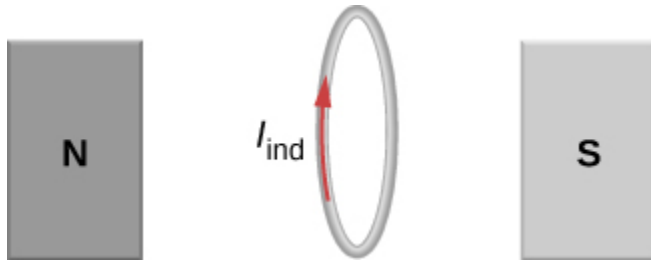
Exercise:

Problem:

A current is induced in a circular loop of radius 1.5 cm between two poles of a horseshoe electromagnet when the current in the electromagnet is varied. The magnetic field in the area of the loop is perpendicular to the area and has a uniform magnitude. If the rate of change of magnetic field is 10 T/s, find the magnitude and direction of the induced current if resistance of the loop is 25Ω .

Solution:

$2.83 \times 10^{-4} \text{ A}$, the direction as follows for increasing magnetic field:



Exercise:

Problem:

A metal bar of length 25 cm is placed perpendicular to a uniform magnetic field of strength 3 T. (a) Determine the induced emf between the ends of the rod when it is not moving. (b) Determine the emf when the rod is moving perpendicular to its length and magnetic field with a speed of 50 cm/s.

Exercise:

Problem:

A coil with 50 turns and area 10 cm^2 is oriented with its plane perpendicular to a 0.75-T magnetic field. If the coil is flipped over (rotated through 180°) in 0.20 s, what is the average emf induced in it?

Solution:

0.375 V

Exercise:

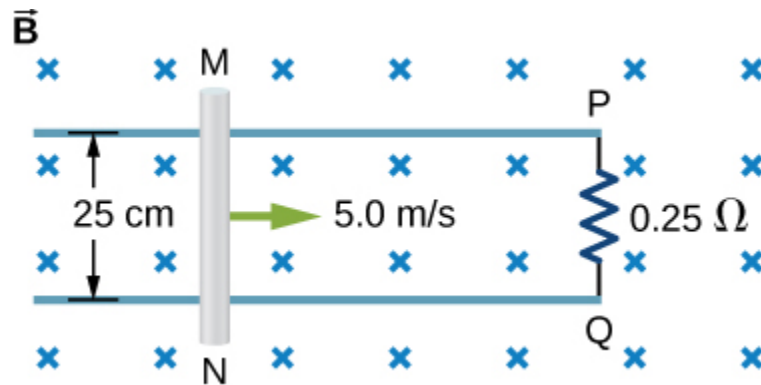
Problem:

A 2-turn planer loop of flexible wire is placed inside a long solenoid of n turns per meter that carries a constant current I_0 . The area A of the loop is changed by pulling on its sides while ensuring that the plane of the loop always remains perpendicular to the axis of the solenoid. If $n = 500$ turns per meter, $I_0 = 20 \text{ A}$, and $A = 20 \text{ cm}^2$, what is the emf induced in the loop when $dA/dt = 100$?

Exercise:

Problem:

The conducting rod shown in the accompanying figure moves along parallel metal rails that are 25-cm apart. The system is in a uniform magnetic field of strength 0.75 T, which is directed into the page. The resistances of the rod and the rails are negligible, but the section PQ has a resistance of $0.25\ \Omega$. (a) What is the emf (including its sense) induced in the rod when it is moving to the right with a speed of 5.0 m/s? (b) What force is required to keep the rod moving at this speed? (c) What is the rate at which work is done by this force? (d) What is the power dissipated in the resistor?

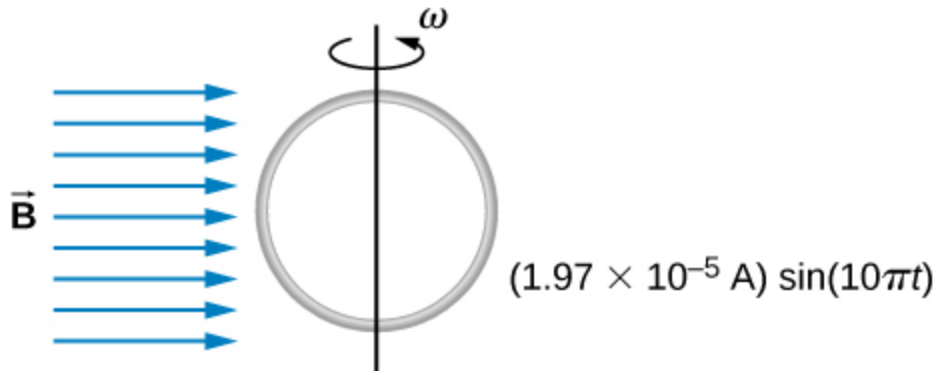


Solution:

a. 0.94 V; b. 0.70 N; c. 3.52 J/s; d. 3.52 W

Exercise:**Problem:**

A circular loop of wire of radius 10 cm is mounted on a vertical shaft and rotated at a frequency of 5 cycles per second in a region of uniform magnetic field of 2 Gauss perpendicular to the axis of rotation. (a) Find an expression for the time-dependent flux through the ring. (b) Determine the time-dependent current through the ring if it has a resistance of $10\ \Omega$.



Exercise:

Problem:

The magnetic field between the poles of a horseshoe electromagnet is uniform and has a cylindrical symmetry about an axis from the middle of the South Pole to the middle of the North Pole. The magnitude of the magnetic field changes as a rate of dB/dt due to the changing current through the electromagnet. Determine the electric field at a distance r from the center.

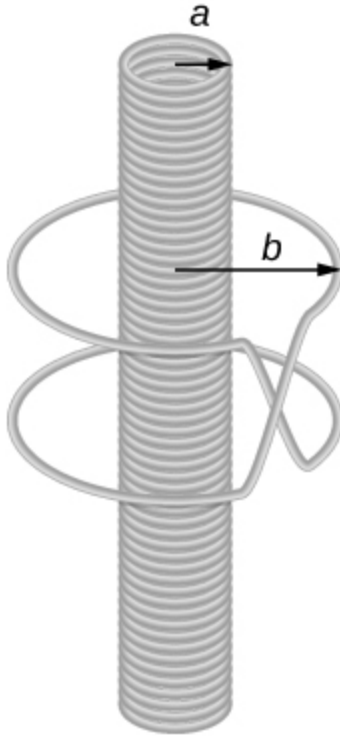
Solution:

$$\left(\frac{dB}{dt} \right) \frac{A}{2\pi r}$$

Exercise:

Problem:

A long solenoid of radius a with n turns per unit length is carrying a time-dependent current $I(t) = I_0 \sin(\omega t)$, where I_0 and ω are constants. The solenoid is surrounded by a wire of resistance R that has two circular loops of radius b with $b > a$ (see the following figure). Find the magnitude and direction of current induced in the outer loops at time $t = 0$.



Exercise:

Problem:

A 120-V, series-wound dc motor draws 0.50 A from its power source when operating at full speed, and it draws 2.0 A when it starts. The resistance of the armature coils is $10\ \Omega$. (a) What is the resistance of the field coils? (b) What is the back emf of the motor when it is running at full speed? (c) The motor operates at a different speed and draws 1.0 A from the source. What is the back emf in this case?

Solution:

a. $R_f + R_a = \frac{120\text{ V}}{2.0\text{ A}} = 60\ \Omega$, so $R_f = 50\ \Omega$;

b. $I = \frac{\varepsilon_s - \varepsilon_i}{R_f + R_a}$, $\Rightarrow \varepsilon_i = 90\text{ V}$;

c. $\varepsilon_i = 60\text{ V}$

Exercise:

Problem:

The armature and field coils of a series-wound motor have a total resistance of $3.0\ \Omega$. When connected to a 120-V source and running at normal speed, the motor draws 4.0 A. (a) How large is the back emf? (b) What current will the motor draw just after it is turned on? Can you suggest a way to avoid this large initial current?

Challenge Problems**Exercise:****Problem:**

A copper wire of length L is fashioned into a circular coil with N turns. When the magnetic field through the coil changes with time, for what value of N is the induced emf a maximum?

Solution:

N is a maximum number of turns allowed.

Exercise:**Problem:**

A 0.50-kg copper sheet drops through a uniform horizontal magnetic field of 1.5 T, and it reaches a terminal velocity of 2.0 m/s. (a) What is the net magnetic force on the sheet after it reaches terminal velocity? (b) Describe the mechanism responsible for this force. (c) How much power is dissipated as Joule heating while the sheet moves at terminal velocity?

Exercise:

Problem:

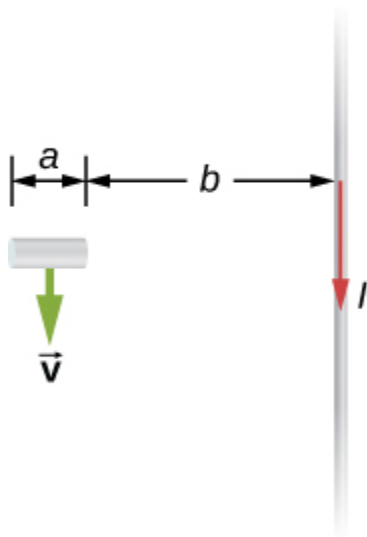
A circular copper disk of radius 7.5 cm rotates at 2400 rpm around the axis through its center and perpendicular to its face. The disk is in a uniform magnetic field \vec{B} of strength 1.2 T that is directed along the axis. What is the potential difference between the rim and the axis of the disk?

Solution:

0.848 V

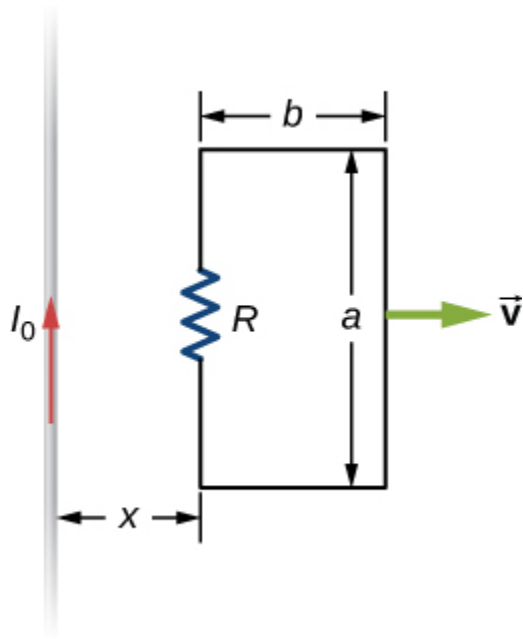
Exercise:**Problem:**

A short rod of length a moves with its velocity \vec{v} parallel to an infinite wire carrying a current I (see below). If the end of the rod nearer the wire is a distance b from the wire, what is the emf induced in the rod?

**Exercise:**

Problem:

A rectangular circuit containing a resistance R is pulled at a constant velocity \vec{v} away from a long, straight wire carrying a current I_0 (see below). Derive an equation that gives the current induced in the circuit as a function of the distance x between the near side of the circuit and the wire.

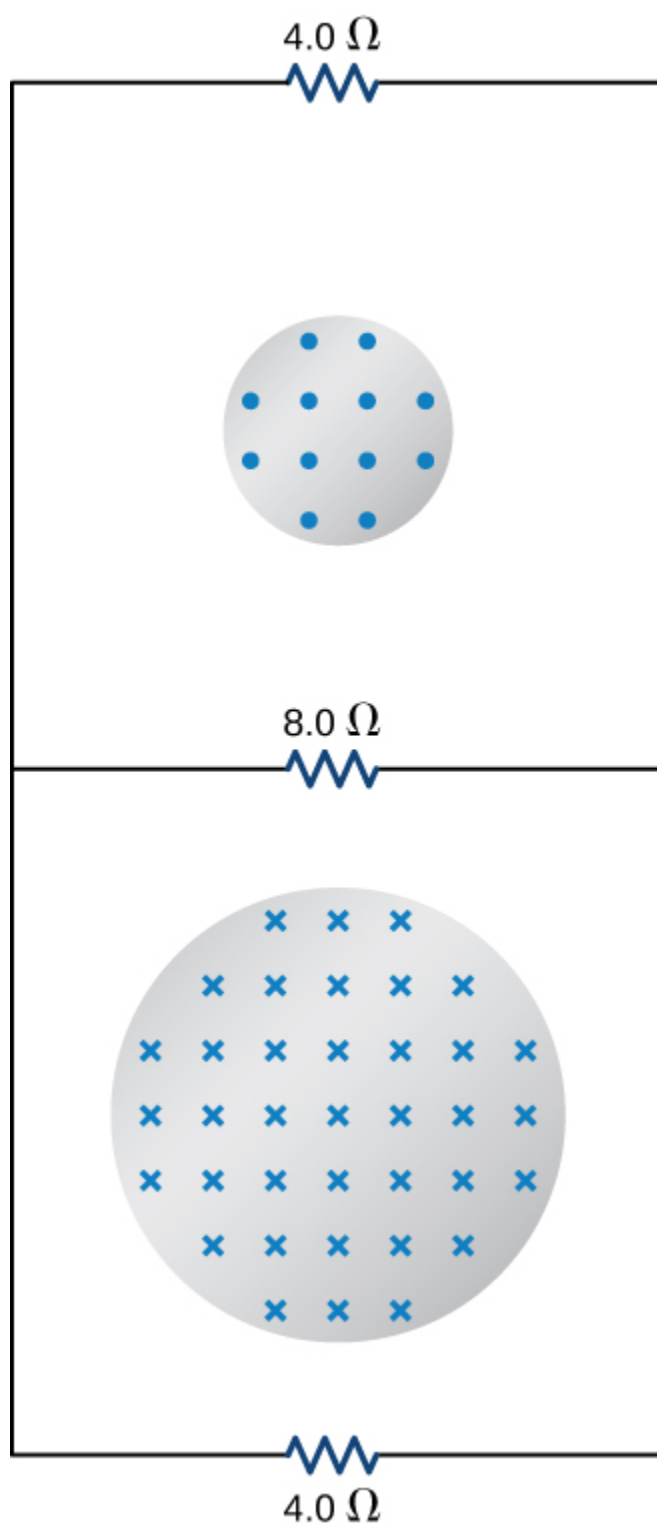
**Solution:**

$$\Phi = \frac{\mu_0 I_0 a}{2\pi} \ln \left(1 + \frac{b}{x} \right), \quad \varepsilon = \frac{\mu_0 I_0 a b v}{2\pi x(x+b)},$$

so $I = \frac{\mu_0 I_0 a b v}{2\pi R x(x+b)}$

Exercise:**Problem:**

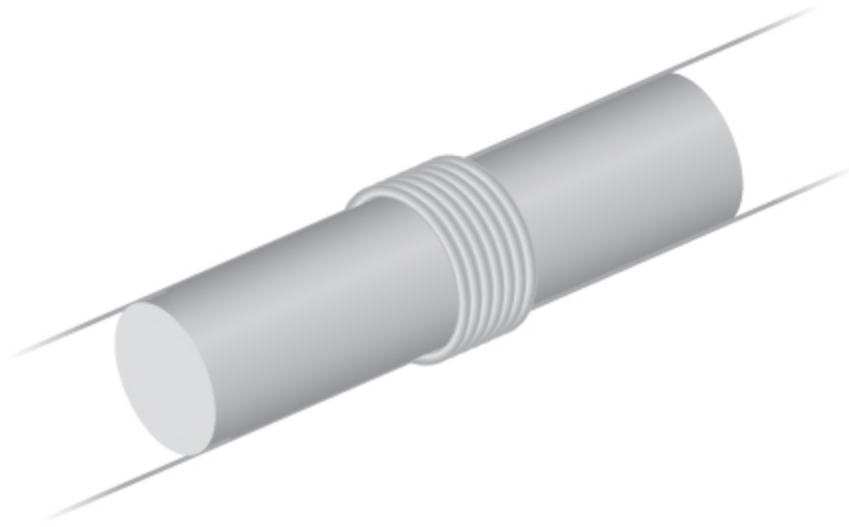
Two infinite solenoids cross the plane of the circuit as shown below. The radii of the solenoids are 0.10 and 0.20 m, respectively, and the current in each solenoid is changing such that $dB/dt = 50.0 \text{ T/s}$. What are the currents in the resistors of the circuit?



Exercise:

Problem:

An eight-turn coil is *tightly wrapped* around the outside of the long solenoid as shown below. The radius of the solenoid is 2.0 cm and it has 10 turns per centimeter. The current through the solenoid increases according to $I = I_0(1 - e^{-\alpha t})$, where $I_0 = 4.0$ A and $\alpha = 2.0 \times 10^{-2} \text{ s}^{-1}$. What is the emf induced in the coil when (a) $t = 0$, (b) $t = 1.0 \times 10^2$ s, and (c) $t \rightarrow \infty$?

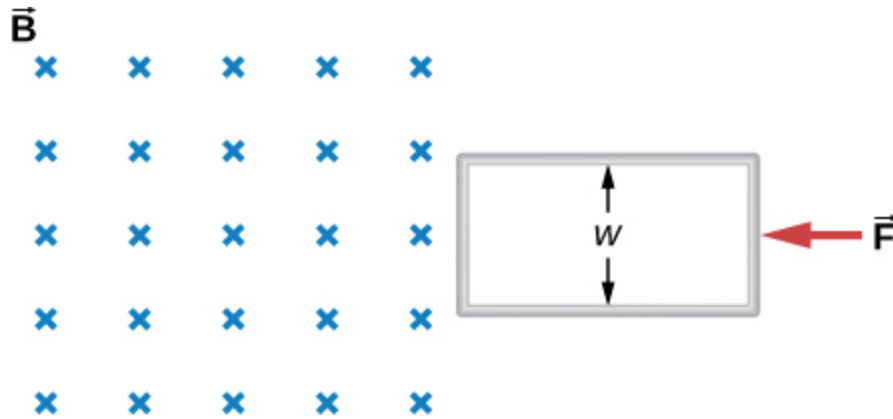


Solution:

a. $1.01 \times 10^{-6} \text{ V}$; b. $1.37 \times 10^{-7} \text{ V}$; c. 0 V

Exercise:**Problem:**

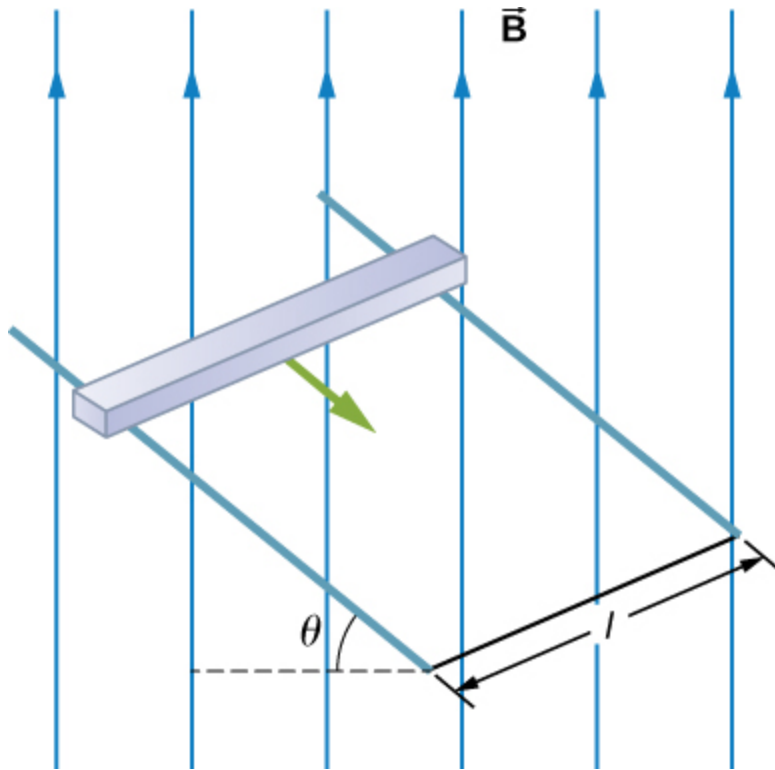
Shown below is a long rectangular loop of width w , length l , mass m , and resistance R . The loop starts from rest at the edge of a uniform magnetic field \vec{B} and is pushed into the field by a constant force \vec{F} . Calculate the speed of the loop as a function of time.



Exercise:

Problem:

A square bar of mass m and resistance R is sliding without friction down very long, parallel conducting rails of negligible resistance (see below). The two rails are a distance l apart and are connected to each other at the bottom of the incline by a zero-resistance wire. The rails are inclined at an angle θ , and there is a uniform vertical magnetic field \vec{B} throughout the region. (a) Show that the bar acquires a terminal velocity given by $v = \frac{mgR \sin \theta}{B^2 l^2 \cos^2 \theta}$. (b) Calculate the work per unit time done by the force of gravity. (c) Compare this with the power dissipated in the Joule heating of the bar. (d) What would happen if \vec{B} were reversed?



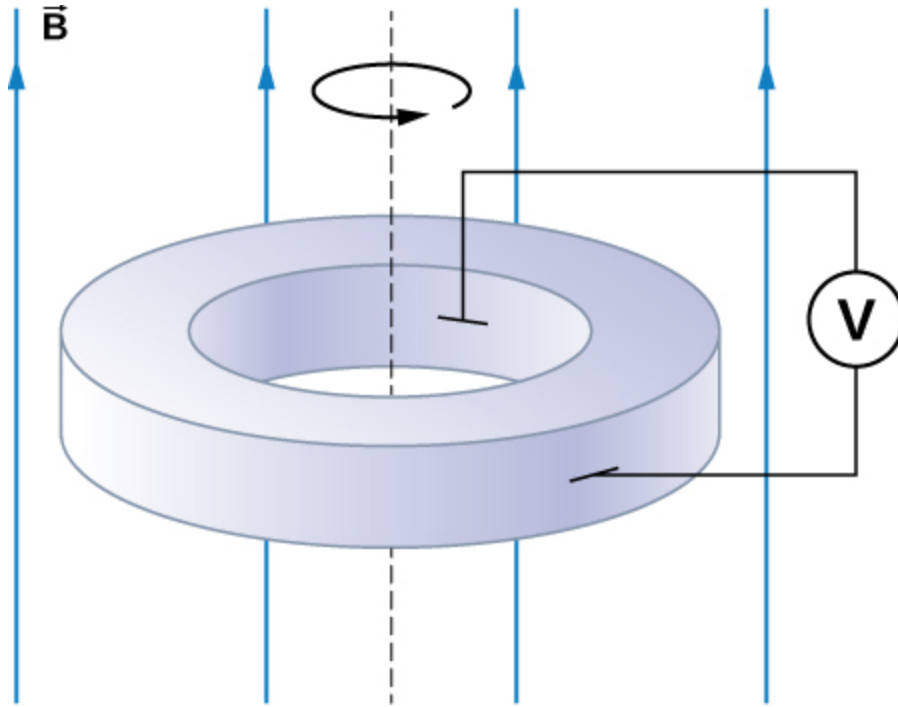
Solution:

a. $v = \frac{mgR \sin \theta}{B^2 l^2 \cos^2 \theta}$; b. $mgv \sin \theta$; c. $mc\Delta T$; d. current would reverse direction but bar would still slide at the same speed

Exercise:

Problem:

The accompanying figure shows a metal disk of inner radius r_1 and outer radius r_2 rotating at an angular velocity $\vec{\omega}$ while in a uniform magnetic field directed parallel to the rotational axis. The brush leads of a voltmeter are connected to the disk's inner and outer surfaces as shown. What is the reading of the voltmeter?



Exercise:

Problem:

A long solenoid with 10 turns per centimeter is placed inside a copper ring such that both objects have the same central axis. The radius of the ring is 10.0 cm, and the radius of the solenoid is 5.0 cm. (a) What is the emf induced in the ring when the current I through the solenoid is 5.0 A and changing at a rate of 100 A/s? (b) What is the emf induced in the ring when $I = 2.0$ A and $dI/dt = 100$ A/s? (c) What is the electric field inside the ring for these two cases? (d) Suppose the ring is moved so that its central axis and the central axis of the solenoid are still parallel but no longer coincide. (You should assume that the solenoid is still inside the ring.) Now what is the emf induced in the ring? (e) Can you calculate the electric field in the ring as you did in part (c)?

Solution:

a.

$$B = \mu_0 n I, \Phi_m = BA = \mu_0 n I A,$$

$$\varepsilon = 9.9 \times 10^{-4} \text{ V};$$

b. $9.9 \times 10^{-4} \text{ V};$

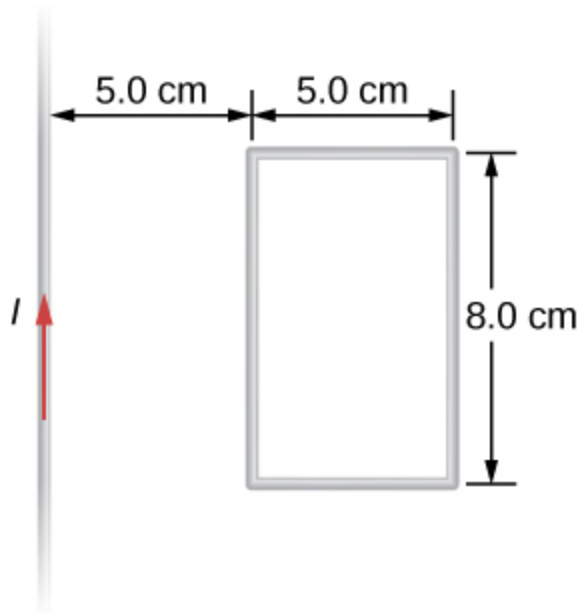
c. $\oint \vec{E} \cdot d\vec{l} = \varepsilon, \Rightarrow E = 1.6 \times 10^{-3} \text{ V/m};$ d. $9.9 \times 10^{-4} \text{ V};$

e. no, because there is no cylindrical symmetry

Exercise:

Problem:

The current in the long, straight wire shown in the accompanying figure is given by $I = I_0 \sin \omega t$, where $I_0 = 15 \text{ A}$ and $\omega = 120\pi \text{ rad/s}$. What is the current induced in the rectangular loop at (a) $t = 0$ and (b) $t = 2.1 \times 10^{-3} \text{ s}$? The resistance of the loop is 2.0Ω .



Exercise:

Problem:

A 500-turn coil with a 0.250-m^2 area is spun in Earth's $5.00 \times 10^{-5}\text{T}$ magnetic field, producing a 12.0-kV maximum emf. (a) At what angular velocity must the coil be spun? (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

a. $1.92 \times 10^6 \text{ rad/s} = 1.83 \times 10^7 \text{ rpm}$; b. This angular velocity is unreasonably high, higher than can be obtained for any mechanical system. c. The assumption that a voltage as great as 12.0 kV could be obtained is unreasonable.

Exercise:**Problem:**

A circular loop of wire of radius 10 cm is mounted on a vertical shaft and rotated at a frequency of 5 cycles per second in a region of uniform magnetic field of $2 \times 10^{-4}\text{T}$ perpendicular to the axis of rotation. (a) Find an expression for the time-dependent flux through the ring (b) Determine the time-dependent current through the ring if it has a resistance of 10Ω .

Exercise:**Problem:**

A long solenoid of radius a with n turns per unit length is carrying a time-dependent current $I(t) = I_0 \sin \omega t$ where I_0 and ω are constants. The solenoid is surrounded by a wire of resistance R that has two circular loops of radius b with $b > a$. Find the magnitude and direction of current induced in the outer loops at time $t = 0$.

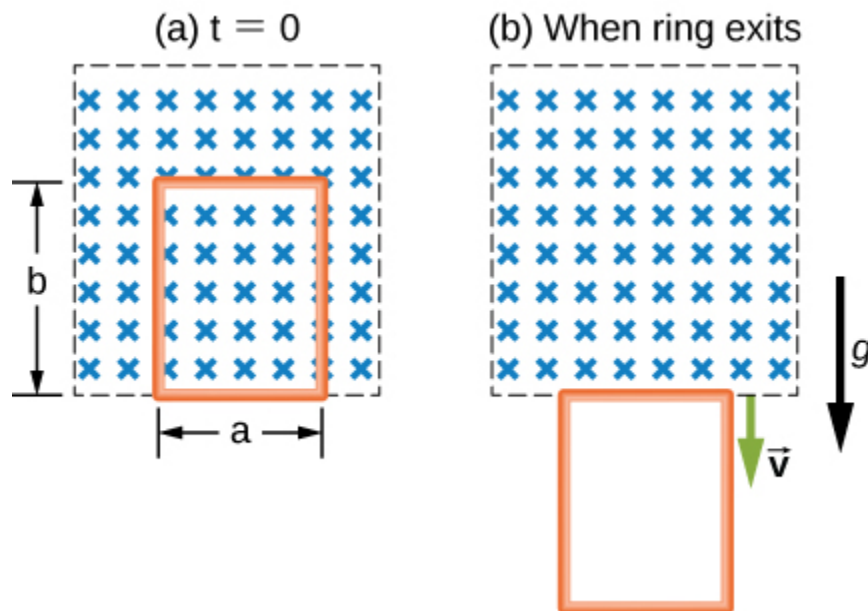
Solution:

$$\frac{2\mu_0\pi a^2 I_0 n \omega}{R}$$

Exercise:

Problem:

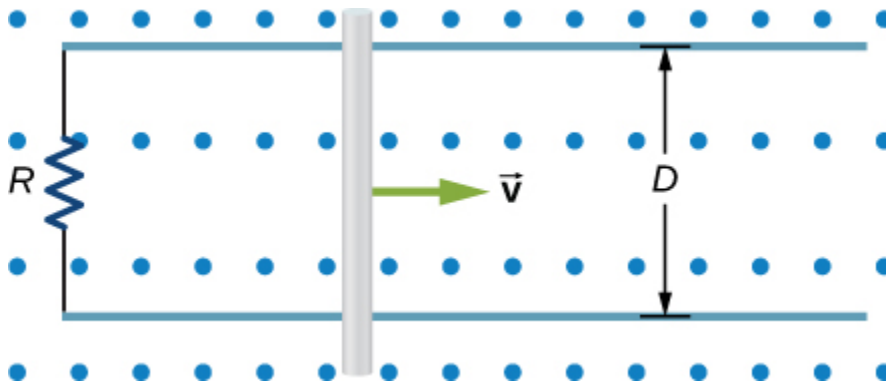
A rectangular copper loop of mass 100 g and resistance $0.2 \, \Omega$ is in a region of uniform magnetic field that is perpendicular to the area enclosed by the ring and horizontal to Earth's surface (see below). The loop is let go from rest when it is at the edge of the nonzero magnetic field region. (a) Find an expression for the speed when the loop just exits the region of uniform magnetic field. (b) If it was let go at $t = 0$, what is the time when it exits the region of magnetic field for the following values: $a = 25 \, \text{cm}$, $b = 50 \, \text{cm}$, $B = 3 \, \text{T}$, $g = 9.8 \, \text{m/s}^2$?



Exercise:

Problem:

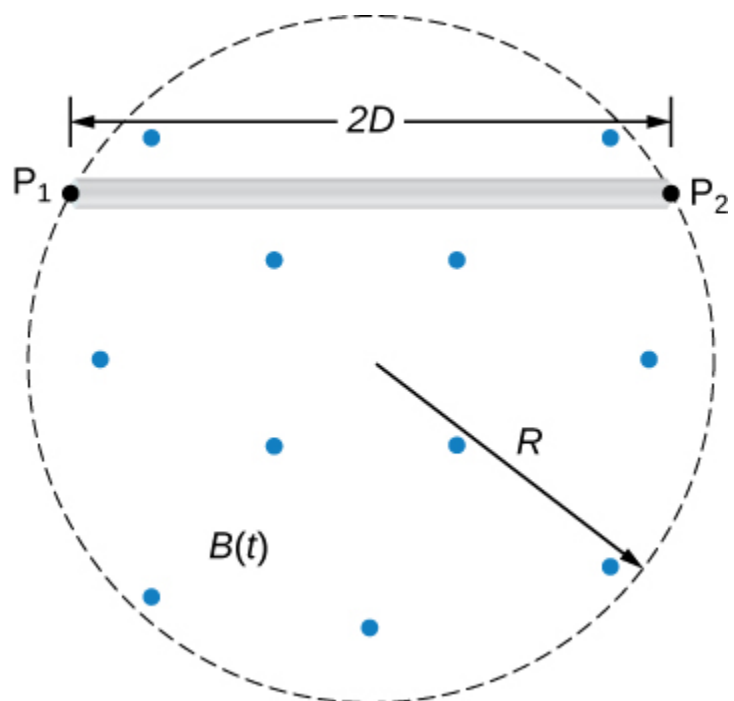
A metal bar of mass m slides without friction over two rails a distance D apart in the region that has a uniform magnetic field of magnitude B_0 and direction perpendicular to the rails (see below). The two rails are connected at one end to a resistor whose resistance is much larger than the resistance of the rails and the bar. The bar is given an initial speed of v_0 . It is found to slow down. How far does the bar go before coming to rest? Assume that the magnetic field of the induced current is negligible compared to B_0 .

**Solution:**

$$\frac{mRv_0}{B^2 D^2}$$

Exercise:**Problem:**

A time-dependent uniform magnetic field of magnitude $B(t)$ is confined in a cylindrical region of radius R . A conducting rod of length $2D$ is placed in the region, as shown below. Show that the emf between the ends of the rod is given by $\frac{dB}{dt} D\sqrt{R^2 - D^2}$. (*Hint: To find the emf between the ends, we need to integrate the electric field from one end to the other. To find the electric field, use Faraday's law as "Ampère's law for E ."*)



Introduction

class="introduction"

A
smartphone
charging
mat contains
a coil that
receives
alternating
current, or
current that
is constantly
increasing
and
decreasing.
The varying
current
induces an
emf in the
smartphone,
which
charges its
battery. Note
that the
black box
containing
the electrical
plug also
contains a
transformer
(discussed in
[Alternating-
Current
Circuits](#))
that
modifies the

current from
the outlet to
suit the
needs of the
smartphone.
(credit:
modification
of work by
“LG”/Flickr
)



In [Electromagnetic Induction](#), we discussed how a time-varying magnetic flux induces an emf in a circuit. In many of our calculations, this flux was due to an applied time-dependent magnetic field. The reverse of this phenomenon also occurs: The current flowing in a circuit produces its own magnetic field.

In [Electric Charges and Fields](#), we saw that induction is the process by which an emf is induced by changing electric flux and separation of a dipole. So far, we have discussed some examples of induction, although some of these applications are more effective than others. The smartphone charging mat in the chapter opener photo also works by induction. Is there a useful physical quantity related to how “effective” a given device is? The answer is yes, and that physical quantity is *inductance*. In this chapter, we look at the applications of inductance in electronic devices and how inductors are used in circuits.

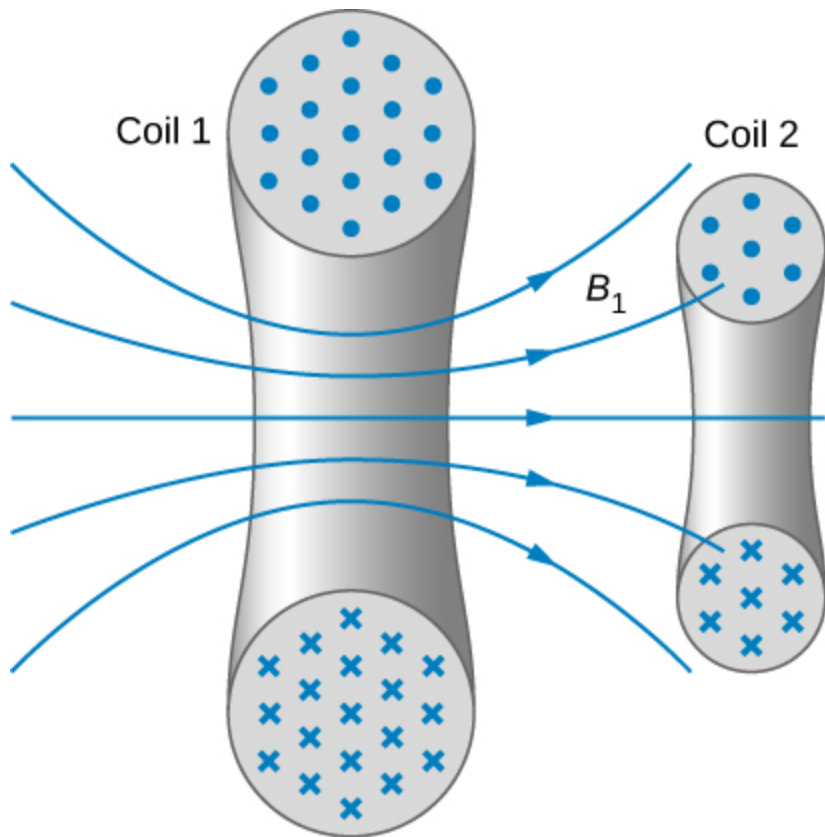
Mutual Inductance

By the end of this section, you will be able to:

- Correlate two nearby circuits that carry time-varying currents with the emf induced in each circuit
- Describe examples in which mutual inductance may or may not be desirable

Inductance is the property of a device that tells us how effectively it induces an emf in another device. In other words, it is a physical quantity that expresses the effectiveness of a given device.

When two circuits carrying time-varying currents are close to one another, the magnetic flux through each circuit varies because of the changing current I in the other circuit. Consequently, an emf is induced in each circuit by the changing current in the other. This type of emf is therefore called a *mutually induced emf*, and the phenomenon that occurs is known as **mutual inductance (M)**. As an example, let's consider two tightly wound coils ([\[link\]](#)). Coils 1 and 2 have N_1 and N_2 turns and carry currents I_1 and I_2 , respectively. The flux through a single turn of coil 2 produced by the magnetic field of the current in coil 1 is Φ_{21} , whereas the flux through a single turn of coil 1 due to the magnetic field of I_2 is Φ_{12} .



Some of the magnetic field lines produced by the current in coil 1 pass through coil 2.

The mutual inductance M_{21} of coil 2 with respect to coil 1 is the ratio of the flux through the N_2 turns of coil 2 produced by the magnetic field of the current in coil 1, divided by that current, that is,

Equation:

$$M_{21} = \frac{N_2 \Phi_{21}}{I_1}.$$

Similarly, the mutual inductance of coil 1 with respect to coil 2 is

Equation:

$$M_{12} = \frac{N_1 \Phi_{12}}{I_2}.$$

Like capacitance, mutual inductance is a geometric quantity. It depends on the shapes and relative positions of the two coils, and it is independent of the currents in the coils. The SI unit for mutual inductance M is called the **henry (H)** in honor of Joseph Henry (1799–1878), an American scientist who discovered induced emf independently of Faraday. Thus, we have $1 \text{ H} = 1 \text{ V} \cdot \text{s}/\text{A}$. From [\[link\]](#) and [\[link\]](#), we can show that $M_{21} = M_{12}$, so we usually drop the subscripts associated with mutual inductance and write

Note:
Equation:

$$M = \frac{N_2 \Phi_{21}}{I_1} = \frac{N_1 \Phi_{12}}{I_2}.$$

The emf developed in either coil is found by combining Faraday's law and the definition of mutual inductance. Since $N_2 \Phi_{21}$ is the total flux through coil 2 due to I_1 , we obtain

Equation:

$$\varepsilon_2 = -\frac{d}{dt}(N_2 \Phi_{21}) = -\frac{d}{dt}(M I_1) = -M \frac{dI_1}{dt}$$

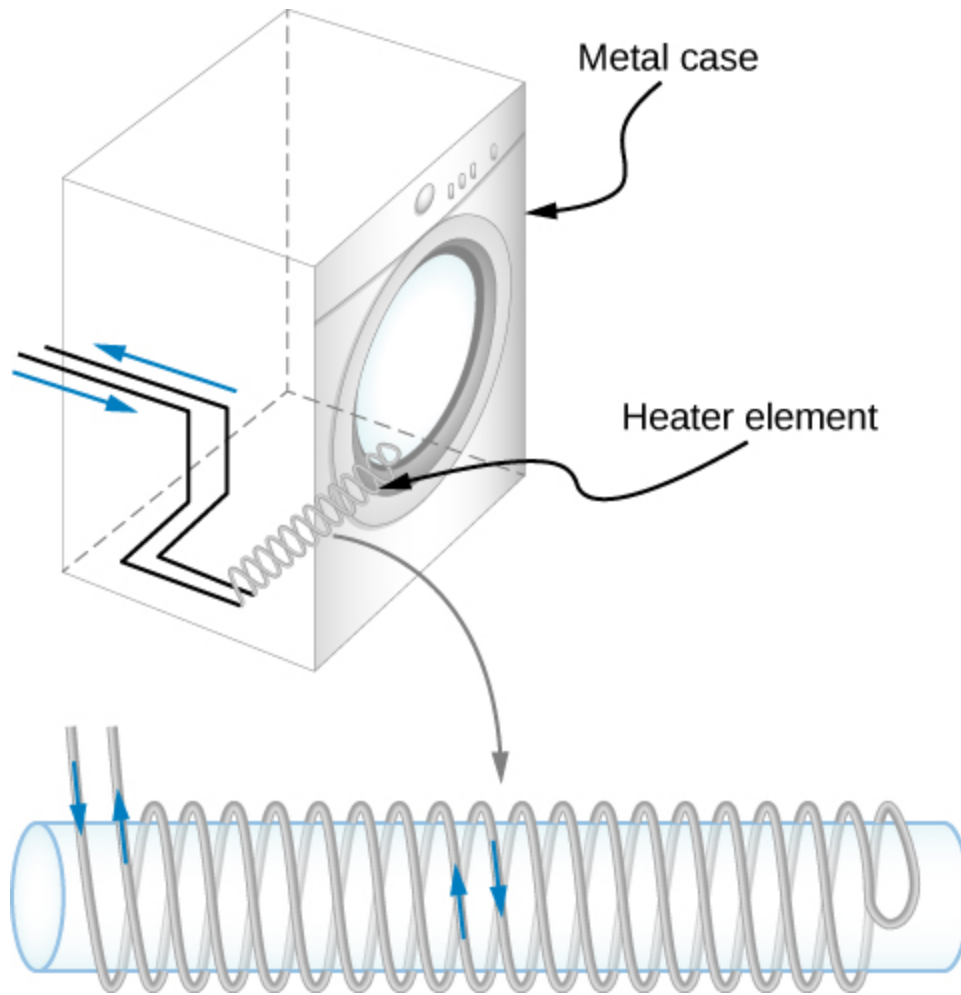
where we have used the fact that M is a time-independent constant because the geometry is time-independent. Similarly, we have

Note:
Equation:

$$\varepsilon_1 = -M \frac{dI_2}{dt}.$$

In [\[link\]](#), we can see the significance of the earlier description of mutual inductance (M) as a geometric quantity. The value of M neatly encapsulates the physical properties of circuit elements and allows us to separate the physical layout of the circuit from the dynamic quantities, such as the emf and the current. [\[link\]](#) defines the mutual inductance in terms of properties in the circuit, whereas the previous definition of mutual inductance in [\[link\]](#) is defined in terms of the magnetic flux experienced, regardless of circuit elements. You should be careful when using [\[link\]](#) and [\[link\]](#) because ε_1 and ε_2 do not necessarily represent the total emfs in the respective coils. Each coil can also have an emf induced in it because of its *self-inductance* (self-inductance will be discussed in more detail in a later section).

A large mutual inductance M may or may not be desirable. We want a transformer to have a large mutual inductance. But an appliance, such as an electric clothes dryer, can induce a dangerous emf on its metal case if the mutual inductance between its coils and the case is large. One way to reduce mutual inductance is to counter-wind coils to cancel the magnetic field produced ([\[link\]](#)).

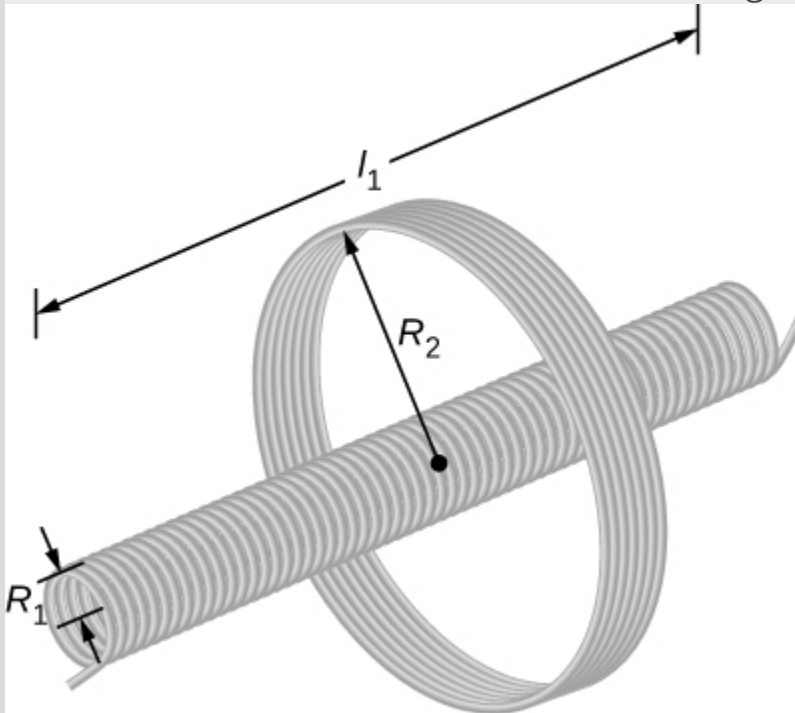


The heating coils of an electric clothes dryer can be counter-wound so that their magnetic fields cancel one another, greatly reducing the mutual inductance with the case of the dryer.

Digital signal processing is another example in which mutual inductance is reduced by counter-winding coils. The rapid on/off emf representing 1s and 0s in a digital circuit creates a complex time-dependent magnetic field. An emf can be generated in neighboring conductors. If that conductor is also carrying a digital signal, the induced emf may be large enough to switch 1s and 0s, with consequences ranging from inconvenient to disastrous.

Example:**Mutual Inductance**

[\[link\]](#) shows a coil of N_2 turns and radius R_2 surrounding a long solenoid of length l_1 , radius R_1 , and N_1 turns. (a) What is the mutual inductance of the two coils? (b) If $N_1 = 500$ turns, $N_2 = 10$ turns, $R_1 = 3.10$ cm, $l_1 = 75.0$ cm, and the current in the solenoid is changing at a rate of 200 A/s, what is the emf induced in the surrounding coil?



A solenoid surrounded by a coil.

Strategy

There is no magnetic field outside the solenoid, and the field inside has magnitude $B_1 = \mu_0(N_1/l_1)I_1$ and is directed parallel to the solenoid's axis. We can use this magnetic field to find the magnetic flux through the surrounding coil and then use this flux to calculate the mutual inductance for part (a), using [\[link\]](#). We solve part (b) by calculating the mutual inductance from the given quantities and using [\[link\]](#) to calculate the induced emf.

Solution

a. The magnetic flux Φ_{21} through the surrounding coil is

Equation:

$$\Phi_{21} = B_1 \pi R_1^2 = \frac{\mu_0 N_1 I_1}{l_1} \pi R_1^2.$$

Now from [\[link\]](#), the mutual inductance is

Equation:

$$M = \frac{N_2 \Phi_{21}}{I_1} = \left(\frac{N_2}{I_1} \right) \left(\frac{\mu_0 N_1 I_1}{l_1} \right) \pi R_1^2 = \frac{\mu_0 N_1 N_2 \pi R_1^2}{l_1}.$$

b. Using the previous expression and the given values, the mutual inductance is

Equation:

$$\begin{aligned} M &= \frac{(4\pi \times 10^{-7} \text{ T}\cdot\text{m/A})(500)(10)\pi(0.0310 \text{ m})^2}{0.750 \text{ m}} \\ &= 2.53 \times 10^{-5} \text{ H}. \end{aligned}$$

Thus, from [\[link\]](#), the emf induced in the surrounding coil is

Equation:

$$\begin{aligned} \varepsilon_2 &= -M \frac{dI_1}{dt} = -(2.53 \times 10^{-5} \text{ H})(200 \text{ A/s}) \\ &= -5.06 \times 10^{-3} \text{ V}. \end{aligned}$$

Significance

Notice that M in part (a) is independent of the radius R_2 of the surrounding coil because the solenoid's magnetic field is confined to its interior. In principle, we can also calculate M by finding the magnetic flux through the solenoid produced by the current in the surrounding coil. This approach is much more difficult because Φ_{12} is so complicated. However, since $M_{12} = M_{21}$, we do know the result of this calculation.

Note:

Exercise:**Problem:**

Check Your Understanding A current

$I(t) = (5.0 \text{ A}) \sin((120\pi \text{ rad/s})t)$ flows through the solenoid of part (b) of [\[link\]](#). What is the maximum emf induced in the surrounding coil?

Solution:

$$4.77 \times 10^{-2} \text{ V}$$

Summary

- Inductance is the property of a device that expresses how effectively it induces an emf in another device.
- Mutual inductance is the effect of two devices inducing emfs in each other.
- A change in current dI_1/dt in one circuit induces an emf (ε_2) in the second:

Equation:

$$\varepsilon_2 = -M \frac{dI_1}{dt},$$

where M is defined to be the mutual inductance between the two circuits and the minus sign is due to Lenz's law.

- Symmetrically, a change in current dI_2/dt through the second circuit induces an emf (ε_1) in the first:

Equation:

$$\varepsilon_1 = -M \frac{dI_2}{dt},$$

where M is the same mutual inductance as in the reverse process.

Conceptual Questions

Exercise:

Problem:

Show that $N\Phi_m/I$ and $\varepsilon/(dI/dt)$, which are both expressions for self-inductance, have the same units.

Solution:

$$\frac{\text{Wb}}{\text{A}} = \frac{\text{T}\cdot\text{m}^2}{\text{A}} = \frac{\text{V}\cdot\text{s}}{\text{A}} = \frac{\text{V}}{\text{A/s}}$$

Exercise:

Problem:

A 10-H inductor carries a current of 20 A. Describe how a 50-V emf can be induced across it.

Exercise:

Problem:

The ignition circuit of an automobile is powered by a 12-V battery. How are we able to generate large voltages with this power source?

Solution:

The induced current from the 12-V battery goes through an inductor, generating a large voltage.

Exercise:

Problem:

When the current through a large inductor is interrupted with a switch, an arc appears across the open terminals of the switch. Explain.

Problems

Exercise:**Problem:**

When the current in one coil changes at a rate of 5.6 A/s, an emf of 6.3×10^{-3} V is induced in a second, nearby coil. What is the mutual inductance of the two coils?

Exercise:**Problem:**

An emf of 9.7×10^{-3} V is induced in a coil while the current in a nearby coil is decreasing at a rate of 2.7 A/s. What is the mutual inductance of the two coils?

Solution:

$$M = 3.6 \times 10^{-3} \text{ H}$$

Exercise:**Problem:**

Two coils close to each other have a mutual inductance of 32 mH. If the current in one coil decays according to $I = I_0 e^{-\alpha t}$, where $I_0 = 5.0$ A and $\alpha = 2.0 \times 10^3 \text{ s}^{-1}$, what is the emf induced in the second coil immediately after the current starts to decay? At $t = 1.0 \times 10^{-3}$ s?

Exercise:**Problem:**

A coil of 40 turns is wrapped around a long solenoid of cross-sectional area $7.5 \times 10^{-3} \text{ m}^2$. The solenoid is 0.50 m long and has 500 turns. (a) What is the mutual inductance of this system? (b) The outer coil is replaced by a coil of 40 turns whose radius is three times that of the solenoid. What is the mutual inductance of this configuration?

Solution:

a. $3.8 \times 10^{-4} \text{ H}$; b. $3.8 \times 10^{-4} \text{ H}$

Exercise:

Problem:

A 600-turn solenoid is 0.55 m long and 4.2 cm in diameter. Inside the solenoid, a small ($1.1 \text{ cm} \times 1.4 \text{ cm}$), single-turn rectangular coil is fixed in place with its face perpendicular to the long axis of the solenoid. What is the mutual inductance of this system?

Exercise:

Problem:

A toroidal coil has a mean radius of 16 cm and a cross-sectional area of 0.25 cm^2 ; it is wound uniformly with 1000 turns. A second toroidal coil of 750 turns is wound uniformly over the first coil. Ignoring the variation of the magnetic field within a toroid, determine the mutual inductance of the two coils.

Solution:

$$M_{21} = 2.3 \times 10^{-5} \text{ H}$$

Exercise:

Problem:

A solenoid of N_1 turns has length l_1 and radius R_1 , and a second smaller solenoid of N_2 turns has length l_2 and radius R_2 . The smaller solenoid is placed completely inside the larger solenoid so that their long axes coincide. What is the mutual inductance of the two solenoids?

Glossary

henry (H)

unit of inductance, $1 \text{ H} = 1 \Omega \cdot \text{s}$; it is also expressed as a volt second per ampere

inductance

property of a device that tells how effectively it induces an emf in another device

mutual inductance

geometric quantity that expresses how effective two devices are at inducing emfs in one another

Self-Inductance and Inductors

By the end of this section, you will be able to:

- Correlate the rate of change of current to the induced emf created by that current in the same circuit
- Derive the self-inductance for a cylindrical solenoid
- Derive the self-inductance for a rectangular toroid

Mutual inductance arises when a current in one circuit produces a changing magnetic field that induces an emf in another circuit. But can the magnetic field affect the current in the original circuit that produced the field? The answer is yes, and this is the phenomenon called *self-inductance*.

Inductors

[\[link\]](#) shows some of the magnetic field lines due to the current in a circular loop of wire. If the current is constant, the magnetic flux through the loop is also constant. However, if the current I were to vary with time—say, immediately after switch S is closed—then the magnetic flux Φ_m would correspondingly change. Then Faraday’s law tells us that an emf ε would be induced in the circuit, where

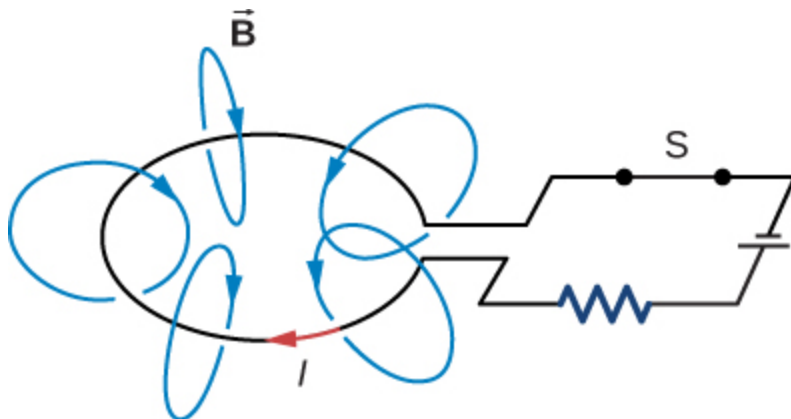
Equation:

$$\varepsilon = -\frac{d\Phi_m}{dt}.$$

Since the magnetic field due to a current-carrying wire is directly proportional to the current, the flux due to this field is also proportional to the current; that is,

Equation:

$$\Phi_m \propto I.$$



A magnetic field is produced by the current I in the loop. If I were to vary with time, the magnetic flux through the loop would also vary and an emf would be induced in the loop.

This can also be written as

Equation:

$$\Phi_m = LI$$

where the constant of proportionality L is known as the **self-inductance** of the wire loop. If the loop has N turns, this equation becomes

Note:

Equation:

$$N\Phi_m = LI.$$

By convention, the positive sense of the normal to the loop is related to the current by the right-hand rule, so in [\[link\]](#), the normal points downward.

With this convention, Φ_m is positive in [\[link\]](#), so L *always has a positive value*.

For a loop with N turns, $\varepsilon = -Nd\Phi_m/dt$, so the induced emf may be written in terms of the self-inductance as

Note:

Equation:

$$\varepsilon = -L \frac{dI}{dt}.$$

When using this equation to determine L , it is easiest to ignore the signs of ε and dI/dt , and calculate L as

Equation:

$$L = \frac{|\varepsilon|}{|dI/dt|}.$$

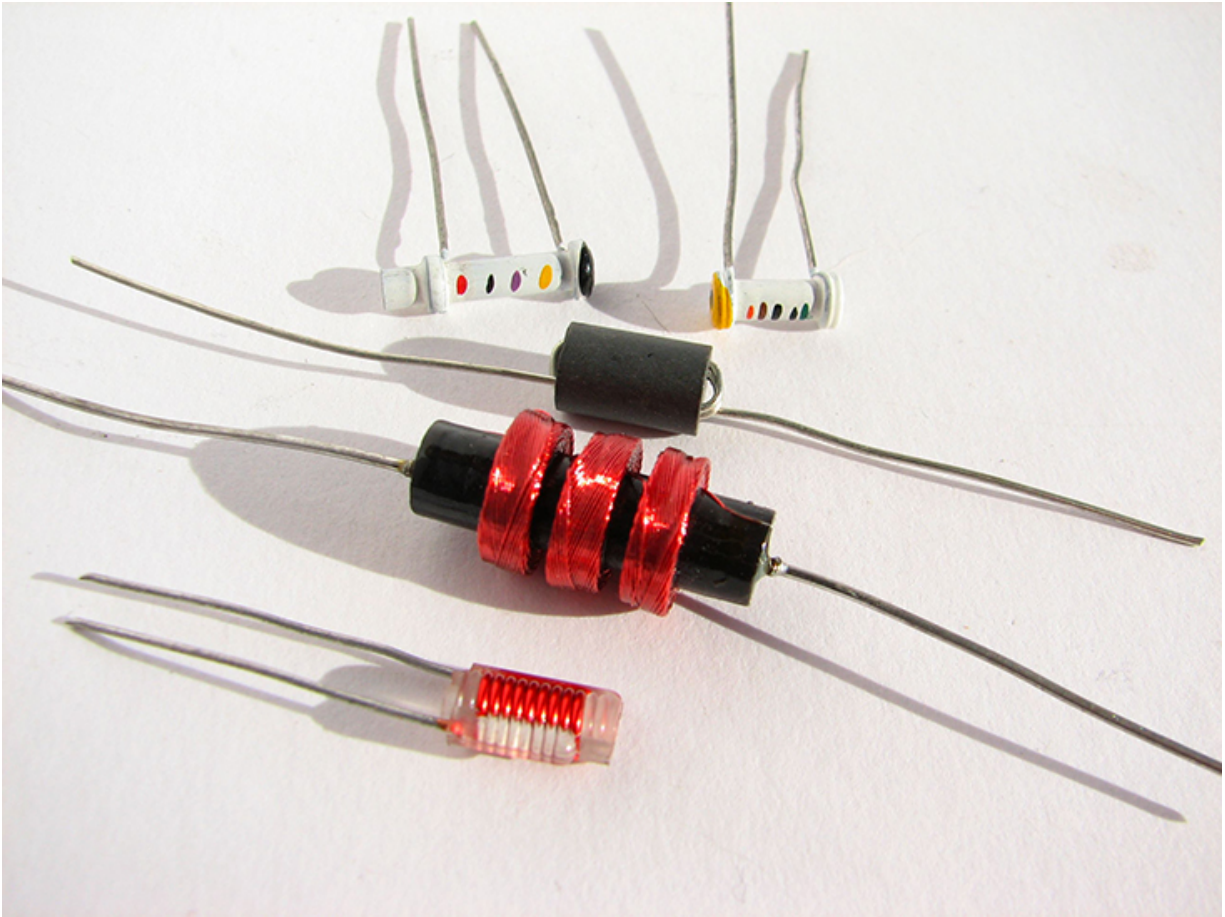
Since self-inductance is associated with the magnetic field produced by a current, any configuration of conductors possesses self-inductance. For example, besides the wire loop, a long, straight wire has self-inductance, as does a coaxial cable. A coaxial cable is most commonly used by the cable television industry and may also be found connecting to your cable modem. Coaxial cables are used due to their ability to transmit electrical signals with minimal distortions. Coaxial cables have two long cylindrical conductors that possess current and a self-inductance that may have undesirable effects.

A circuit element used to provide self-inductance is known as an **inductor**. It is represented by the symbol shown in [\[link\]](#), which resembles a coil of

wire, the basic form of the inductor. [\[link\]](#) shows several types of inductors commonly used in circuits.

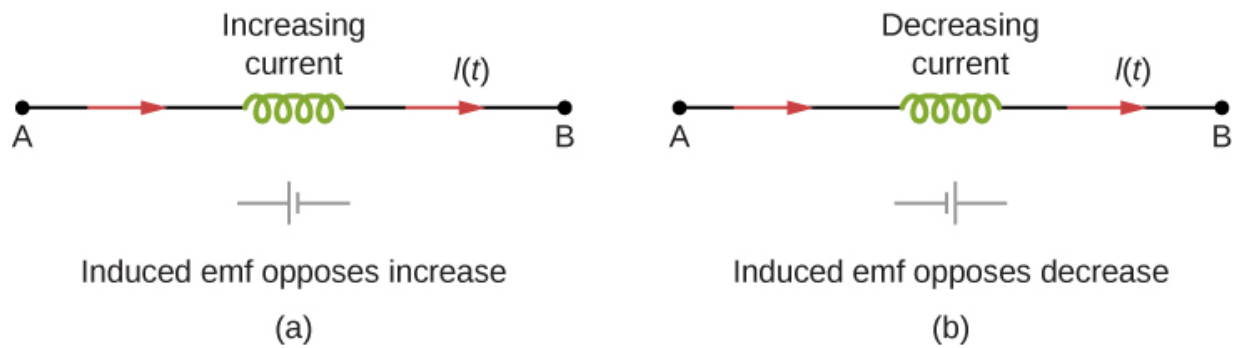


Symbol
used to
represent
an
inductor
in a
circuit.



A variety of inductors. Whether they are encapsulated like the top three shown or wound around in a coil like the bottom-most one, each is simply a relatively long coil of wire. (credit: Windell Oskay)

In accordance with Lenz's law, the negative sign in [\[link\]](#) indicates that the induced emf across an inductor always has a polarity that *opposes* the change in the current. For example, if the current flowing from *A* to *B* in [\[link\]](#)(a) were increasing, the induced emf (represented by the imaginary battery) would have the polarity shown in order to oppose the increase. If the current from *A* to *B* were decreasing, then the induced emf would have the opposite polarity, again to oppose the change in current ([\[link\]](#)(b)). Finally, if the current through the inductor were constant, no emf would be induced in the coil.



The induced emf across an inductor always acts to oppose the change in the current. This can be visualized as an imaginary battery causing current to flow to oppose the change in (a) and reinforce the change in (b).

One common application of inductance is to allow traffic signals to sense when vehicles are waiting at a street intersection. An electrical circuit with an inductor is placed in the road underneath the location where a waiting car will stop. The body of the car increases the inductance and the circuit changes, sending a signal to the traffic lights to change colors. Similarly, metal detectors used for airport security employ the same technique. A coil or inductor in the metal detector frame acts as both a transmitter and a receiver. The pulsed signal from the transmitter coil induces a signal in the receiver. The self-inductance of the circuit is affected by any metal object in the path ([link](#)). Metal detectors can be adjusted for sensitivity and can also sense the presence of metal on a person.



The familiar security gate at an airport not only detects metals, but can also indicate their approximate height above the floor. (credit: “Alexbuids”/Wikimedia Commons)

Large induced voltages are found in camera flashes. Camera flashes use a battery, two inductors that function as a transformer, and a switching system or *oscillator* to induce large voltages. Recall from [Oscillations](#) on oscillations that “oscillation” is defined as the fluctuation of a quantity, or repeated regular fluctuations of a quantity, between two extreme values around an average value. Also recall (from [Electromagnetic Induction](#) on electromagnetic induction) that we need a changing magnetic field, brought about by a changing current, to induce a voltage in another coil. The oscillator system does this many times as the battery voltage is boosted to over 1000 volts. (You may hear the high-pitched whine from the

transformer as the capacitor is being charged.) A capacitor stores the high voltage for later use in powering the flash.

Example:**Self-Inductance of a Coil**

An induced emf of 20 mV is measured across a coil of 50 closely wound turns while the current through it increases uniformly from 0.0 to 5.0 A in 0.10 s. (a) What is the self-inductance of the coil? (b) With the current at 5.0 A, what is the flux through each turn of the coil?

Strategy

Both parts of this problem give all the information needed to solve for the self-inductance in part (a) or the flux through each turn of the coil in part (b). The equations needed are [\[link\]](#) for part (a) and [\[link\]](#) for part (b).

Solution

- a. Ignoring the negative sign and using magnitudes, we have, from [\[link\]](#),

Equation:

$$L = \frac{\varepsilon}{dI/dt} = \frac{20 \text{ mV}}{5.0 \text{ A}/0.10 \text{ s}} = 4.0 \times 10^{-4} \text{ H}.$$

- b. From [\[link\]](#), the flux is given in terms of the current by $\Phi_m = LI/N$,
so

Equation:

$$\Phi_m = \frac{(4.0 \times 10^{-4} \text{ H})(5.0 \text{ A})}{50 \text{ turns}} = 4.0 \times 10^{-5} \text{ Wb}.$$

Significance

The self-inductance and flux calculated in parts (a) and (b) are typical values for coils found in contemporary devices. If the current is not changing over time, the flux is not changing in time, so no emf is induced.

Note:

Exercise:

Problem:

Check Your Understanding Current flows through the inductor in [\[link\]](#) from B to A instead of from A to B as shown. Is the current increasing or decreasing in order to produce the emf given in diagram (a)? In diagram (b)?

Solution:

a. decreasing; b. increasing; Since the current flows in the opposite direction of the diagram, in order to get a positive emf on the left-hand side of diagram (a), we need to decrease the current to the left, which creates a reinforced emf where the positive end is on the left-hand side. To get a positive emf on the right-hand side of diagram (b), we need to increase the current to the left, which creates a reinforced emf where the positive end is on the right-hand side.

Note:

Exercise:

Problem:

Check Your Understanding A changing current induces an emf of 10 V across a 0.25-H inductor. What is the rate at which the current is changing?

Solution:

40 A/s

A good approach for calculating the self-inductance of an inductor consists of the following steps:

Note:

Self-Inductance

1. Assume a current I is flowing through the inductor.
2. Determine the magnetic field \vec{B} produced by the current. If there is appropriate symmetry, you may be able to do this with Ampère's law.
3. Obtain the magnetic flux, Φ_m .
4. With the flux known, the self-inductance can be found from [\[link\]](#),
 $L = N\Phi_m/I$.

To demonstrate this procedure, we now calculate the self-inductances of two inductors.

Cylindrical Solenoid

Consider a long, cylindrical solenoid with length l , cross-sectional area A , and N turns of wire. We assume that the length of the solenoid is so much larger than its diameter that we can take the magnetic field to be $B = \mu_0 nI$ throughout the interior of the solenoid, that is, we ignore end effects in the solenoid. With a current I flowing through the coils, the magnetic field produced within the solenoid is

Equation:

$$B = \mu_0 \left(\frac{N}{l} \right) I,$$

so the magnetic flux through one turn is

Equation:

$$\Phi_m = BA = \frac{\mu_0 N A}{l} I.$$

Using [\[link\]](#), we find for the self-inductance of the solenoid,

Note:

Equation:

$$L_{\text{solenoid}} = \frac{N\Phi_m}{I} = \frac{\mu_0 N^2 A}{l}.$$

If $n = N/l$ is the number of turns per unit length of the solenoid, we may write [\[link\]](#) as

Equation:

$$L = \mu_0 \left(\frac{N}{l} \right)^2 Al = \mu_0 n^2 Al = \mu_0 n^2 (V),$$

where $V = Al$ is the volume of the solenoid. Notice that *the self-inductance of a long solenoid depends only on its physical properties* (such as the number of turns of wire per unit length and the volume), and not on the magnetic field or the current. This is true for inductors in general.

Rectangular Toroid

A toroid with a rectangular cross-section is shown in [\[link\]](#). The inner and outer radii of the toroid are R_1 and R_2 , and h is the height of the toroid. Applying Ampère's law in the same manner as we did in [\[link\]](#) for a toroid with a circular cross-section, we find the magnetic field inside a rectangular toroid is also given by

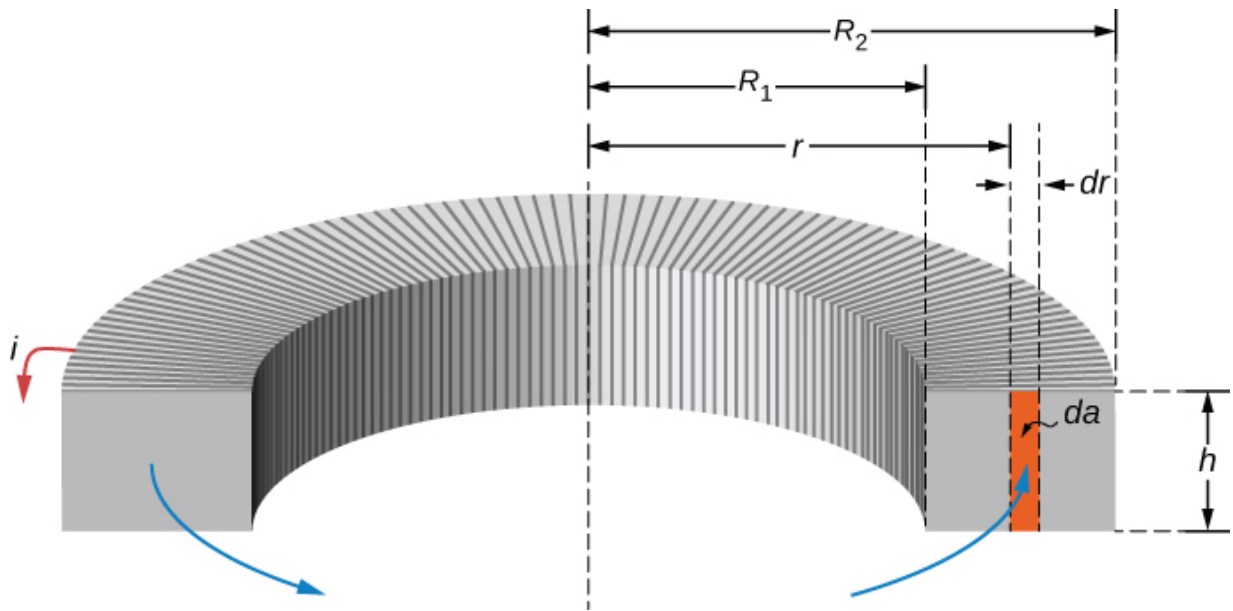
Equation:

$$B = \frac{\mu_0 N I}{2\pi r},$$

where r is the distance from the central axis of the toroid. Because the field changes within the toroid, we must calculate the flux by integrating over the toroid's cross-section. Using the infinitesimal cross-sectional area element $da = h dr$ shown in [\[link\]](#), we obtain

Equation:

$$\Phi_m = \int B da = \int_{R_1}^{R_2} \left(\frac{\mu_0 N I}{2\pi r} \right) (h dr) = \frac{\mu_0 N h I}{2\pi} \ln \frac{R_2}{R_1}.$$



Calculating the self-inductance of a rectangular toroid.

Now from [\[link\]](#), we obtain for the self-inductance of a rectangular toroid

Note:

Equation:

$$L = \frac{N\Phi_m}{I} = \frac{\mu_0 N^2 h}{2\pi} \ln \frac{R_2}{R_1}.$$

As expected, the self-inductance is a constant determined by only the physical properties of the toroid.

Note:

Exercise:

Problem:

Check Your Understanding (a) Calculate the self-inductance of a solenoid that is tightly wound with wire of diameter 0.10 cm, has a cross-sectional area of 0.90 cm^2 , and is 40 cm long. (b) If the current through the solenoid decreases uniformly from 10 to 0 A in 0.10 s, what is the emf induced between the ends of the solenoid?

Solution:

a. $4.5 \times 10^{-5} \text{ H}$; b. $4.5 \times 10^{-3} \text{ V}$

Note:

Exercise:

Problem:

Check Your Understanding (a) What is the magnetic flux through one turn of a solenoid of self-inductance $8.0 \times 10^{-5} \text{ H}$ when a current of 3.0 A flows through it? Assume that the solenoid has 1000 turns and is wound from wire of diameter 1.0 mm . (b) What is the cross-sectional area of the solenoid?

Solution:

a. $2.4 \times 10^{-7} \text{ Wb}$; b. $6.4 \times 10^{-5} \text{ m}^2$

Summary

- Current changes in a device induce an emf in the device itself, called self-inductance,

Equation:

$$\varepsilon = -L \frac{dI}{dt},$$

where L is the self-inductance of the inductor and dI/dt is the rate of change of current through it. The minus sign indicates that emf opposes the change in current, as required by Lenz's law. The unit of self-inductance and inductance is the henry (H), where $1 \text{ H} = 1 \Omega \cdot \text{s}$.

- The self-inductance of a solenoid is

Equation:

$$L = \frac{\mu_0 N^2 A}{l},$$

where N is its number of turns in the solenoid, A is its cross-sectional area, l is its length, and $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ is the permeability of free space.

- The self-inductance of a toroid is

Equation:

$$L = \frac{\mu_0 N^2 h}{2\pi} \ln \frac{R_2}{R_1},$$

where N is its number of turns in the toroid, R_1 and R_2 are the inner and outer radii of the toroid, h is the height of the toroid, and $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ is the permeability of free space.

Conceptual Questions**Exercise:****Problem:**

Does self-inductance depend on the value of the magnetic flux? Does it depend on the current through the wire? Correlate your answers with the equation $N\Phi_m = LI$.

Solution:

Self-inductance is proportional to the magnetic flux and inversely proportional to the current. However, since the magnetic flux depends on the current I , these effects cancel out. This means that the self-inductance does not depend on the current. If the emf is induced across an element, it does depend on how the current changes with time.

Exercise:**Problem:**

Would the self-inductance of a 1.0 m long, tightly wound solenoid differ from the self-inductance per meter of an infinite, but otherwise identical, solenoid?

Exercise:

Problem:

Discuss how you might determine the self-inductance per unit length of a long, straight wire.

Solution:

Consider the ends of a wire a part of an RL circuit and determine the self-inductance from this circuit.

Exercise:**Problem:**

The self-inductance of a coil is zero if there is no current passing through the windings. True or false?

Exercise:**Problem:**

How does the self-inductance per unit length near the center of a solenoid (away from the ends) compare with its value near the end of the solenoid?

Solution:

The magnetic field will flare out at the end of the solenoid so there is less flux through the last turn than through the middle of the solenoid.

Problems**Exercise:****Problem:**

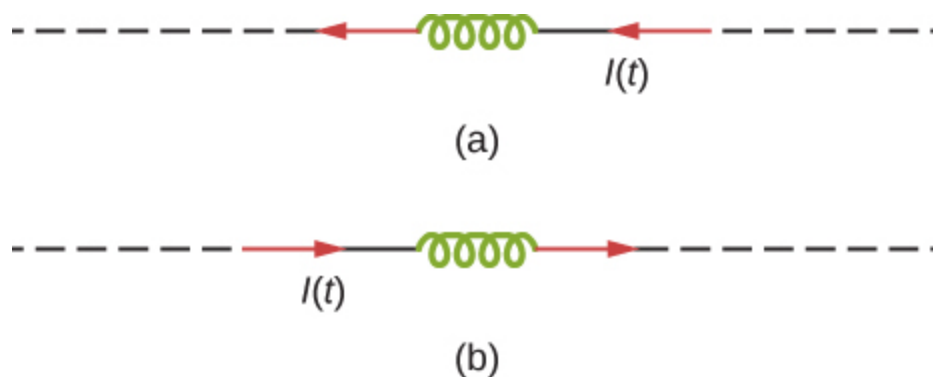
An emf of 0.40 V is induced across a coil when the current through it changes uniformly from 0.10 A to 0.60 A in 0.30 s. What is the self-inductance of the coil?

Solution:

0.24 H

Exercise:**Problem:**

The current shown in part (a) below is increasing, whereas that shown in part (b) is decreasing. In each case, determine which end of the inductor is at the higher potential.

**Exercise:****Problem:**

What is the rate at which the current through a 0.30-H coil is changing if an emf of 0.12 V is induced across the coil?

Solution:

0.4 A/s

Exercise:**Problem:**

When a camera uses a flash, a fully charged capacitor discharges through an inductor. In what time must the 0.100-A current through a 2.00-mH inductor be switched on or off to induce a 500-V emf?

Exercise:

Problem:

A coil with a self-inductance of 2.0 H carries a current that varies with time according to $I(t) = (2.0 \text{ A})\sin 120\pi t$. Find an expression for the emf induced in the coil.

Solution:

$$\varepsilon = 480\pi \sin(120\pi t - \pi/2) \text{ V}$$

Exercise:**Problem:**

A solenoid 50 cm long is wound with 500 turns of wire. The cross-sectional area of the coil is 2.0 cm^2 . What is the self-inductance of the solenoid?

Exercise:**Problem:**

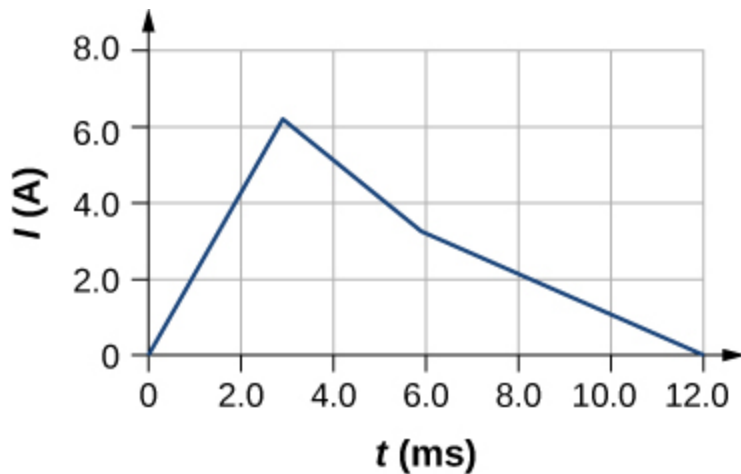
A coil with a self-inductance of 3.0 H carries a current that decreases at a uniform rate $dI/dt = -0.050 \text{ A/s}$. What is the emf induced in the coil? Describe the polarity of the induced emf.

Solution:

0.15 V. This is the same polarity as the emf driving the current.

Exercise:**Problem:**

The current $I(t)$ through a 5.0-mH inductor varies with time, as shown below. The resistance of the inductor is 5.0Ω . Calculate the voltage across the inductor at $t = 2.0 \text{ ms}$, $t = 4.0 \text{ ms}$, and $t = 8.0 \text{ ms}$.



Exercise:

Problem:

A long, cylindrical solenoid with 100 turns per centimeter has a radius of 1.5 cm. (a) Neglecting end effects, what is the self-inductance per unit length of the solenoid? (b) If the current through the solenoid changes at the rate 5.0 A/s, what is the emf induced per unit length?

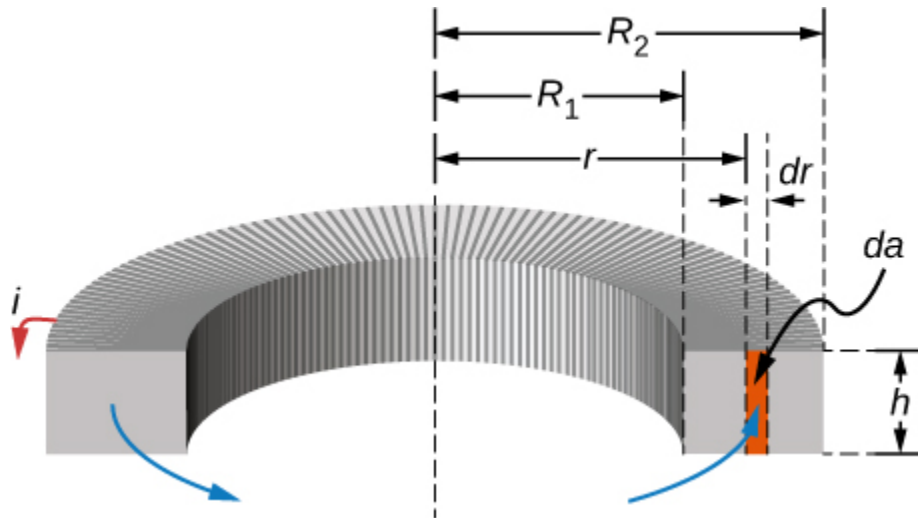
Solution:

a. 0.089 H/m; b. 0.44 V/m

Exercise:

Problem:

Suppose that a rectangular toroid has 2000 windings and a self-inductance of 0.040 H. If $h = 0.10$ m, what is the ratio of its outer radius to its inner radius?



Exercise:

Problem:

What is the self-inductance per meter of a coaxial cable whose inner radius is 0.50 mm and whose outer radius is 4.00 mm?

Solution:

$$\frac{L}{l} = 4.16 \times 10^{-7} \text{ H/m}$$

Glossary

inductor

part of an electrical circuit to provide self-inductance, which is symbolized by a coil of wire

self-inductance

effect of the device inducing emf in itself

Energy in a Magnetic Field

By the end of this section, you will be able to:

- Explain how energy can be stored in a magnetic field
- Derive the equation for energy stored in a coaxial cable given the magnetic energy density

The energy of a capacitor is stored in the electric field between its plates. Similarly, an inductor has the capability to store energy, but in its magnetic field. This energy can be found by integrating the **magnetic energy density**,

Equation:

$$u_m = \frac{B^2}{2\mu_0}$$

over the appropriate volume. To understand where this formula comes from, let's consider the long, cylindrical solenoid of the previous section. Again using the infinite solenoid approximation, we can assume that the magnetic field is essentially constant and given by $B = \mu_0 n I$ everywhere inside the solenoid. Thus, the energy stored in a solenoid or the magnetic energy density times volume is equivalent to

Equation:

$$U = u_m(V) = \frac{(\mu_0 n I)^2}{2\mu_0} (Al) = \frac{1}{2} (\mu_0 n^2 Al) I^2.$$

With the substitution of [\[link\]](#), this becomes

Note:

Equation:

$$U = \frac{1}{2}LI^2.$$

Although derived for a special case, this equation gives the energy stored in the magnetic field of *any* inductor. We can see this by considering an arbitrary inductor through which a changing current is passing. At any instant, the magnitude of the induced emf is $\varepsilon = Ldi/dt$, where i is the induced current at that instance. Therefore, the power absorbed by the inductor is

Equation:

$$P = \varepsilon i = L \frac{di}{dt} i.$$

The total energy stored in the magnetic field when the current increases from 0 to I in a time interval from 0 to t can be determined by integrating this expression:

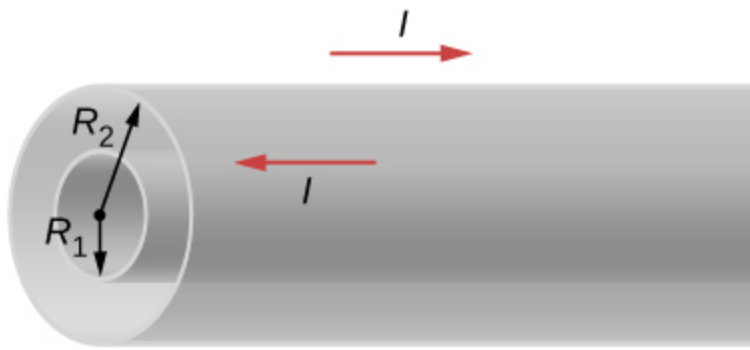
Equation:

$$U = \int_0^t P dt' = \int_0^t L \frac{di}{dt'} i dt' = L \int_0^I i di = \frac{1}{2}LI^2.$$

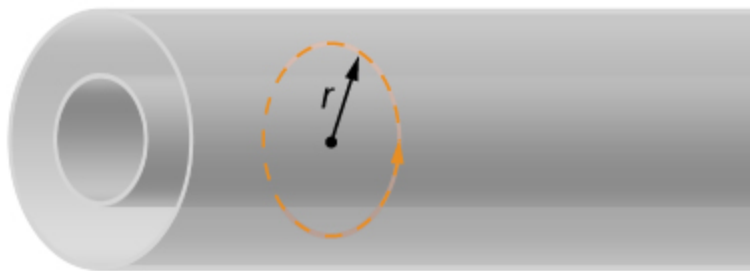
Example:

Self-Inductance of a Coaxial Cable

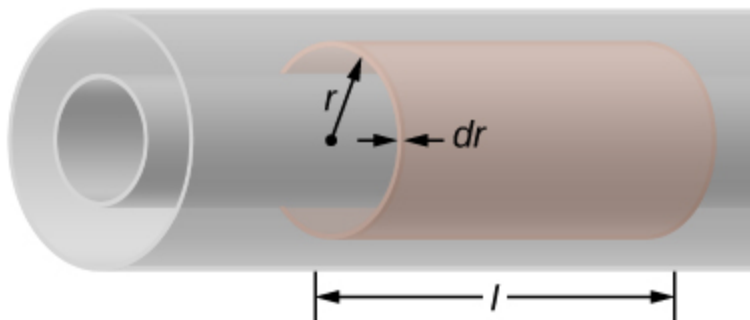
[\[link\]](#) shows two long, concentric cylindrical shells of radii R_1 and R_2 . As discussed in [Capacitance](#) on capacitance, this configuration is a simplified representation of a coaxial cable. The capacitance per unit length of the cable has already been calculated. Now (a) determine the magnetic energy stored per unit length of the coaxial cable and (b) use this result to find the self-inductance per unit length of the cable.



(a)



(b)



(c)

(a) A coaxial cable is represented here by two hollow, concentric cylindrical conductors along which electric current flows in opposite directions. (b) The magnetic field between the conductors can be found by applying Ampère's law to the dashed path. (c) The cylindrical shell is

used to find the magnetic energy stored in a length l of the cable.

Strategy

The magnetic field both inside and outside the coaxial cable is determined by Ampère's law. Based on this magnetic field, we can use [\[link\]](#) to calculate the energy density of the magnetic field. The magnetic energy is calculated by an integral of the magnetic energy density times the differential volume over the cylindrical shell. After the integration is carried out, we have a closed-form solution for part (a). The self-inductance per unit length is determined based on this result and [\[link\]](#).

Solution

- a. We determine the magnetic field between the conductors by applying Ampère's law to the dashed circular path shown in [\[link\]](#)(b). Because of the cylindrical symmetry, \vec{B} is constant along the path, and

Equation:

$$\oint \vec{B} \cdot d\vec{l} = B(2\pi r) = \mu_0 I.$$

This gives us

Equation:

$$B = \frac{\mu_0 I}{2\pi r}.$$

In the region outside the cable, a similar application of Ampère's law shows that $B = 0$, since no net current crosses the area bounded by a circular path where $r > R_2$. This argument also holds when $r < R_1$; that is, in the region within the inner cylinder. All the magnetic energy of the cable is therefore stored between the two conductors. Since the energy density of the magnetic field is

Equation:

$$u_m = \frac{B^2}{2\mu_0}$$

the energy stored in a cylindrical shell of inner radius r , outer radius $r + dr$, and length l (see part (c) of the figure) is

Equation:

$$u_m = \frac{\mu_0 I^2}{8\pi^2 r^2}.$$

Thus, the total energy of the magnetic field in a length l of the cable is

Equation:

$$U = \int_{R_1}^{R_2} dU = \int_{R_1}^{R_2} \frac{\mu_0 I^2}{8\pi^2 r^2} (2\pi r l) dr = \frac{\mu_0 I^2 l}{4\pi} \ln \frac{R_2}{R_1},$$

and the energy per unit length is $(\mu_0 I^2 / 4\pi) \ln(R_2 / R_1)$.

b. From [\[link\]](#),

Equation:

$$U = \frac{1}{2} L I^2,$$

where L is the self-inductance of a length l of the coaxial cable.

Equating the previous two equations, we find that the self-inductance per unit length of the cable is

Equation:

$$\frac{L}{l} = \frac{\mu_0}{2\pi} \ln \frac{R_2}{R_1}.$$

Significance

The inductance per unit length depends only on the inner and outer radii as seen in the result. To increase the inductance, we could either increase the outer radius (R_2) or decrease the inner radius (R_1). In the limit as the two radii become equal, the inductance goes to zero. In this limit, there is no

coaxial cable. Also, the magnetic energy per unit length from part (a) is proportional to the square of the current.

Note:

Exercise:

Problem:

Check Your Understanding How much energy is stored in the inductor of [\[link\]](#) after the current reaches its maximum value?

Solution:

0.50 J

Summary

- The energy stored in an inductor U is

Equation:

$$U = \frac{1}{2}LI^2.$$

- The self-inductance per unit length of coaxial cable is

Equation:

$$\frac{L}{l} = \frac{\mu_0}{2\pi} \ln \frac{R_2}{R_1}.$$

Conceptual Questions

Exercise:

Problem: Show that $LI^2/2$ has units of energy.

Problems

Exercise:

Problem:

At the instant a current of 0.20 A is flowing through a coil of wire, the energy stored in its magnetic field is 6.0×10^{-3} J. What is the self-inductance of the coil?

Exercise:

Problem:

Suppose that a rectangular toroid has 2000 windings and a self-inductance of 0.040 H. If $h = 0.10$ m, what is the current flowing through a rectangular toroid when the energy in its magnetic field is 2.0×10^{-6} J?

Solution:

0.01 A

Exercise:

Problem:

Solenoid A is tightly wound while solenoid B has windings that are evenly spaced with a gap equal to the diameter of the wire. The solenoids are otherwise identical. Determine the ratio of the energies stored per unit length of these solenoids when the same current flows through each.

Exercise:

Problem:

A 10-H inductor carries a current of 20 A. How much ice at 0° C could be melted by the energy stored in the magnetic field of the inductor? (*Hint:* Use the value $L_f = 334 \text{ J/g}$ for ice.)

Solution:

6.0 g

Exercise:**Problem:**

A coil with a self-inductance of 3.0 H and a resistance of 100 Ω carries a steady current of 2.0 A. (a) What is the energy stored in the magnetic field of the coil? (b) What is the energy per second dissipated in the resistance of the coil?

Exercise:**Problem:**

A current of 1.2 A is flowing in a coaxial cable whose outer radius is five times its inner radius. What is the magnetic field energy stored in a 3.0-m length of the cable?

Solution:

$$U_m = 7.0 \times 10^{-7} \text{ J}$$

Glossary

magnetic energy density

energy stored per volume in a magnetic field

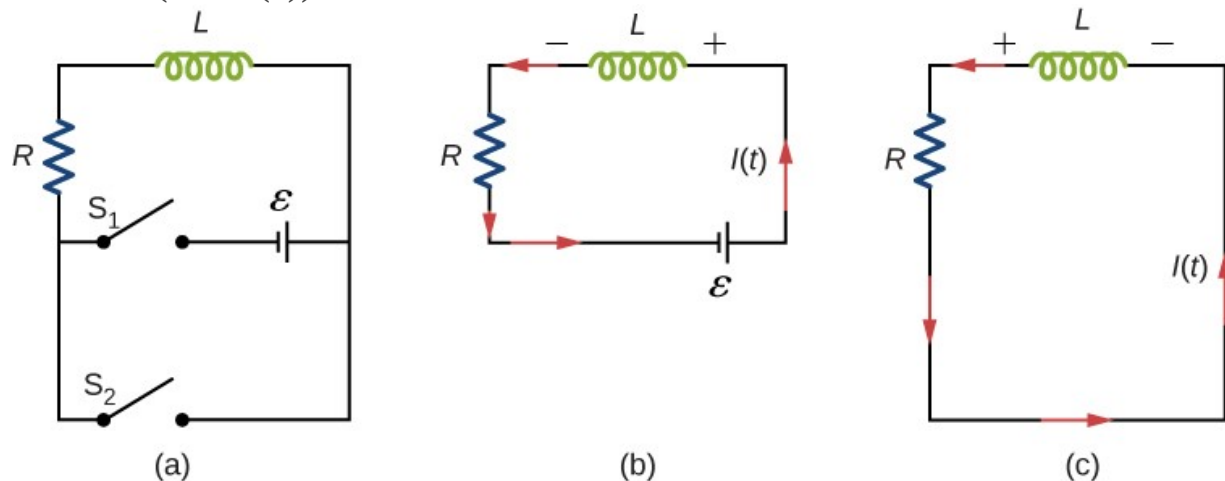
RL Circuits

By the end of this section, you will be able to:

- Analyze circuits that have an inductor and resistor in series
- Describe how current and voltage exponentially grow or decay based on the initial conditions

A circuit with resistance and self-inductance is known as an *RL* circuit.

[\[link\]](#)(a) shows an *RL* circuit consisting of a resistor, an inductor, a constant source of emf, and switches S_1 and S_2 . When S_1 is closed, the circuit is equivalent to a single-loop circuit consisting of a resistor and an inductor connected across a source of emf ([\[link\]](#)(b)). When S_1 is opened and S_2 is closed, the circuit becomes a single-loop circuit with only a resistor and an inductor ([\[link\]](#)(c)).



(a) An *RL* circuit with switches S_1 and S_2 . (b) The equivalent circuit with S_1 closed and S_2 open. (c) The equivalent circuit after S_1 is opened and S_2 is closed.

We first consider the *RL* circuit of [\[link\]](#)(b). Once S_1 is closed and S_2 is open, the source of emf produces a current in the circuit. If there were no self-inductance in the circuit, the current would rise immediately to a steady value of \mathcal{E}/R . However, from Faraday's law, the increasing current produces an emf $V_L = -L(dI/dt)$ across the inductor. In accordance with

Lenz's law, the induced emf counteracts the increase in the current and is directed as shown in the figure. As a result, $I(t)$ starts at zero and increases asymptotically to its final value.

Applying Kirchhoff's loop rule to this circuit, we obtain

Equation:

$$\varepsilon - L \frac{dI}{dt} - IR = 0,$$

which is a first-order differential equation for $I(t)$. Notice its similarity to the equation for a capacitor and resistor in series (See [RC Circuits](#)).

Similarly, the solution to [\[link\]](#) can be found by making substitutions in the equations relating the capacitor to the inductor. This gives

Note:

Equation:

$$I(t) = \frac{\varepsilon}{R} \left(1 - e^{-Rt/L} \right) = \frac{\varepsilon}{R} \left(1 - e^{-t/\tau_L} \right),$$

where

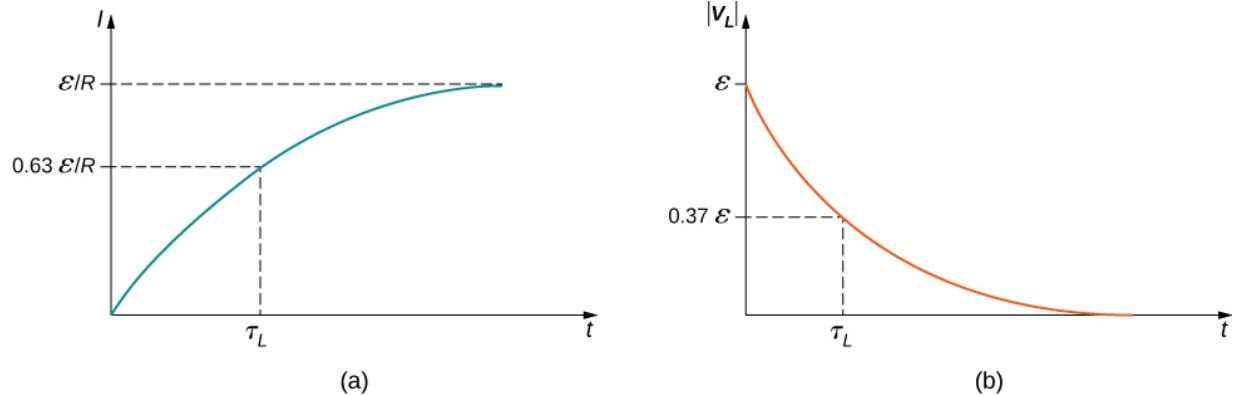
Note:

Equation:

$$\tau_L = L/R$$

is the **inductive time constant** of the circuit.

The current $I(t)$ is plotted in [\[link\]](#)(a). It starts at zero, and as $t \rightarrow \infty$, $I(t)$ approaches ε/R asymptotically. The induced emf $V_L(t)$ is directly proportional to dI/dt , or the slope of the curve. Hence, while at its greatest immediately after the switches are thrown, the induced emf decreases to zero with time as the current approaches its final value of ε/R . The circuit then becomes equivalent to a resistor connected across a source of emf.



Time variation of (a) the electric current and (b) the magnitude of the induced voltage across the coil in the circuit of [\[link\]](#)(b).

The energy stored in the magnetic field of an inductor is

Equation:

$$U_L = \frac{1}{2}LI^2.$$

Thus, as the current approaches the maximum current ε/R , the stored energy in the inductor increases from zero and asymptotically approaches a maximum of $L(\varepsilon/R)^2/2$.

The time constant τ_L tells us how rapidly the current increases to its final value. At $t = \tau_L$, the current in the circuit is, from [\[link\]](#),

Equation:

$$I(\tau_L) = \frac{\varepsilon}{R}(1 - e^{-1}) = 0.63\frac{\varepsilon}{R},$$

which is 63% of the final value ε/R . The smaller the inductive time constant $\tau_L = L/R$, the more rapidly the current approaches ε/R .

We can find the time dependence of the induced voltage across the inductor in this circuit by using $V_L(t) = -L(dI/dt)$ and [\[link\]](#):

Equation:

$$V_L(t) = -L \frac{dI}{dt} = -\varepsilon e^{-t/\tau_L}.$$

The magnitude of this function is plotted in [\[link\]](#)(b). The greatest value of $L(dI/dt)$ is ε ; it occurs when dI/dt is greatest, which is immediately after S_1 is closed and S_2 is opened. In the approach to steady state, dI/dt decreases to zero. As a result, the voltage across the inductor also vanishes as $t \rightarrow \infty$.

The time constant τ_L also tells us how quickly the induced voltage decays. At $t = \tau_L$, the magnitude of the induced voltage is

Equation:

$$|V_L(\tau_L)| = \varepsilon e^{-1} = 0.37\varepsilon = 0.37V(0).$$

The voltage across the inductor therefore drops to about 37% of its initial value after one time constant. The shorter the time constant τ_L , the more rapidly the voltage decreases.

After enough time has elapsed so that the current has essentially reached its final value, the positions of the switches in [\[link\]](#)(a) are reversed, giving us the circuit in part (c). At $t = 0$, the current in the circuit is $I(0) = \varepsilon/R$.

With Kirchhoff's loop rule, we obtain

Equation:

$$IR + L \frac{dI}{dt} = 0.$$

The solution to this equation is similar to the solution of the equation for a discharging capacitor, with similar substitutions. The current at time t is then

Equation:

$$I(t) = \frac{\varepsilon}{R} e^{-t/\tau_L}.$$

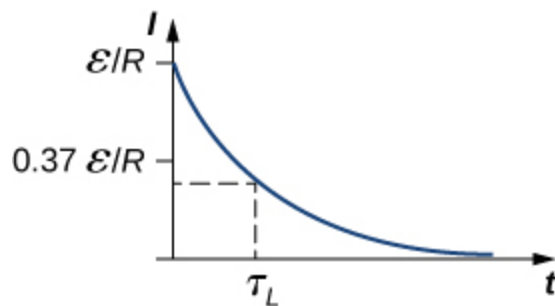
The current starts at $I(0) = \varepsilon/R$ and decreases with time as the energy stored in the inductor is depleted ([\[link\]](#)).

The time dependence of the voltage across the inductor can be determined from $V_L = -L(dI/dt)$:

Equation:

$$V_L(t) = \varepsilon e^{-t/\tau_L}.$$

This voltage is initially $V_L(0) = \varepsilon$, and it decays to zero like the current. The energy stored in the magnetic field of the inductor, $LI^2/2$, also decreases exponentially with time, as it is dissipated by Joule heating in the resistance of the circuit.



Time variation of electric current in the RL circuit of [\[link\]](#)(c). The induced

voltage across the coil also decays exponentially.

Example:

An RL Circuit with a Source of emf

In the circuit of [\[link\]](#)(a), let $\varepsilon = 2.0\text{ V}$, $R = 4.0\ \Omega$, and $L = 4.0\text{ H}$. With S_1 closed and S_2 open ([\[link\]](#)(b)), (a) what is the time constant of the circuit? (b) What are the current in the circuit and the magnitude of the induced emf across the inductor at $t = 0$, at $t = 2.0\tau_L$, and as $t \rightarrow \infty$?

Strategy

The time constant for an inductor and resistor in a series circuit is calculated using [\[link\]](#). The current through and voltage across the inductor are calculated by the scenarios detailed from [\[link\]](#) and [\[link\]](#).

Solution

- a. The inductive time constant is

Equation:

$$\tau_L = \frac{L}{R} = \frac{4.0\text{ H}}{4.0\ \Omega} = 1.0\text{ s}.$$

- b. The current in the circuit of [\[link\]](#)(b) increases according to [\[link\]](#):

Equation:

$$I(t) = \frac{\varepsilon}{R}(1 - e^{-t/\tau_L}).$$

At $t = 0$,

Equation:

$$(1 - e^{-t/\tau_L}) = (1 - 1) = 0; \text{ so } I(0) = 0.$$

At $t = 2.0\tau_L$ and $t \rightarrow \infty$, we have, respectively,

Equation:

$$I(2.0\tau_L) = \frac{\varepsilon}{R}(1 - e^{-2.0}) = (0.50 \text{ A})(0.86) = 0.43 \text{ A},$$

and

Equation:

$$I(\infty) = \frac{\varepsilon}{R} = 0.50 \text{ A}.$$

From [\[link\]](#), the magnitude of the induced emf decays as

Equation:

$$|V_L(t)| = \varepsilon e^{-t/\tau_L}.$$

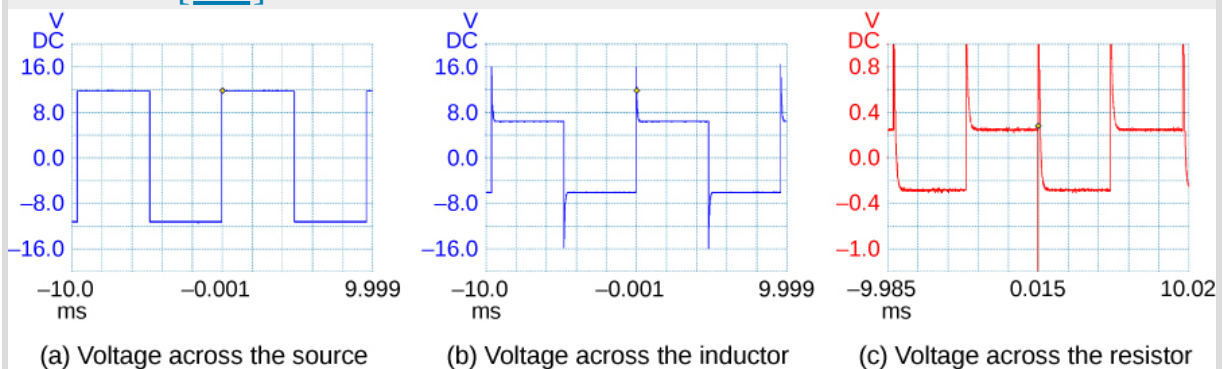
At $t = 0$, $t = 2.0\tau_L$, and as $t \rightarrow \infty$, we obtain

Equation:

$$\begin{aligned} |V_L(0)| &= \varepsilon = 2.0 \text{ V}, \\ |V_L(2.0\tau_L)| &= (2.0 \text{ V}) e^{-2.0} = 0.27 \text{ V} \\ &\text{and} \\ |V_L(\infty)| &= 0. \end{aligned}$$

Significance

If the time of the measurement were much larger than the time constant, we would not see the decay or growth of the voltage across the inductor or resistor. The circuit would quickly reach the asymptotic values for both of these. See [\[link\]](#).



A generator in an RL circuit produces a square-pulse output in which the voltage oscillates between zero and some set value. These oscilloscope traces show (a) the voltage across the source; (b) the voltage across the inductor; (c) the voltage across the resistor.

Example:

An RL Circuit without a Source of emf

After the current in the RL circuit of [\[link\]](#) has reached its final value, the positions of the switches are reversed so that the circuit becomes the one shown in [\[link\]](#)(c). (a) How long does it take the current to drop to half its initial value? (b) How long does it take before the energy stored in the inductor is reduced to 1.0% of its maximum value?

Strategy

The current in the inductor will now decrease as the resistor dissipates this energy. Therefore, the current falls as an exponential decay. We can also use that same relationship as a substitution for the energy in an inductor formula to find how the energy decreases at different time intervals.

Solution

- a. With the switches reversed, the current decreases according to

Equation:

$$I(t) = \frac{\varepsilon}{R} e^{-t/\tau_L} = I(0) e^{-t/\tau_L}.$$

At a time t when the current is one-half its initial value, we have

Equation:

$$I(t) = 0.50I(0) \text{ so } e^{-t/\tau_L} = 0.50,$$

and

Equation:

$$t = -[\ln(0.50)]\tau_L = 0.69(1.0 \text{ s}) = 0.69 \text{ s},$$

where we have used the inductive time constant found in [\[link\]](#).

b. The energy stored in the inductor is given by

Equation:

$$U_L(t) = \frac{1}{2} L [I(t)]^2 = \frac{1}{2} L \left(\frac{\varepsilon}{R} e^{-t/\tau_L} \right)^2 = \frac{L \varepsilon^2}{2 R^2} e^{-2t/\tau_L}.$$

If the energy drops to 1.0% of its initial value at a time t , we have

Equation:

$$U_L(t) = (0.010) U_L(0) \text{ or } \frac{L \varepsilon^2}{2 R^2} e^{-2t/\tau_L} = (0.010) \frac{L \varepsilon^2}{2 R^2}.$$

Upon canceling terms and taking the natural logarithm of both sides, we obtain

Equation:

$$-\frac{2t}{\tau_L} = \ln(0.010),$$

so

Equation:

$$t = -\frac{1}{2} \tau_L \ln(0.010).$$

Since $\tau_L = 1.0$ s, the time it takes for the energy stored in the inductor to decrease to 1.0% of its initial value is

Equation:

$$t = -\frac{1}{2} (1.0 \text{ s}) \ln(0.010) = 2.3 \text{ s}.$$

Significance

This calculation only works if the circuit is at maximum current in situation (b) prior to this new situation. Otherwise, we start with a lower initial current, which will decay by the same relationship.

Note:

Exercise:

Problem:

Check Your Understanding Verify that RC and L/R have the dimensions of time.

Note:

Exercise:

Problem:

Check Your Understanding (a) If the current in the circuit of in [\[link\]](#) (b) increases to 90% of its final value after 5.0 s, what is the inductive time constant? (b) If $R = 20\ \Omega$, what is the value of the self-inductance? (c) If the $20\text{-}\Omega$ resistor is replaced with a $100\text{-}\Omega$ resistor, what is the time taken for the current to reach 90% of its final value?

Solution:

a. 2.2 s; b. 43 H; c. 1.0 s

Note:

Exercise:

Problem:

Check Your Understanding For the circuit of in [\[link\]](#) (b), show that when steady state is reached, the difference in the total energies produced by the battery and dissipated in the resistor is equal to the energy stored in the magnetic field of the coil.

Summary

- When a series connection of a resistor and an inductor—an RL circuit—is connected to a voltage source, the time variation of the current is $I(t) = \frac{\varepsilon}{R}(1 - e^{-Rt/L}) = \frac{\varepsilon}{R}(1 - e^{-t/\tau_L})$ (turning on), where the initial current is $I_0 = \varepsilon/R$.
- The characteristic time constant τ is $\tau_L = L/R$, where L is the inductance and R is the resistance.
- In the first time constant τ , the current rises from zero to $0.632I_0$, and to 0.632 of the remainder in every subsequent time interval τ .
- When the inductor is shorted through a resistor, current decreases as $I(t) = \frac{\varepsilon}{R}e^{-t/\tau_L}$ (turning off).
Current falls to $0.368I_0$ in the first time interval τ , and to 0.368 of the remainder toward zero in each subsequent time τ .

Conceptual Questions

Exercise:

Problem:

Use Lenz's law to explain why the initial current in the RL circuit of [\[link\]](#)(b) is zero.

Solution:

As current flows through the inductor, there is a back current by Lenz's law that is created to keep the net current at zero amps, the initial current.

Exercise:

Problem:

When the current in the RL circuit of [\[link\]](#)(b) reaches its final value ε/R , what is the voltage across the inductor? Across the resistor?

Exercise:

Problem:

Does the time required for the current in an RL circuit to reach any fraction of its steady-state value depend on the emf of the battery?

Solution:

no

Exercise:**Problem:**

An inductor is connected across the terminals of a battery. Does the current that eventually flows through the inductor depend on the internal resistance of the battery? Does the time required for the current to reach its final value depend on this resistance?

Exercise:**Problem:**

At what time is the voltage across the inductor of the RL circuit of [\[link\]](#)(b) a maximum?

Solution:

At $t = 0$, or when the switch is first thrown.

Exercise:**Problem:**

In the simple RL circuit of [\[link\]](#)(b), can the emf induced across the inductor ever be greater than the emf of the battery used to produce the current?

Exercise:

Problem:

If the emf of the battery of [\[link\]](#)(b) is reduced by a factor of 2, by how much does the steady-state energy stored in the magnetic field of the inductor change?

Solution:

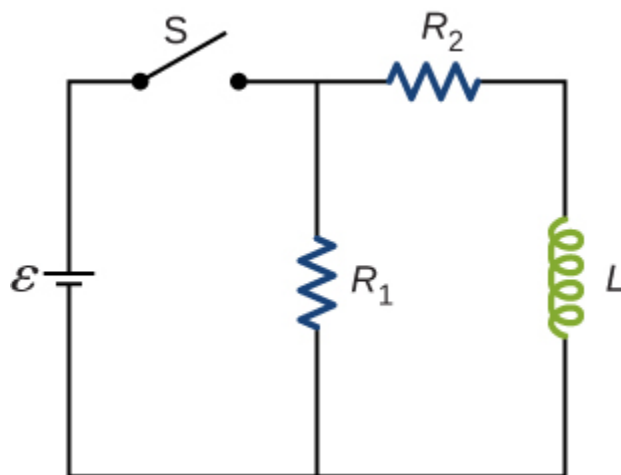
1/4

Exercise:**Problem:**

A steady current flows through a circuit with a large inductive time constant. When a switch in the circuit is opened, a large spark occurs across the terminals of the switch. Explain.

Exercise:**Problem:**

Describe how the currents through R_1 and R_2 shown below vary with time after switch S is closed.

**Solution:**

Initially, $I_{R1} = \frac{\varepsilon}{R_1}$ and $I_{R2} = 0$, and after a long time has passed, $I_{R1} = \frac{\varepsilon}{R_1}$ and $I_{R2} = \frac{\varepsilon}{R_2}$.

Exercise:

Problem: Discuss possible practical applications of RL circuits.

Problems

Exercise:

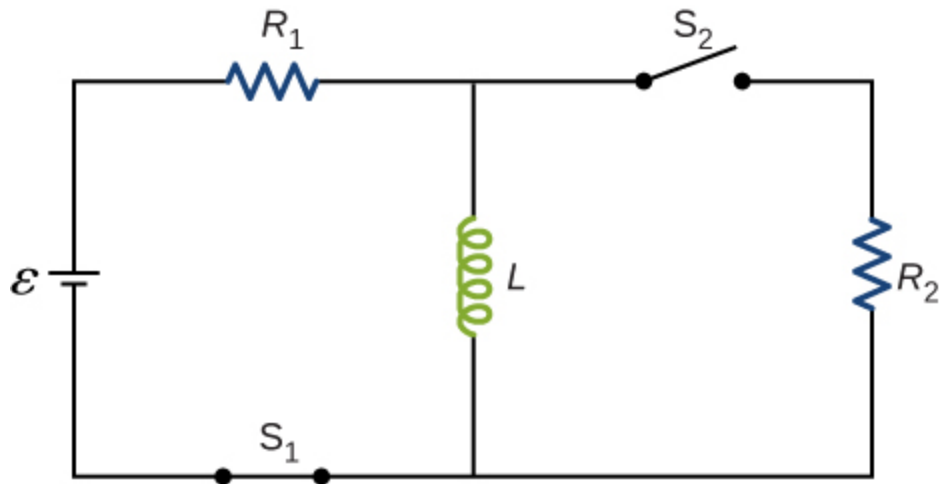
Problem:

In [\[link\]](#), $\varepsilon = 12 \text{ V}$, $L = 20 \text{ mH}$, and $R = 5.0 \Omega$. Determine (a) the time constant of the circuit, (b) the initial current through the resistor, (c) the final current through the resistor, (d) the current through the resistor when $t = 2\tau_L$, and (e) the voltages across the inductor and the resistor when $t = 2\tau_L$.

Exercise:

Problem:

For the circuit shown below, $\varepsilon = 20 \text{ V}$, $L = 4.0 \text{ mH}$, and $R = 5.0 \Omega$. After steady state is reached with S_1 closed and S_2 open, S_2 is closed and immediately thereafter (at $t = 0$) S_1 is opened. Determine (a) the current through L at $t = 0$, (b) the current through L at $t = 4.0 \times 10^{-4} \text{ s}$, and (c) the voltages across L and R_1 at $t = 4.0 \times 10^{-4} \text{ s}$. $R_1 = R_2 = R$.



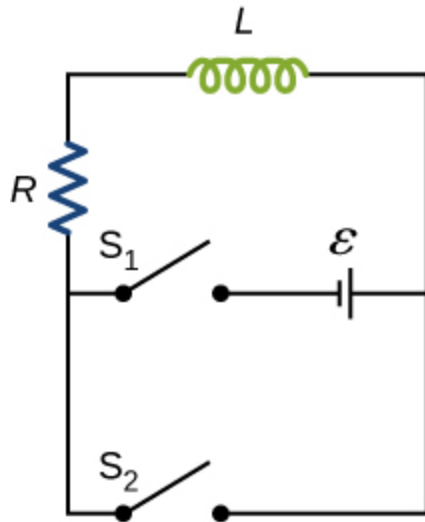
Solution:

a. 4.0 A; b. 2.4 A; c. on R : $V = 12$ V; on L : $V = 7.9$ V

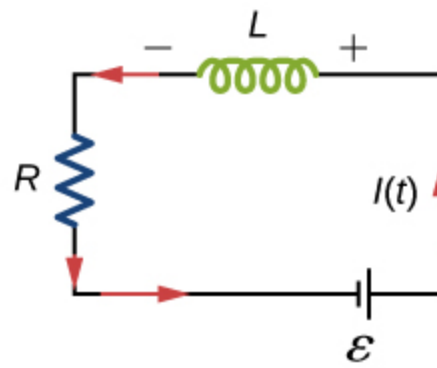
Exercise:

Problem:

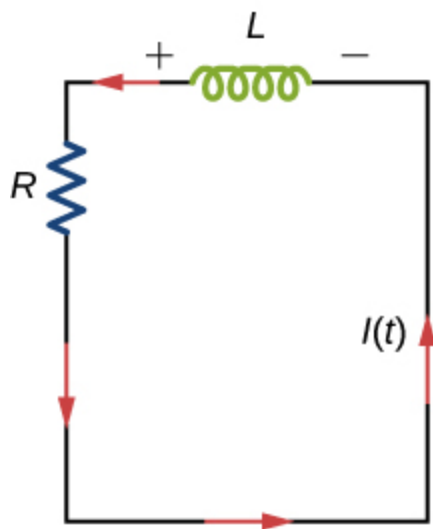
The current in the RL circuit shown here increases to 40% of its steady-state value in 2.0 s. What is the time constant of the circuit?



(a)



(b)

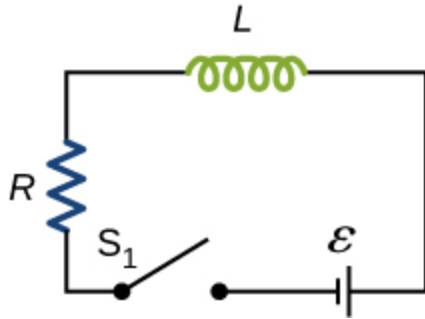


(c)

Exercise:

Problem:

How long after switch S_1 is thrown does it take the current in the circuit shown to reach half its maximum value? Express your answer in terms of the time constant of the circuit.



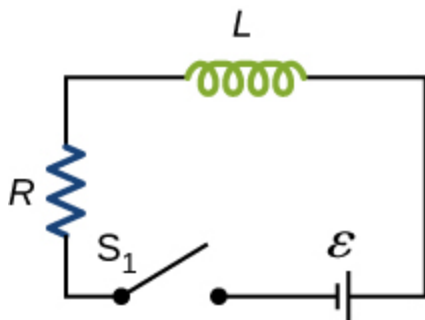
Solution:

$$0.69\tau$$

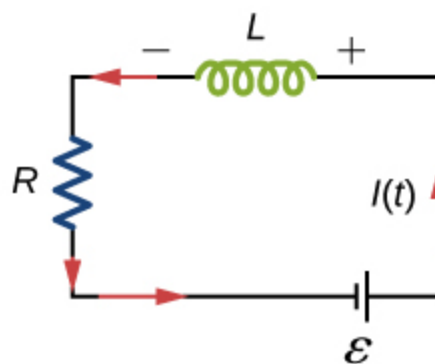
Exercise:

Problem:

Examine the circuit shown below in part (a). Determine dI/dt at the instant after the switch is thrown in the circuit of (a), thereby producing the circuit of (b). Show that if I were to continue to increase at this initial rate, it would reach its maximum \mathcal{E}/R in one time constant.



(a)

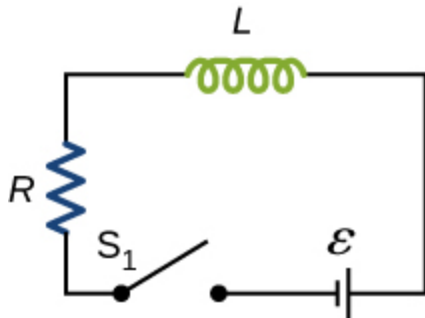


(b)

Exercise:

Problem:

The current in the RL circuit shown below reaches half its maximum value in 1.75 ms after the switch S_1 is thrown. Determine (a) the time constant of the circuit and (b) the resistance of the circuit if $L = 250$ mH.

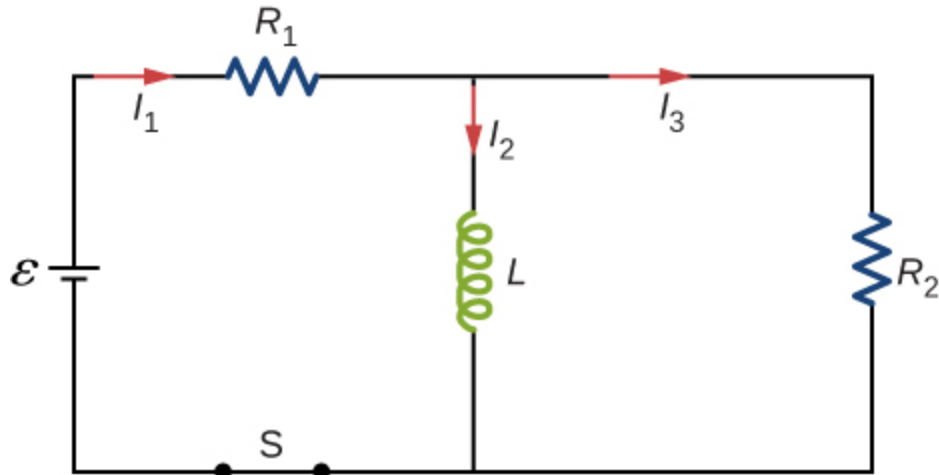


Solution:

a. 2.52 ms; b. 99.2 Ω

Exercise:**Problem:**

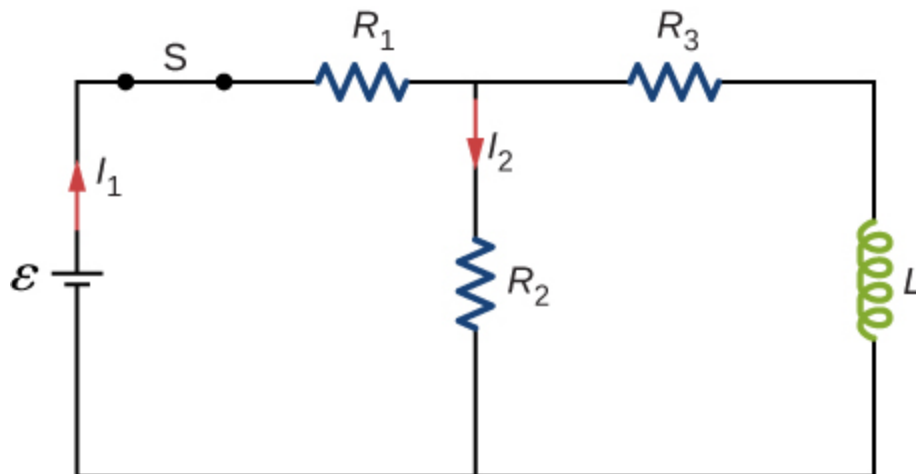
Consider the circuit shown below. Find I_1 , I_2 , and I_3 when (a) the switch S is first closed, (b) after the currents have reached steady-state values, and (c) at the instant the switch is reopened (after being closed for a long time).



Exercise:

Problem:

For the circuit shown below, $\varepsilon = 50 \text{ V}$, $R_1 = 10 \Omega$, $R_2 = R_3 = 19.4 \Omega$, and $L = 2.0 \text{ mH}$. Find the values of I_1 and I_2 (a) immediately after switch S is closed, (b) a long time after S is closed, (c) immediately after S is reopened, and (d) a long time after S is reopened.



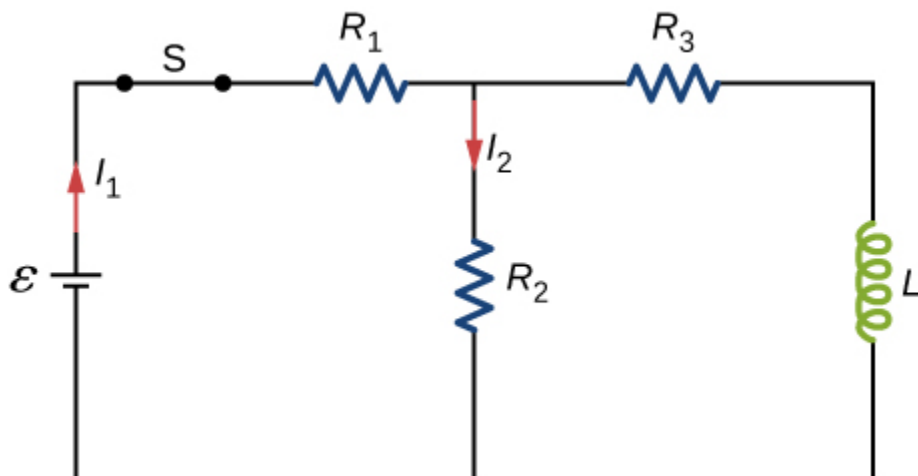
Solution:

a. $I_1 = I_2 = 1.7 \text{ A}$; b. $I_1 = 2.73 \text{ A}$, $I_2 = 1.36 \text{ A}$; c. $I_1 = 0$, $I_2 = 0.54 \text{ A}$; d. $I_1 = I_2 = 0$

Exercise:

Problem:

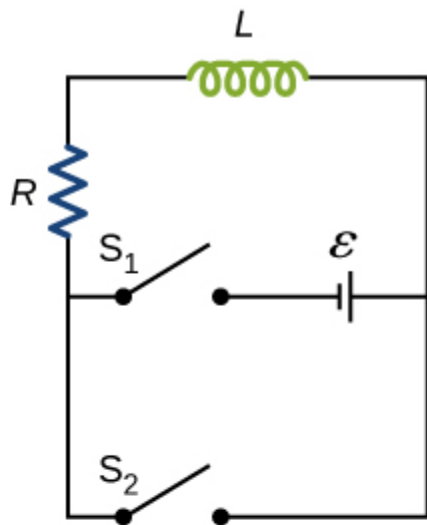
For the circuit shown below, find the current through the inductor 2.0×10^{-5} s after the switch is reopened.



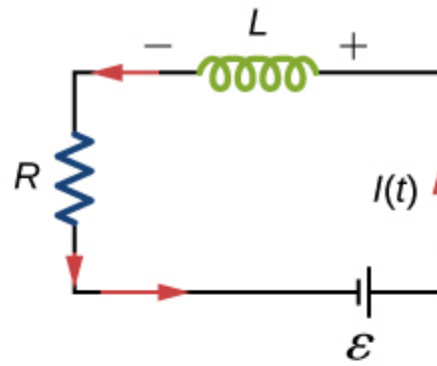
Exercise:

Problem:

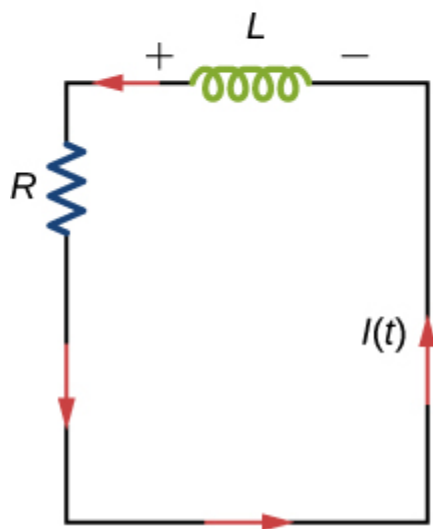
Show that for the circuit shown below, the initial energy stored in the inductor, $LI^2(0)/2$, is equal to the total energy eventually dissipated in the resistor, $\int_0^\infty I^2(t)Rdt$.



(a)



(b)



(c)

Solution:

proof

Glossary

inductive time constant

denoted by τ , the characteristic time given by quantity L/R of a particular series RL circuit

Oscillations in an LC Circuit

By the end of this section, you will be able to:

- Explain why charge or current oscillates between a capacitor and inductor, respectively, when wired in series
- Describe the relationship between the charge and current oscillating between a capacitor and inductor wired in series

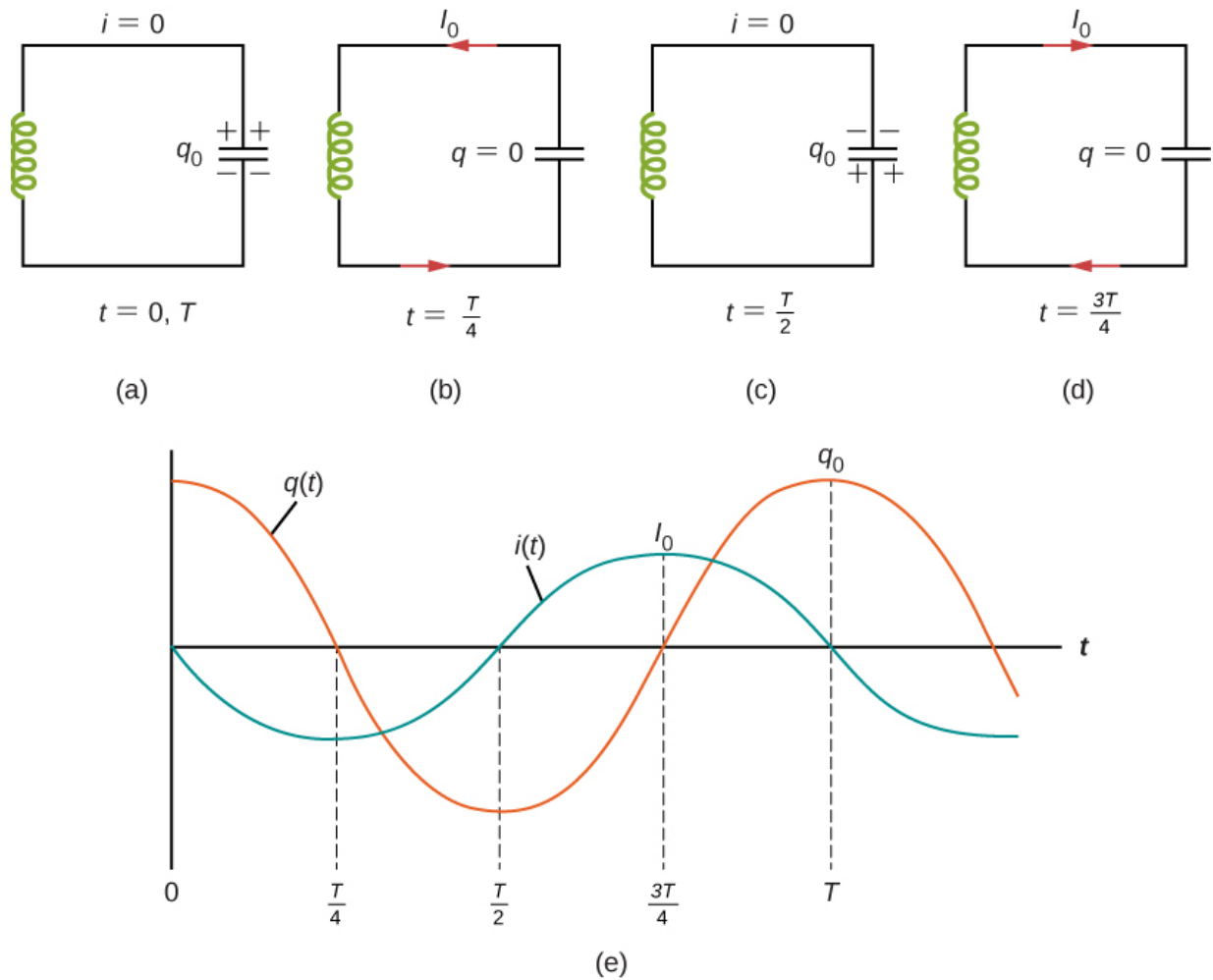
It is worth noting that both capacitors and inductors store energy, in their electric and magnetic fields, respectively. A circuit containing both an inductor (L) and a capacitor (C) can oscillate without a source of emf by shifting the energy stored in the circuit between the electric and magnetic fields. Thus, the concepts we develop in this section are directly applicable to the exchange of energy between the electric and magnetic fields in electromagnetic waves, or light. We start with an idealized circuit of zero resistance that contains an inductor and a capacitor, an ***LC circuit***.

An *LC* circuit is shown in [\[link\]](#). If the capacitor contains a charge q_0 before the switch is closed, then all the energy of the circuit is initially stored in the electric field of the capacitor ([\[link\]](#)(a)). This energy is

Equation:

$$U_C = \frac{1}{2} \frac{q_0^2}{C}.$$

When the switch is closed, the capacitor begins to discharge, producing a current in the circuit. The current, in turn, creates a magnetic field in the inductor. The net effect of this process is a transfer of energy from the capacitor, with its diminishing electric field, to the inductor, with its increasing magnetic field.



(a–d) The oscillation of charge storage with changing directions of current in an LC circuit. (e) The graphs show the distribution of charge and current between the capacitor and inductor.

In [\[link\]](#)(b), the capacitor is completely discharged and all the energy is stored in the magnetic field of the inductor. At this instant, the current is at its maximum value I_0 and the energy in the inductor is

Equation:

$$U_L = \frac{1}{2}LI_0^2.$$

Since there is no resistance in the circuit, no energy is lost through Joule heating; thus, the maximum energy stored in the capacitor is equal to the maximum energy stored at a later time in the inductor:

Equation:

$$\frac{1}{2} \frac{q_0^2}{C} = \frac{1}{2} L I_0^2.$$

At an arbitrary time when the capacitor charge is $q(t)$ and the current is $i(t)$, the total energy U in the circuit is given by

Equation:

$$\frac{q^2(t)}{2C} + \frac{L i^2(t)}{2}.$$

Because there is no energy dissipation,

Equation:

$$U = \frac{1}{2} \frac{q^2}{C} + \frac{1}{2} L i^2 = \frac{1}{2} \frac{q_0^2}{C} = \frac{1}{2} L I_0^2.$$

After reaching its maximum I_0 , the current $i(t)$ continues to transport charge between the capacitor plates, thereby recharging the capacitor. Since the inductor resists a change in current, current continues to flow, even though the capacitor is discharged. This continued current causes the capacitor to charge with opposite polarity. The electric field of the capacitor increases while the magnetic field of the inductor diminishes, and the overall effect is a transfer of energy from the inductor *back* to the capacitor. From the law of energy conservation, the maximum charge that the capacitor re-acquires is q_0 . However, as [\[link\]\(c\)](#) shows, the capacitor plates are charged *opposite* to what they were initially.

When fully charged, the capacitor once again transfers its energy to the inductor until it is again completely discharged, as shown in [\[link\]\(d\)](#). Then, in the last part of this cyclic process, energy flows back to the capacitor, and the initial state of the circuit is restored.

We have followed the circuit through one complete cycle. Its electromagnetic oscillations are analogous to the mechanical oscillations of a mass at the end of a spring. In this latter case, energy is transferred back and forth between the mass, which has kinetic energy $mv^2/2$, and the spring, which has potential energy $kx^2/2$. With the absence of friction in the mass-spring system, the oscillations would continue indefinitely. Similarly, the oscillations of an LC circuit with no resistance would continue forever if undisturbed; however, this ideal zero-resistance LC circuit is not practical, and any LC circuit will have at least a small resistance, which will radiate and lose energy over time.

The frequency of the oscillations in a resistance-free LC circuit may be found by analogy with the mass-spring system. For the circuit, $i(t) = dq(t)/dt$, the total electromagnetic energy U is

Equation:

$$U = \frac{1}{2}Li^2 + \frac{1}{2}\frac{q^2}{C}.$$

For the mass-spring system, $v(t) = dx(t)/dt$, the total mechanical energy E is

Equation:

$$E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2.$$

The equivalence of the two systems is clear. To go from the mechanical to the electromagnetic system, we simply replace m by L , v by i , k by $1/C$, and x by q . Now $x(t)$ is given by

Equation:

$$x(t) = A \cos(\omega t + \phi)$$

where $\omega = \sqrt{k/m}$. Hence, the charge on the capacitor in an LC circuit is given by

Note:

Equation:

$$q(t) = q_0 \cos(\omega t + \phi)$$

where the angular frequency of the oscillations in the circuit is

Note:

Equation:

$$\omega = \sqrt{\frac{1}{LC}}.$$

Finally, the current in the LC circuit is found by taking the time derivative of $q(t)$:

Note:

Equation:

$$i(t) = \frac{dq(t)}{dt} = -\omega q_0 \sin(\omega t + \phi).$$

The time variations of q and I are shown in [\[link\]](#)(e) for $\phi = 0$.

Example:

An LC Circuit

In an LC circuit, the self-inductance is 2.0×10^{-2} H and the capacitance is 8.0×10^{-6} F. At $t = 0$, all of the energy is stored in the capacitor, which has charge 1.2×10^{-5} C. (a) What is the angular frequency of the oscillations in the circuit? (b) What is the maximum current flowing through circuit? (c) How long does it take the capacitor to become completely discharged? (d) Find an equation that represents $q(t)$.

Strategy

The angular frequency of the LC circuit is given by [\[link\]](#). To find the maximum current, the maximum energy in the capacitor is set equal to the maximum energy in the inductor. The time for the capacitor to become discharged if it is initially charged is a quarter of the period of the cycle, so if we calculate the period of the oscillation, we can find out what a quarter of that is to find this time. Lastly, knowing the initial charge and angular frequency, we can set up a cosine equation to find $q(t)$.

Solution

- a. From [\[link\]](#), the angular frequency of the oscillations is

Equation:

$$\omega = \sqrt{\frac{1}{LC}} = \sqrt{\frac{1}{(2.0 \times 10^{-2} \text{ H})(8.0 \times 10^{-6} \text{ F})}} = 2.5 \times 10^3 \text{ rad/s}.$$

- b. The current is at its maximum I_0 when all the energy is stored in the inductor. From the law of energy conservation,

Equation:

$$\frac{1}{2}LI_0^2 = \frac{1}{2}\frac{q_0^2}{C},$$

so

Equation:

$$I_0 = \sqrt{\frac{1}{LC}}q_0 = (2.5 \times 10^3 \text{ rad/s})(1.2 \times 10^{-5} \text{ C}) = 3.0 \times 10^{-2} \text{ A}.$$

This result can also be found by an analogy to simple harmonic motion, where current and charge are the velocity and position of an oscillator.

- c. The capacitor becomes completely discharged in one-fourth of a cycle, or during a time $T/4$, where T is the period of the oscillations. Since

Equation:

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{2.5 \times 10^3 \text{ rad/s}} = 2.5 \times 10^{-3} \text{ s},$$

the time taken for the capacitor to become fully discharged is $(2.5 \times 10^{-3} \text{ s})/4 = 6.3 \times 10^{-4} \text{ s}$.

- d. The capacitor is completely charged at $t = 0$, so $q(0) = q_0$. Using [\[link\]](#), we obtain

Equation:

$$q(0) = q_0 = q_0 \cos \phi.$$

Thus, $\phi = 0$, and

Equation:

$$q(t) = (1.2 \times 10^{-5} \text{ C})\cos(2.5 \times 10^3 t).$$

Significance

The energy relationship set up in part (b) is not the only way we can equate energies. At most times, some energy is stored in the capacitor and some energy is stored in the inductor. We can put both terms on each side of the equation. By examining the circuit only when there is no charge on the capacitor or no current in the inductor, we simplify the energy equation.

Note:

Exercise:

Problem:

Check Your Understanding The angular frequency of the oscillations in an LC circuit is $2.0 \times 10^3 \text{ rad/s}$. (a) If $L = 0.10 \text{ H}$, what is C ? (b) Suppose that at $t = 0$, all the energy is stored in the inductor. What is the value of ϕ ? (c) A second identical capacitor is connected in parallel with the original capacitor. What is the angular frequency of this circuit?

Solution:

a. $2.5\mu\text{F}$; b. $\pi/2$ rad or $3\pi/2$ rad; c. 1.4×10^3 rad/s

Summary

- The energy transferred in an oscillatory manner between the capacitor and inductor in an LC circuit occurs at an angular frequency $\omega = \sqrt{\frac{1}{LC}}$.

- The charge and current in the circuit are given by
Equation:

$$\begin{aligned}q(t) &= q_0 \cos(\omega t + \phi), \\i(t) &= -\omega q_0 \sin(\omega t + \phi).\end{aligned}$$

Conceptual Questions

Exercise:

Problem:

Do Kirchhoff's rules apply to circuits that contain inductors and capacitors?

Solution:

yes

Exercise:

Problem: Can a circuit element have both capacitance and inductance?

Exercise:

Problem:

In an LC circuit, what determines the frequency and the amplitude of the energy oscillations in either the inductor or capacitor?

Solution:

The amplitude of energy oscillations depend on the initial energy of the system. The frequency in a LC circuit depends on the values of inductance and capacitance.

Problems

Exercise:

Problem:

A 5000-pF capacitor is charged to 100 V and then quickly connected to an 80-mH inductor. Determine (a) the maximum energy stored in the magnetic field of the inductor, (b) the peak value of the current, and (c) the frequency of oscillation of the circuit.

Exercise:

Problem:

The self-inductance and capacitance of an LC circuit are 0.20 mH and 5.0 pF. What is the angular frequency at which the circuit oscillates?

Solution:

$$\omega = 3.2 \times 10^7 \text{ rad/s}$$

Exercise:

Problem:

What is the self-inductance of an LC circuit that oscillates at 60 Hz when the capacitance is 10 μF ?

Exercise:

Problem:

In an oscillating LC circuit, the maximum charge on the capacitor is $2.0 \times 10^{-6} \text{ C}$ and the maximum current through the inductor is 8.0 mA. (a) What is the period of the oscillations? (b) How much time elapses between an instant when the capacitor is uncharged and the next instant when it is fully charged?

Solution:

a. $1.57 \times 10^{-6} \text{ s}$; b. $3.93 \times 10^{-7} \text{ s}$

Exercise:**Problem:**

The self-inductance and capacitance of an oscillating LC circuit are $L = 20 \text{ mH}$ and $C = 1.0 \mu\text{F}$, respectively. (a) What is the frequency of the oscillations? (b) If the maximum potential difference between the plates of the capacitor is 50 V , what is the maximum current in the circuit?

Exercise:**Problem:**

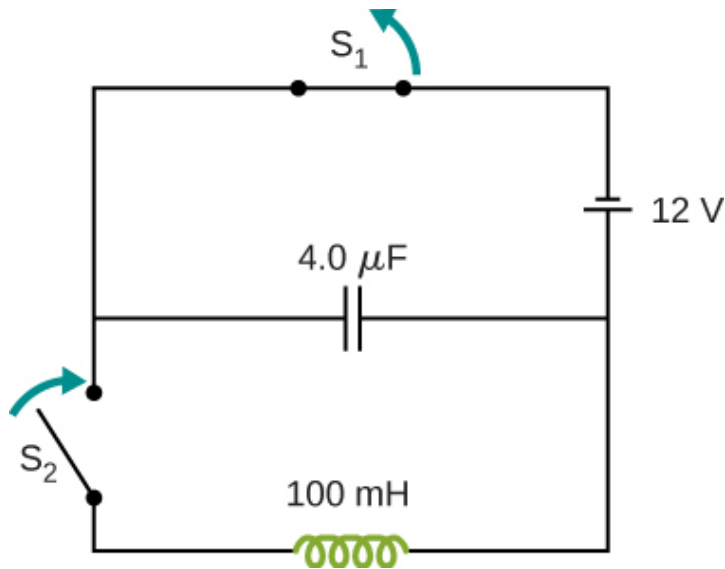
In an oscillating LC circuit, the maximum charge on the capacitor is q_m . Determine the charge on the capacitor and the current through the inductor when energy is shared equally between the electric and magnetic fields. Express your answer in terms of q_m , L , and C .

Solution:

$$q = \frac{q_m}{\sqrt{2}}, I = \frac{q_m}{\sqrt{2LC}}$$

Exercise:**Problem:**

In the circuit shown below, S_1 is opened and S_2 is closed simultaneously. Determine (a) the frequency of the resulting oscillations, (b) the maximum charge on the capacitor, (c) the maximum current through the inductor, and (d) the electromagnetic energy of the oscillating circuit.



Exercise:

Problem:

An LC circuit in an AM tuner (in a car stereo) uses a coil with an inductance of 2.5 mH and a variable capacitor. If the natural frequency of the circuit is to be adjustable over the range 540 to 1600 kHz (the AM broadcast band), what range of capacitance is required?

Solution:

$$C = \frac{1}{4\pi^2 f^2 L}$$

$$f_1 = 540 \text{ Hz}; \quad C_1 = 3.5 \times 10^{-11} \text{ F}$$

$$f_2 = 1600 \text{ Hz}; \quad C_2 = 4.0 \times 10^{-12} \text{ F}$$

Glossary

LC circuit

circuit composed of an ac source, inductor, and capacitor

RLC Series Circuits

By the end of this section, you will be able to:

- Determine the angular frequency of oscillation for a resistor, inductor, capacitor (RLC) series circuit
- Relate the RLC circuit to a damped spring oscillation

When the switch is closed in the **RLC circuit** of [\[link\]](#)(a), the capacitor begins to discharge and electromagnetic energy is dissipated by the resistor at a rate $i^2 R$. With U given by [\[link\]](#), we have

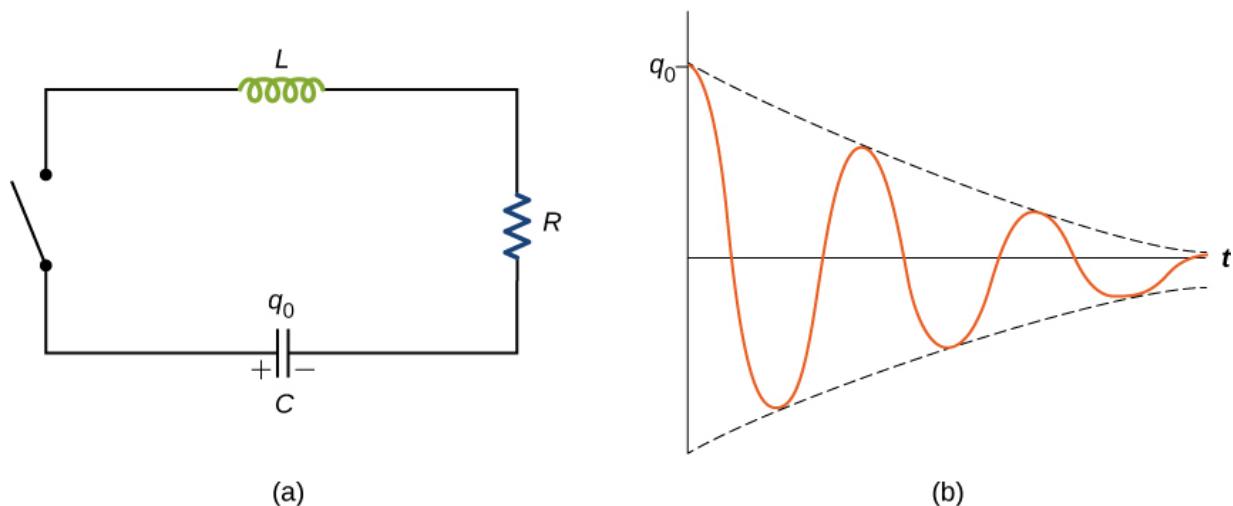
Equation:

$$\frac{dU}{dt} = \frac{q}{C} \frac{dq}{dt} + Li \frac{di}{dt} = -i^2 R$$

where i and q are time-dependent functions. This reduces to

Equation:

$$L \frac{d^2 q}{dt^2} + R \frac{dq}{dt} + \frac{1}{C} q = 0.$$



(a) An RLC circuit. Electromagnetic oscillations begin when the switch is closed. The capacitor is fully charged initially. (b) Damped oscillations of the capacitor charge are shown in this curve of charge

versus time, or q versus t . The capacitor contains a charge q_0 before the switch is closed.

This equation is analogous to

Equation:

$$m \frac{d^2x}{dt^2} + b \frac{dx}{dt} + kx = 0,$$

which is the equation of motion for a *damped mass-spring system* (you first encountered this equation in [Oscillations](#)). As we saw in that chapter, it can be shown that the solution to this differential equation takes three forms, depending on whether the angular frequency of the undamped spring is greater than, equal to, or less than $b/2m$. Therefore, the result can be underdamped ($\sqrt{k/m} > b/2m$), critically damped ($\sqrt{k/m} = b/2m$), or overdamped ($\sqrt{k/m} < b/2m$). By analogy, the solution $q(t)$ to the *RLC* differential equation has the same feature. Here we look only at the case of under-damping. By replacing m by L , b by R , k by $1/C$, and x by q in [\[link\]](#), and assuming $\sqrt{1/LC} > R/2L$, we obtain

Note:

Equation:

$$q(t) = q_0 e^{-Rt/2L} \cos(\omega t + \phi)$$

where the angular frequency of the oscillations is given by

Note:

Equation:

$$\omega' = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}$$

This underdamped solution is shown in [\[link\]](#)(b). Notice that the amplitude of the oscillations decreases as energy is dissipated in the resistor. [\[link\]](#) can be confirmed experimentally by measuring the voltage across the capacitor as a function of time. This voltage, multiplied by the capacitance of the capacitor, then gives $q(t)$.

Note:

Try an [interactive circuit construction kit](#) that allows you to graph current and voltage as a function of time. You can add inductors and capacitors to work with any combination of R , L , and C circuits with both dc and ac sources.

Note:

Try out a [circuit-based java applet website](#) that has many problems with both dc and ac sources that will help you practice circuit problems.

Note:**Exercise:**

Problem:

Check Your Understanding In an RLC circuit, $L = 5.0 \text{ mH}$, $C = 6.0 \mu\text{F}$, and $R = 200 \Omega$. (a) Is the circuit underdamped, critically damped, or overdamped? (b) If the circuit starts oscillating with a charge of $3.0 \times 10^{-3} \text{ C}$ on the capacitor, how much energy has been dissipated in the resistor by the time the oscillations cease?

Solution:

a. overdamped; b. 0.75 J

Summary

- The underdamped solution for the capacitor charge in an RLC circuit is
Equation:

$$q(t) = q_0 e^{-Rt/2L} \cos(\omega' t + \phi).$$

- The angular frequency given in the underdamped solution for the RLC circuit is
Equation:

$$\omega' = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}.$$

Key Equations

Mutual inductance by flux	$M = \frac{N_2\Phi_{21}}{I_1} = \frac{N_1\Phi_{12}}{I_2}$
Mutual inductance in circuits	$\varepsilon_1 = -M \frac{dI_2}{dt}$
Self-inductance in terms of magnetic flux	$N\Phi_m = LI$
Self-inductance in terms of emf	$\varepsilon = -L \frac{dI}{dt}$
Self-inductance of a solenoid	$L_{\text{solenoid}} = \frac{\mu_0 N^2 A}{l}$
Self-inductance of a toroid	$L_{\text{toroid}} = \frac{\mu_0 N^2 h}{2\pi} \ln \frac{R_2}{R_1}.$
Energy stored in an inductor	$U = \frac{1}{2} LI^2$
Current as a function of time for a RL circuit	$I(t) = \frac{\varepsilon}{R} (1 - e^{-t/\tau_L})$
Time constant for a RL circuit	$\tau_L = L/R$
Charge oscillation in LC circuits	$q(t) = q_0 \cos(\omega t + \phi)$
Angular frequency in LC circuits	$\omega = \frac{1}{\sqrt{LC}}$
Current oscillations in LC circuits	$i(t) = -\omega q_0 \sin(\omega t + \phi)$
Charge as a function of time in RLC circuit	$q(t) = q_0 e^{-Rt/2L} \cos(\omega' t + \phi)$
Angular frequency in RLC circuit	$\omega' = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}$

Conceptual Questions

Exercise:

Problem:

When a wire is connected between the two ends of a solenoid, the resulting circuit can oscillate like an RLC circuit. Describe what causes the capacitance in this circuit.

Exercise:

Problem:

Describe what effect the resistance of the connecting wires has on an oscillating LC circuit.

Solution:

This creates an RLC circuit that dissipates energy, causing oscillations to decrease in amplitude slowly or quickly depending on the value of resistance.

Exercise:

Problem:

Suppose you wanted to design an LC circuit with a frequency of 0.01 Hz. What problems might you encounter?

Exercise:

Problem:

A radio receiver uses an RLC circuit to pick out particular frequencies to listen to in your house or car without hearing other unwanted frequencies. How would someone design such a circuit?

Solution:

You would have to pick out a resistance that is small enough so that only one station at a time is picked up, but big enough so that the tuner

doesn't have to be set at exactly the correct frequency. The inductance or capacitance would have to be varied to tune into the station however practically speaking, variable capacitors are a lot easier to build in a circuit.

Problems

Exercise:

Problem:

In an oscillating RLC circuit, $R = 5.0\ \Omega$, $L = 5.0\ \text{mH}$, and $C = 500\ \mu\text{F}$. What is the angular frequency of the oscillations?

Exercise:

Problem:

In an oscillating RLC circuit with $L = 10\ \text{mH}$, $C = 1.5\ \mu\text{F}$, and $R = 2.0\ \Omega$, how much time elapses before the amplitude of the oscillations drops to half its initial value?

Solution:

6.9 ms

Exercise:

Problem:

What resistance R must be connected in series with a 200-mH inductor and a $10\ \mu\text{F}$ capacitor of the resulting RLC oscillating circuit is to decay to 50% of its initial value of charge in 50 cycles? To 0.10% of its initial value in 50 cycles?

Additional Problems

Exercise:

Problem:

Show that the self-inductance per unit length of an infinite, straight, thin wire is infinite.

Solution:

Let a equal the radius of the long, thin wire, r the location where the magnetic field is measured, and R the upper limit of the problem where we will take R as it approaches infinity.

$$\text{Outside, } B = \frac{\mu_0 I}{2\pi r} \quad \text{Inside, } B = \frac{\mu_0 I r}{2\pi a^2}$$

$$\text{proof} \quad U = \frac{\mu_0 I^2 l}{4\pi} \left(\frac{1}{4} + \ln \frac{R}{a} \right)$$

$$\text{So, } \frac{2U}{I^2} = \frac{\mu_0 l}{2\pi} \left(\frac{1}{4} + \ln \frac{R}{a} \right) \quad \text{and} \quad L = \infty$$

Exercise:**Problem:**

Two long, parallel wires carry equal currents in opposite directions. The radius of each wire is a , and the distance between the centers of the wires is d . Show that if the magnetic flux within the wires themselves can be ignored, the self-inductance of a length l of such a pair of wires is

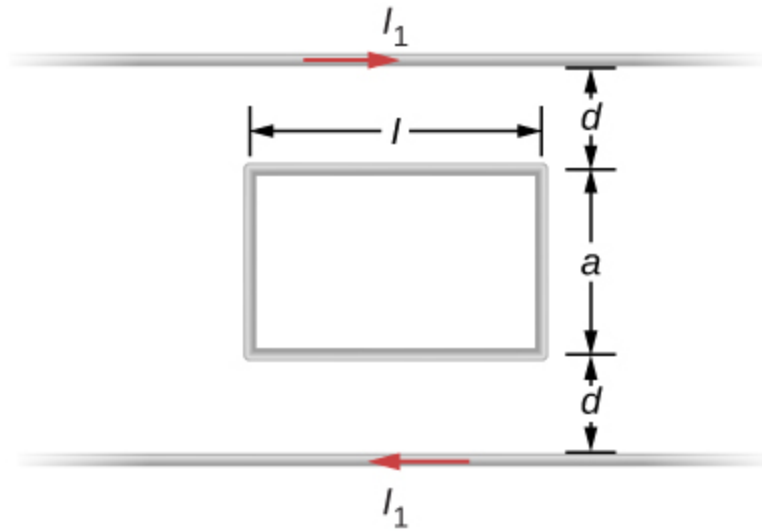
$$L = \frac{\mu_0 l}{\pi} \ln \frac{d-a}{a}.$$

(Hint: Calculate the magnetic flux through a rectangle of length l between the wires and then use $L = N\Phi/I$.)

Exercise:

Problem:

A small, rectangular single loop of wire with dimensions l , and a is placed, as shown below, in the plane of a much larger, rectangular single loop of wire. The two short sides of the larger loop are so far from the smaller loop that their magnetic fields over the smaller loop can be ignored. What is the mutual inductance of the two loops?

**Solution:**

$$M = \frac{\mu_0 l}{\pi} \ln \frac{d+a}{d}$$

Exercise:**Problem:**

Suppose that a cylindrical solenoid is wrapped around a core of iron whose magnetic susceptibility is x . Using [\[link\]](#), show that the self-inductance of the solenoid is given by

$$L = \frac{(1+x)\mu_0 N^2 A}{l},$$

where l is its length, A its cross-sectional area, and N its total number of turns.

Exercise:

Problem:

A solenoid with 4×10^7 turns/m has an iron core placed in it whose magnetic susceptibility is 4.0×10^3 . (a) If a current of 2.0 A flows through the solenoid, what is the magnetic field in the iron core? (b) What is the effective surface current formed by the aligned atomic current loops in the iron core? (c) What is the self-inductance of the filled solenoid?

Solution:

a. 100 T; b. 2 A; c. 0.50 H

Exercise:

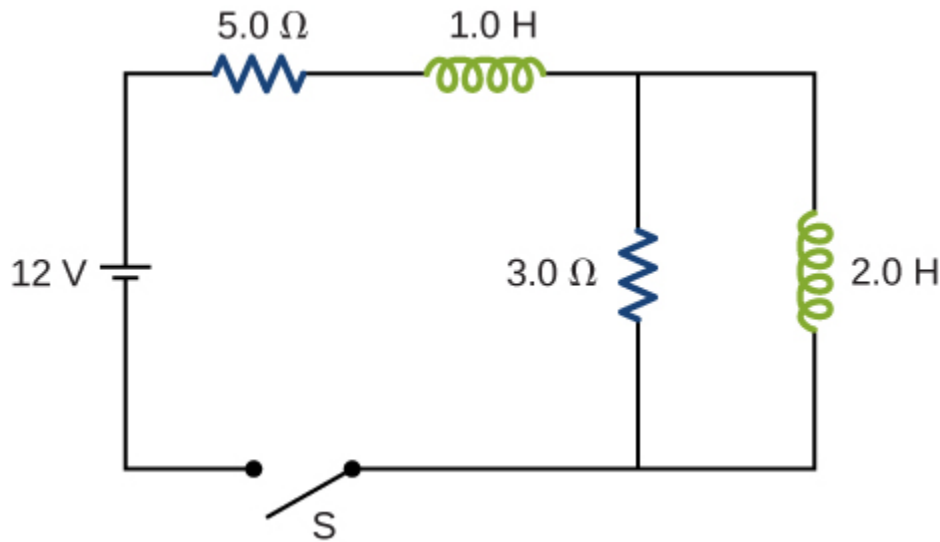
Problem:

A rectangular toroid with inner radius $R_1 = 7.0$ cm, outer radius $R_2 = 9.0$ cm, height $h = 3.0$, and $N = 3000$ turns is filled with an iron core of magnetic susceptibility 5.2×10^3 . (a) What is the self-inductance of the toroid? (b) If the current through the toroid is 2.0 A, what is the magnetic field at the center of the core? (c) For this same 2.0-A current, what is the effective surface current formed by the aligned atomic current loops in the iron core?

Exercise:

Problem:

The switch S of the circuit shown below is closed at $t = 0$. Determine (a) the initial current through the battery and (b) the steady-state current through the battery.



Solution:

a. 0 A; b. 2.4 A

Exercise:

Problem:

In an oscillating RLC circuit, $R = 7.0\ \Omega$, $L = 10\ \text{mH}$, and $C = 3.0\ \mu\text{F}$. Initially, the capacitor has a charge of $8.0\ \mu\text{C}$ and the current is zero. Calculate the charge on the capacitor (a) five cycles later and (b) 50 cycles later.

Exercise:

Problem:

A 25.0-H inductor has 100 A of current turned off in 1.00 ms. (a) What voltage is induced to oppose this? (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

a. $2.50 \times 10^6\ \text{V}$; (b) The voltage is so extremely high that arcing would occur and the current would not be reduced so rapidly. (c) It is

not reasonable to shut off such a large current in such a large inductor in such an extremely short time.

Challenge Problems

Exercise:

Problem:

A coaxial cable has an inner conductor of radius a , and outer thin cylindrical shell of radius b . A current I flows in the inner conductor and returns in the outer conductor. The self-inductance of the structure will depend on how the current in the inner cylinder tends to be distributed. Investigate the following two extreme cases. (a) Let current in the inner conductor be distributed only on the surface and find the self-inductance. (b) Let current in the inner cylinder be distributed uniformly over its cross-section and find the self-inductance. Compare with your results in (a).

Exercise:

Problem:

In a damped oscillating circuit the energy is dissipated in the resistor. The Q -factor is a measure of the persistence of the oscillator against the dissipative loss. (a) Prove that for a lightly damped circuit the energy, U , in the circuit decreases according to the following equation.

$$\frac{dU}{dt} = -2\beta U, \text{ where } \beta = \frac{R}{2L}.$$

(b) Using the definition of the Q -factor as energy divided by the loss over the next cycle, prove that Q -factor of a lightly damped oscillator as defined in this problem is

$$Q \equiv \frac{U_{\text{begin}}}{\Delta U_{\text{one cycle}}} = \frac{1}{2\pi R} \frac{\overline{L}}{C}.$$

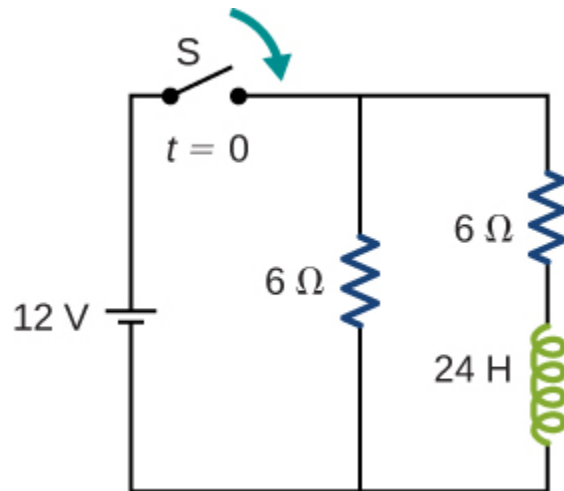
(Hint: For (b), to obtain Q , divide E at the beginning of one cycle by the change ΔE over the next cycle.)

Solution:

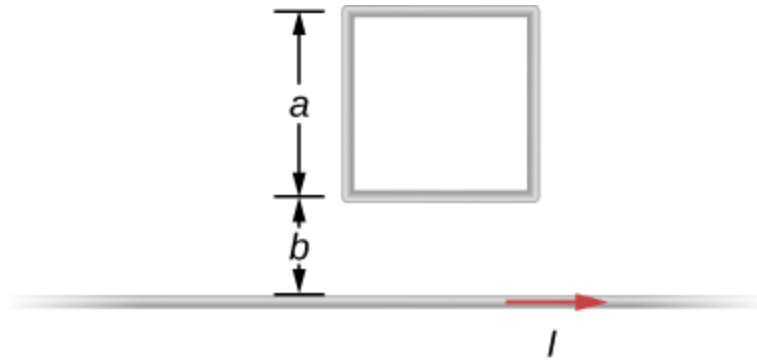
proof

Exercise:**Problem:**

The switch in the circuit shown below is closed at $t = 0$ s. Find currents through (a) R_1 , (b) R_2 , and (c) the battery as function of time.

**Exercise:****Problem:**

A square loop of side 2 cm is placed 1 cm from a long wire carrying a current that varies with time at a constant rate of 3 A/s as shown below. (a) Use Ampère's law and find the magnetic field. (b) Determine the magnetic flux through the loop. (c) If the loop has a resistance of $3\ \Omega$, how much induced current flows in the loop?



Solution:

a. $\frac{dB}{dt} = 6 \times 10^{-6} \text{ T/s}$; b. $\Phi = \frac{\mu_0 a I}{2\pi} \ln \left(\frac{a+b}{b} \right)$; c. 4.4 nA

Glossary

RLC circuit

circuit with an ac source, resistor, inductor, and capacitor all in series.

Introduction

class="introduction"

The current we draw into our houses is an alternating current (ac). Power lines transmit ac to our neighborhoods, where local power stations and transformers distribute it to our homes. In this chapter, we discuss how a transformer works and how it allows us to transmit power at very high voltages and minimal heating losses across the lines.



Electric power is delivered to our homes by alternating current (ac) through high-voltage transmission lines. As explained in [Transformers](#), transformers can then change the amplitude of the alternating potential difference to a more useful form. This lets us transmit power at very high voltages, minimizing resistive heating losses in the lines, and then furnish that power to homes at lower, safer voltages. Because constant potential differences are unaffected by transformers, this capability is more difficult to achieve with direct-current transmission.

In this chapter, we use Kirchhoff's laws to analyze four simple circuits in which ac flows. We have discussed the use of the resistor, capacitor, and inductor in circuits with batteries. These components are also part of ac circuits. However, because ac is required, the constant source of emf supplied by a battery is replaced by an ac voltage source, which produces an oscillating emf.

AC Sources

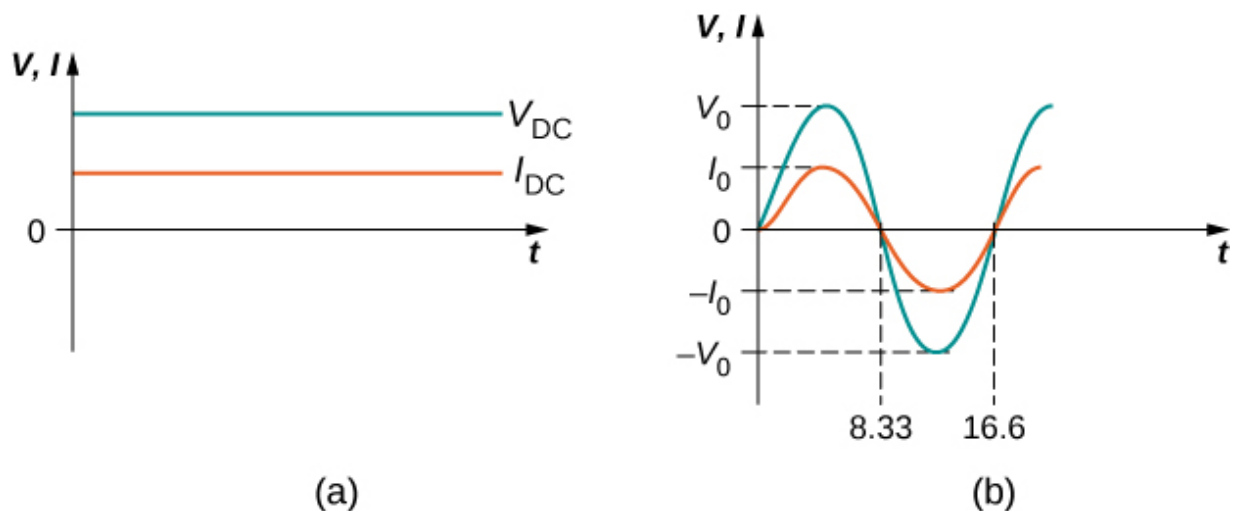
By the end of the section, you will be able to:

- Explain the differences between direct current (dc) and alternating current (ac)
- Define characteristic features of alternating current and voltage, such as the amplitude or peak and the frequency

Most examples dealt with so far in this book, particularly those using batteries, have constant-voltage sources. Thus, once the current is established, it is constant. **Direct current (dc)** is the flow of electric charge in only one direction. It is the steady state of a constant-voltage circuit.

Most well-known applications, however, use a time-varying voltage source. **Alternating current (ac)** is the flow of electric charge that periodically reverses direction. An ac is produced by an alternating emf, which is generated in a power plant, as described in [Induced Electric Fields](#). If the ac source varies periodically, particularly sinusoidally, the circuit is known as an ac circuit. Examples include the commercial and residential power that serves so many of our needs.

The ac voltages and frequencies commonly used in businesses and homes vary around the world. In a typical house, the potential difference between the two sides of an electrical outlet alternates sinusoidally with a frequency of 60 or 50 Hz and an amplitude of 170 or 311 V, depending on whether you live in the United States or Europe, respectively. Most people know the potential difference for electrical outlets is 120 V or 220 V in the US or Europe, but as explained later in the chapter, these voltages are not the peak values given here but rather are related to the common voltages we see in our electrical outlets. [\[link\]](#) shows graphs of voltage and current versus time for typical dc and ac power in the United States.



(a) The dc voltage and current are constant in time, once the current is established. (b) The voltage and current versus time are quite different for ac power. In this example, which shows 60-Hz ac power and time t in milliseconds, voltage and current are sinusoidal and are in phase for a simple resistance circuit. The frequencies and peak voltages of ac sources differ greatly.

Suppose we hook up a resistor to an ac voltage source and determine how the voltage and current vary in time across the resistor. [\[link\]](#) shows a schematic of a simple circuit with an ac voltage source. The voltage fluctuates sinusoidally with time at a fixed frequency, as shown, on either the battery terminals or the resistor. Therefore, the **ac voltage**, or the “voltage at a plug,” can be given by

Note:

Equation:

$$v = V_0 \sin \omega t,$$

where v is the voltage at time t , V_0 is the peak voltage, and ω is the angular frequency in radians per second. For a typical house in the United States, $V_0 = 170\text{ V}$ and $\omega = 120\pi\text{ rad/s}$, whereas in Europe, $V_0 = 311\text{ V}$ and $\omega = 100\pi\text{ rad/s}$.

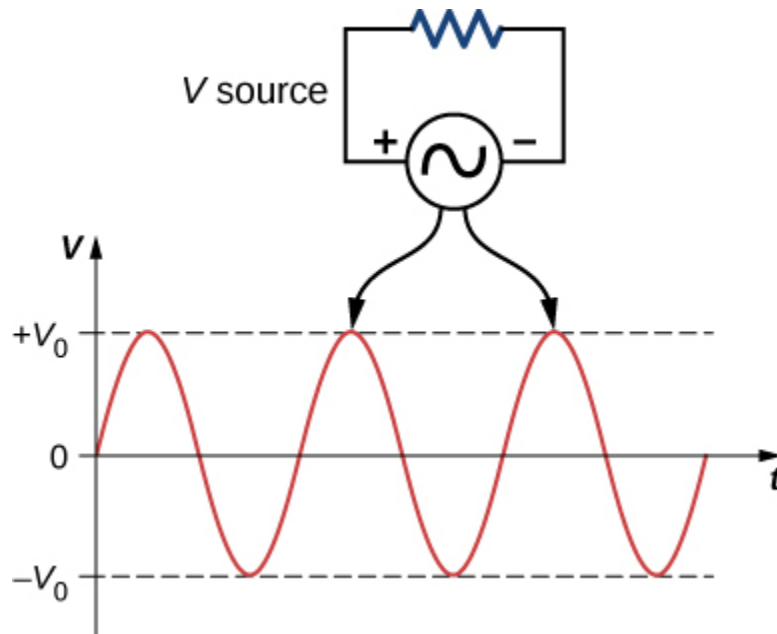
For this simple resistance circuit, $I = V/R$, so the **ac current**, meaning the current that fluctuates sinusoidally with time at a fixed frequency, is

Note:

Equation:

$$i = I_0 \sin \omega t,$$

where i is the current at time t and I_0 is the peak current and is equal to V_0/R . For this example, the voltage and current are said to be in phase, meaning that their sinusoidal functional forms have peaks, troughs, and nodes in the same place. They oscillate in sync with each other, as shown in [\[link\]](#)(b). In these equations, and throughout this chapter, we use lowercase letters (such as i) to indicate instantaneous values and capital letters (such as I) to indicate maximum, or peak, values.



The potential difference V between the terminals of an ac voltage source fluctuates, so the source and the resistor have ac sine waves on top of each other. The mathematical expression for v is given by $v = V_0 \sin \omega t$.

Current in the resistor alternates back and forth just like the driving voltage, since $I = V/R$. If the resistor is a fluorescent light bulb, for example, it brightens and dims 120 times per second as the current repeatedly goes through zero. A 120-Hz flicker is too rapid for your eyes to detect, but if you wave your hand back and forth between your face and a fluorescent light, you will see the stroboscopic effect of ac.

Note:

Exercise:

Problem:

Check Your Understanding If a European ac voltage source is considered, what is the time difference between the zero crossings on an ac voltage-versus-time graph?

Solution:

10 ms

Summary

- Direct current (dc) refers to systems in which the source voltage is constant.
- Alternating current (ac) refers to systems in which the source voltage varies periodically, particularly sinusoidally.
- The voltage source of an ac system puts out a voltage that is calculated from the time, the peak voltage, and the angular frequency.
- In a simple circuit, the current is found by dividing the voltage by the resistance. An ac current is calculated using the peak current (determined by dividing the peak voltage by the resistance), the angular frequency, and the time.

Conceptual Questions

Exercise:**Problem:**

What is the relationship between frequency and angular frequency?

Solution:

Angular frequency is 2π times frequency.

Problems

Exercise:

Problem:

Write an expression for the output voltage of an ac source that has an amplitude of 12 V and a frequency of 200 Hz.

Glossary

ac current

current that fluctuates sinusoidally with time at a fixed frequency

ac voltage

voltage that fluctuates sinusoidally with time at a fixed frequency

alternating current (ac)

flow of electric charge that periodically reverses direction

direct current (dc)

flow of electric charge in only one direction

Simple AC Circuits

By the end of the section, you will be able to:

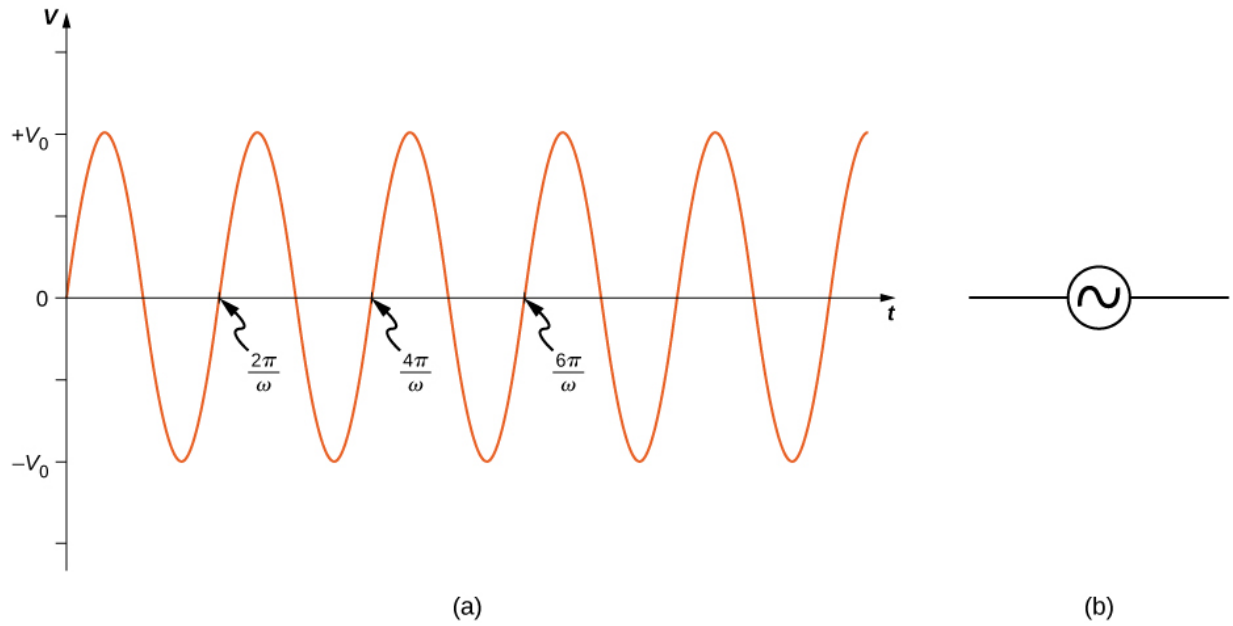
- Interpret phasor diagrams and apply them to ac circuits with resistors, capacitors, and inductors
- Define the reactance for a resistor, capacitor, and inductor to help understand how current in the circuit behaves compared to each of these devices

In this section, we study simple models of ac voltage sources connected to three circuit components: (1) a resistor, (2) a capacitor, and (3) an inductor. The power furnished by an ac voltage source has an emf given by

Equation:

$$v(t) = V_0 \sin \omega t,$$

as shown in [\[link\]](#). This sine function assumes we start recording the voltage when it is $v = 0$ V at a time of $t = 0$ s. A phase constant may be involved that shifts the function when we start measuring voltages, similar to the phase constant in the waves we studied in [Waves](#). However, because we are free to choose when we start examining the voltage, we can ignore this phase constant for now. We can measure this voltage across the circuit components using one of two methods: (1) a quantitative approach based on our knowledge of circuits, or (2) a graphical approach that is explained in the coming sections.



(a) The output $v(t) = V_0 \sin \omega t$ of an ac generator. (b) Symbol used to represent an ac voltage source in a circuit diagram.

Resistor

First, consider a resistor connected across an ac voltage source. From Kirchhoff's loop rule, the instantaneous voltage across the resistor of [\[link\]](#) (a) is

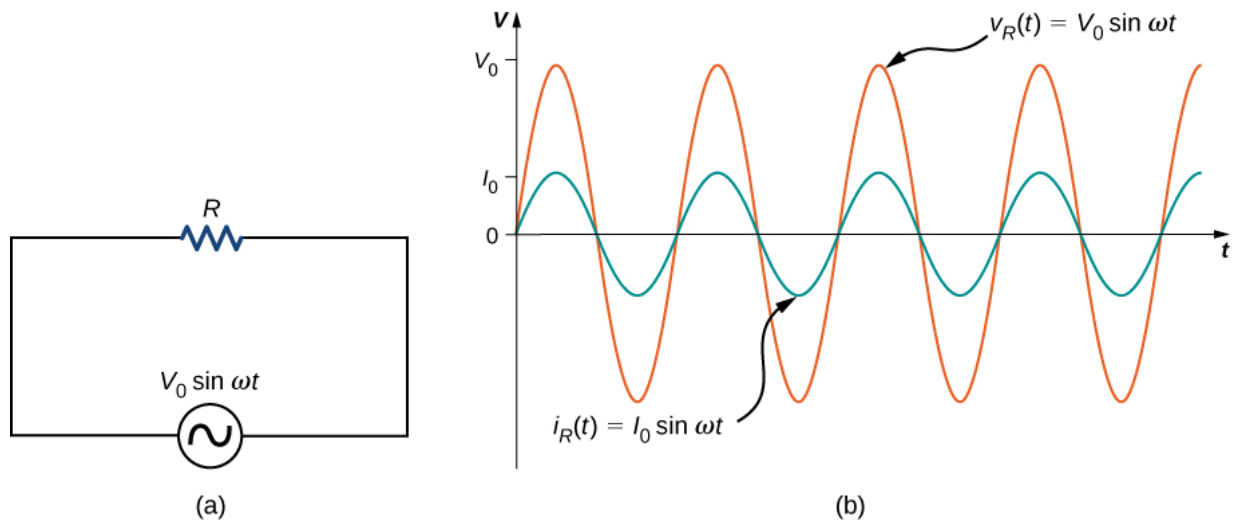
Equation:

$$v_R(t) = V_0 \sin \omega t$$

and the instantaneous current through the resistor is

Equation:

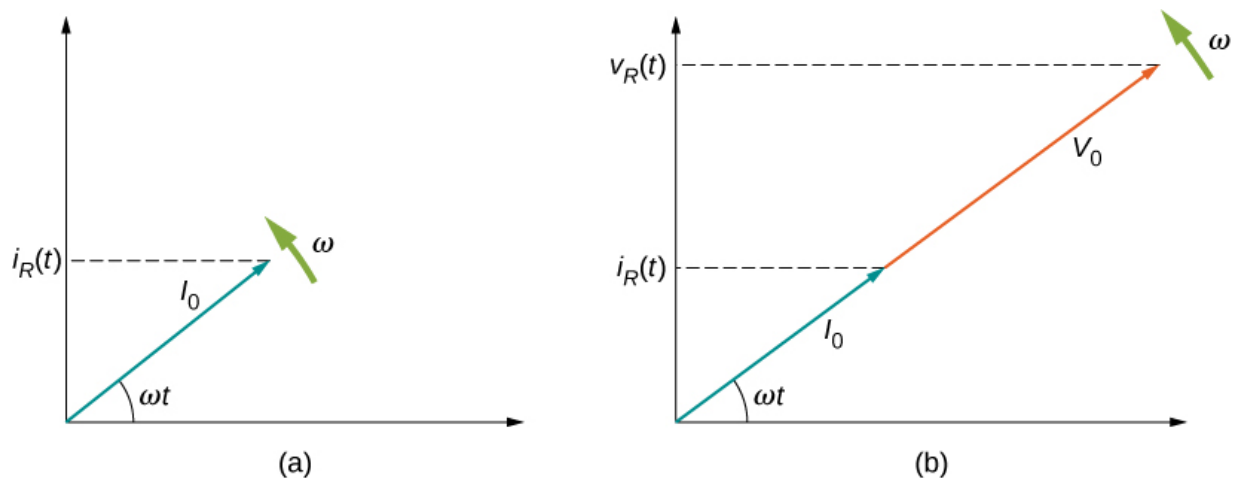
$$i_R(t) = \frac{v_R(t)}{R} = \frac{V_0}{R} \sin \omega t = I_0 \sin \omega t.$$



(a) A resistor connected across an ac voltage source. (b) The current $i_R(t)$ through the resistor and the voltage $v_R(t)$ across the resistor. The two quantities are in phase.

Here, $I_0 = V_0/R$ is the amplitude of the time-varying current. Plots of $i_R(t)$ and $v_R(t)$ are shown in [link](#)(b). Both curves reach their maxima and minima at the same times, that is, the current through and the voltage across the resistor are in phase.

Graphical representations of the phase relationships between current and voltage are often useful in the analysis of ac circuits. Such representations are called *phasor diagrams*. The phasor diagram for $i_R(t)$ is shown in [link](#)(a), with the current on the vertical axis. The arrow (or phasor) is rotating counterclockwise at a constant angular frequency ω , so we are viewing it at one instant in time. If the length of the arrow corresponds to the current amplitude I_0 , the projection of the rotating arrow onto the vertical axis is $i_R(t) = I_0 \sin \omega t$, which is the instantaneous current.



(a) The phasor diagram representing the current through the resistor of [\[link\]](#). (b) The phasor diagram representing both $i_R(t)$ and $v_R(t)$.

The vertical axis on a phasor diagram could be either the voltage or the current, depending on the phasor that is being examined. In addition, several quantities can be depicted on the same phasor diagram. For example, both the current $i_R(t)$ and the voltage $v_R(t)$ are shown in the diagram of [\[link\]](#)(b). Since they have the same frequency and are in phase, their phasors point in the same direction and rotate together. The relative lengths of the two phasors are arbitrary because they represent different quantities; however, the ratio of the lengths of the two phasors can be represented by the resistance, since one is a voltage phasor and the other is a current phasor.

Capacitor

Now let's consider a capacitor connected across an ac voltage source. From Kirchhoff's loop rule, the instantaneous voltage across the capacitor of [\[link\]](#)(a) is

Equation:

$$v_C(t) = V_0 \sin \omega t.$$

Recall that the charge in a capacitor is given by $Q = CV$. This is true at any time measured in the ac cycle of voltage. Consequently, the instantaneous charge on the capacitor is

Equation:

$$q(t) = Cv_C(t) = CV_0 \sin \omega t.$$

Since the current in the circuit is the rate at which charge enters (or leaves) the capacitor,

Equation:

$$i_C(t) = \frac{dq(t)}{dt} = \omega CV_0 \cos \omega t = I_0 \cos \omega t,$$

where $I_0 = \omega CV_0$ is the current amplitude. Using the trigonometric relationship $\cos \omega t = \sin (\omega t + \pi/2)$, we may express the instantaneous current as

Equation:

$$i_C(t) = I_0 \sin \left(\omega t + \frac{\pi}{2} \right).$$

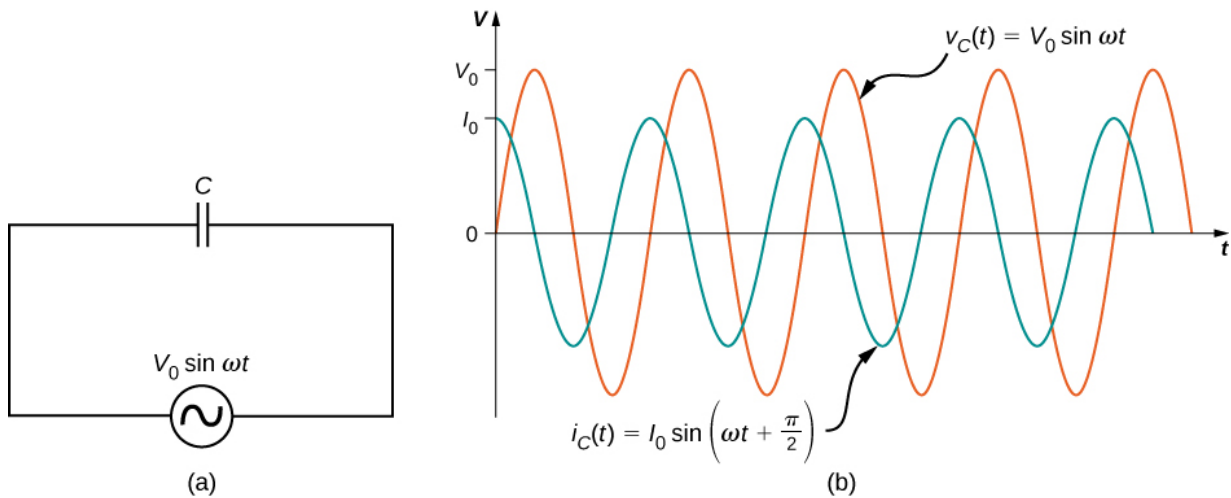
Dividing V_0 by I_0 , we obtain an equation that looks similar to Ohm's law:

Note:

Equation:

$$\frac{V_0}{I_0} = \frac{1}{\omega C} = X_C.$$

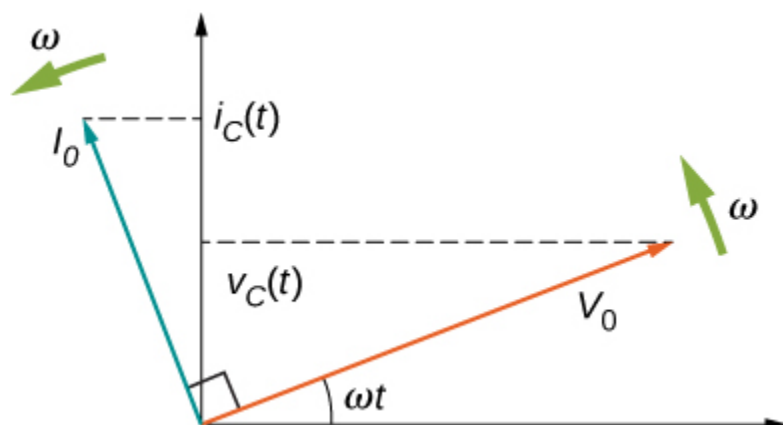
The quantity X_C is analogous to resistance in a dc circuit in the sense that both quantities are a ratio of a voltage to a current. As a result, they have the same unit, the ohm. Keep in mind, however, that a capacitor stores and discharges electric energy, whereas a resistor dissipates it. The quantity X_C is known as the **capacitive reactance** of the capacitor, or the opposition of a capacitor to a change in current. It depends inversely on the frequency of the ac source—high frequency leads to low capacitive reactance.



(a) A capacitor connected across an ac generator. (b) The current $i_C(t)$ through the capacitor and the voltage $v_C(t)$ across the capacitor. Notice that $i_C(t)$ leads $v_C(t)$ by $\pi/2$ rad.

A comparison of the expressions for $v_C(t)$ and $i_C(t)$ shows that there is a phase difference of $\pi/2$ rad between them. When these two quantities are plotted together, the current peaks a quarter cycle (or $\pi/2$ rad) ahead of the voltage, as illustrated in [\[link\]](#)(b). The current through a capacitor leads the voltage across a capacitor by $\pi/2$ rad, or a quarter of a cycle.

The corresponding phasor diagram is shown in [\[link\]](#). Here, the relationship between $i_C(t)$ and $v_C(t)$ is represented by having their phasors rotate at the same angular frequency, with the current phasor leading by $\pi/2$ rad.



The phasor diagram for the capacitor of [\[link\]](#). The current phasor leads the voltage phasor by $\pi/2$ rad as they both rotate with the same angular frequency.

To this point, we have exclusively been using peak values of the current or voltage in our discussion, namely, I_0 and V_0 . However, if we average out the values of current or voltage, these values are zero. Therefore, we often use a second convention called the root mean square value, or rms value, in discussions of current and voltage. The rms operates in reverse of the terminology. First, you square the function, next, you take the mean, and then, you find the square root. As a result, the rms values of current and voltage are not zero. Appliances and devices are commonly quoted with rms values for their operations, rather than peak values. We indicate rms values with a subscript attached to a capital letter (such as I_{rms}).

Although a capacitor is basically an open circuit, an **rms current**, or the root mean square of the current, appears in a circuit with an ac voltage applied to a capacitor. Consider that

Note:

Equation:

$$I_{\text{rms}} = \frac{I_0}{\sqrt{2}},$$

where I_0 is the peak current in an ac system. The **rms voltage**, or the root mean square of the voltage, is

Note:

Equation:

$$V_{\text{rms}} = \frac{V_0}{\sqrt{2}},$$

where V_0 is the peak voltage in an ac system. The rms current appears because the voltage is continually reversing, charging, and discharging the capacitor. If the frequency goes to zero, which would be a dc voltage, X_C tends to infinity, and the current is zero once the capacitor is charged. At very high frequencies, the capacitor's reactance tends to zero—it has a negligible reactance and does not impede the current (it acts like a simple wire).

Inductor

Lastly, let's consider an inductor connected to an ac voltage source. From Kirchhoff's loop rule, the voltage across the inductor L of [\[link\]](#)(a) is

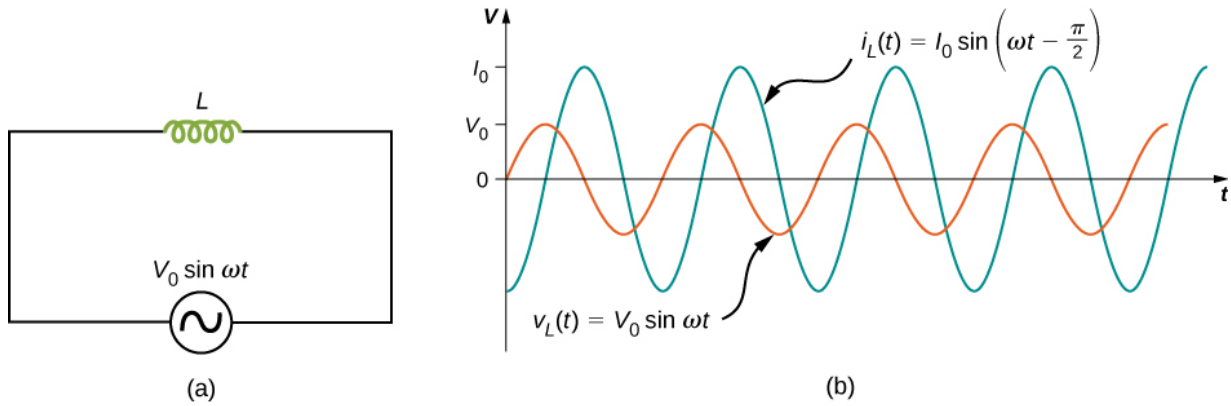
Equation:

$$v_L(t) = V_0 \sin \omega t.$$

The emf across an inductor is equal to $\varepsilon = -L (di_L/dt)$; however, the potential difference across the inductor is $v_L(t) = L di_L(t)/dt$, because if we consider that the voltage around the loop must equal zero, the voltage gained from the ac source must dissipate through the inductor. Therefore, connecting this with the ac voltage source, we have

Equation:

$$\frac{di_L(t)}{dt} = \frac{V_0}{L} \sin \omega t.$$



(a) An inductor connected across an ac generator. (b) The current $i_L(t)$ through the inductor and the voltage $v_L(t)$ across the inductor. Here $i_L(t)$ lags $v_L(t)$ by $\pi/2$ rad.

The current $i_L(t)$ is found by integrating this equation. Since the circuit does not contain a source of constant emf, there is no steady current in the circuit. Hence, we can set the constant of integration, which represents the steady current in the circuit, equal to zero, and we have

Equation:

$$i_L(t) = -\frac{V_0}{\omega L} \cos \omega t = \frac{V_0}{\omega L} \sin \left(\omega t - \frac{\pi}{2} \right) = I_0 \sin \left(\omega t - \frac{\pi}{2} \right),$$

where $I_0 = V_0/\omega L$. The relationship between V_0 and I_0 may also be written in a form analogous to Ohm's law:

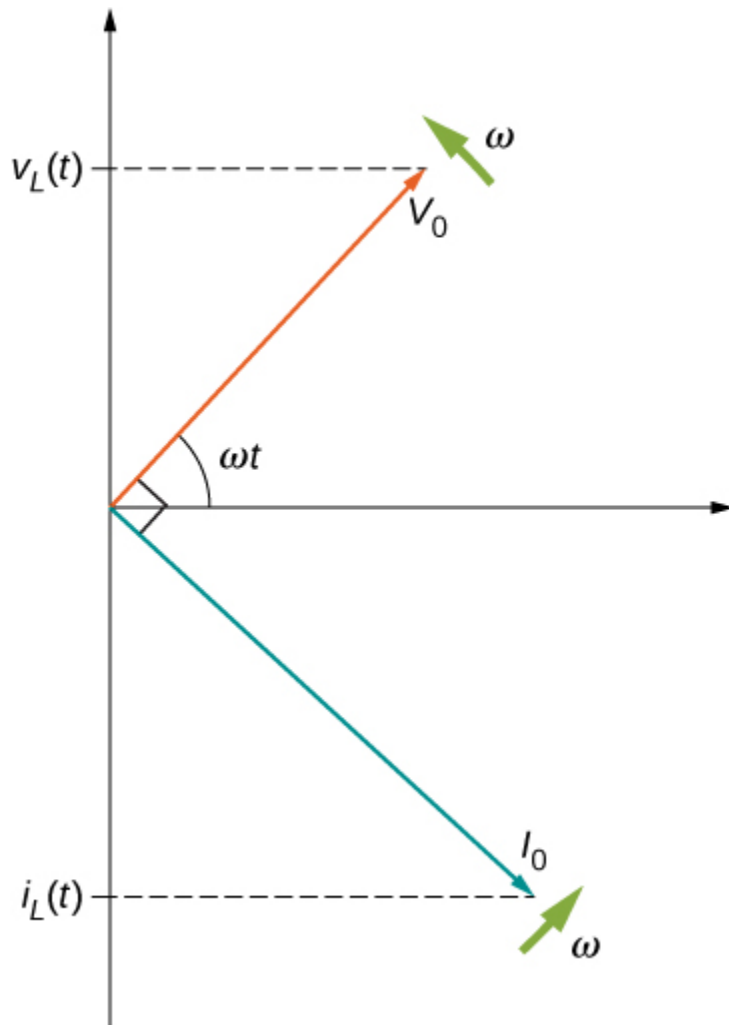
Note:

Equation:

$$\frac{V_0}{I_0} = \omega L = X_L.$$

The quantity X_L is known as the **inductive reactance** of the inductor, or the opposition of an inductor to a change in current; its unit is also the ohm. Note that X_L varies directly as the frequency of the ac source—high frequency causes high inductive reactance.

A phase difference of $\pi/2$ rad occurs between the current through and the voltage across the inductor. From [\[link\]](#) and [\[link\]](#), the current through an inductor lags the potential difference across an inductor by $\pi/2$ rad, or a quarter of a cycle. The phasor diagram for this case is shown in [\[link\]](#).



The phasor diagram for the inductor of [\[link\]](#). The current phasor lags the voltage phasor by $\pi/2$ rad as they both rotate with the same angular frequency.

Note:

An animation from the University of New South Wales [AC Circuits](#) illustrates some of the concepts we discuss in this chapter. They also include wave and phasor diagrams that evolve over time so that you can get a better picture of how each changes over time.

Example:**Simple AC Circuits**

An ac generator produces an emf of amplitude 10 V at a frequency $f = 60$ Hz. Determine the voltages across and the currents through the circuit elements when the generator is connected to (a) a $100\text{ }\Omega$ resistor, (b) a $10\text{ }\mu\text{F}$ capacitor, and (c) a 15-mH inductor.

Strategy

The entire AC voltage across each device is the same as the source voltage. We can find the currents by finding the reactance X of each device and solving for the peak current using $I_0 = V_0/X$.

Solution

The voltage across the terminals of the source is

Equation:

$$v(t) = V_0 \sin \omega t = (10\text{ V}) \sin 120\pi t,$$

where $\omega = 2\pi f = 120\pi$ rad/s is the angular frequency. Since $v(t)$ is also the voltage across each of the elements, we have

Equation:

$$v(t) = v_R(t) = v_C(t) = v_L(t) = (10\text{ V}) \sin 120\pi t.$$

a. When $R = 100\text{ }\Omega$, the amplitude of the current through the resistor is

Equation:

$$I_0 = V_0/R = 10\text{ V}/100\text{ }\Omega = 0.10\text{ A},$$

so

Equation:

$$i_R(t) = (0.10\text{ A}) \sin 120\pi t.$$

b. From [\[link\]](#), the capacitive reactance is

Equation:

$$X_C = \frac{1}{\omega C} = \frac{1}{(120\pi\text{ rad/s})(10 \times 10^{-6}\text{ F})} = 265\text{ }\Omega,$$

so the maximum value of the current is

Equation:

$$I_0 = \frac{V_0}{X_C} = \frac{10 \text{ V}}{265 \Omega} = 3.8 \times 10^{-2} \text{ A}$$

and the instantaneous current is given by

Equation:

$$i_C(t) = (3.8 \times 10^{-2} \text{ A}) \sin \left(120\pi t + \frac{\pi}{2} \right).$$

c. From [\[link\]](#), the inductive reactance is

Equation:

$$X_L = \omega L = (120\pi \text{ rad/s})(15 \times 10^{-3} \text{ H}) = 5.7 \Omega.$$

The maximum current is therefore

Equation:

$$I_0 = \frac{10 \text{ V}}{5.7 \Omega} = 1.8 \text{ A}$$

and the instantaneous current is

Equation:

$$i_L(t) = (1.8 \text{ A}) \sin \left(120\pi t - \frac{\pi}{2} \right).$$

Significance

Although the voltage across each device is the same, the peak current has different values, depending on the reactance. The reactance for each device depends on the values of resistance, capacitance, or inductance.

Note:

Exercise:

Problem:

Check Your Understanding Repeat [\[link\]](#) for an ac source of amplitude 20 V and frequency 100 Hz.

Solution:

a. $(20 \text{ V}) \sin 200\pi t$, $(0.20 \text{ A}) \sin 200\pi t$; b. $(20 \text{ V}) \sin 200\pi t$, $(0.13 \text{ A}) \sin (200\pi t + \pi/2)$; c. $(20 \text{ V}) \sin 200\pi t$, $(2.1 \text{ A}) \sin (200\pi t - \pi/2)$

Summary

- For resistors, the current through and the voltage across are in phase.
- For capacitors, we find that when a sinusoidal voltage is applied to a capacitor, the voltage follows the current by one-fourth of a cycle. Since a capacitor can stop current when fully charged, it limits current and offers another form of ac resistance, called capacitive reactance, which has units of ohms.
- For inductors in ac circuits, we find that when a sinusoidal voltage is applied to an inductor, the voltage leads the current by one-fourth of a cycle.
- The opposition of an inductor to a change in current is expressed as a type of ac reactance. This inductive reactance, which has units of ohms, varies with the frequency of the ac source.

Conceptual Questions

Exercise:**Problem:**

Explain why at high frequencies a capacitor acts as an ac short, whereas an inductor acts as an open circuit.

Problems

Exercise:

Problem:

Calculate the reactance of a $5.0\text{-}\mu\text{F}$ capacitor at (a) 60 Hz, (b) 600 Hz, and (c) 6000 Hz.

Solution:

a. $530\ \Omega$; b. $53\ \Omega$; c. $5.3\ \Omega$

Exercise:

Problem:

What is the capacitance of a capacitor whose reactance is $10\ \Omega$ at 60 Hz?

Exercise:

Problem:

Calculate the reactance of a 5.0-mH inductor at (a) 60 Hz, (b) 600 Hz, and (c) 6000 Hz.

Solution:

a. $1.9\ \Omega$; b. $19\ \Omega$; c. $190\ \Omega$

Exercise:

Problem:

What is the self-inductance of a coil whose reactance is $10\ \Omega$ at 60 Hz?

Exercise:

Problem:

At what frequency is the reactance of a $20\text{-}\mu\text{F}$ capacitor equal to that of a 10-mH inductor?

Solution:

360 Hz

Exercise:**Problem:**

At 1000 Hz, the reactance of a 5.0-mH inductor is equal to the reactance of a particular capacitor. What is the capacitance of the capacitor?

Exercise:**Problem:**

A $50\text{-}\Omega$ resistor is connected across the emf $v(t) = (160\text{ V}) \sin(120\pi t)$. Write an expression for the current through the resistor.

Solution:

$$i(t) = (3.2\text{ A}) \sin(120\pi t)$$

Exercise:**Problem:**

A $25\text{-}\mu\text{F}$ capacitor is connected to an emf given by $v(t) = (160\text{ V}) \sin(120\pi t)$. (a) What is the reactance of the capacitor? (b) Write an expression for the current output of the source.

Exercise:**Problem:**

A 100-mH inductor is connected across the emf of the preceding problem. (a) What is the reactance of the inductor? (b) Write an expression for the current through the inductor.

Solution:

a. $38\ \Omega$; b. $i(t) = (4.24\text{A}) \sin(120\pi t - \pi/2)$

Glossary

capacitive reactance

opposition of a capacitor to a change in current

inductive reactance

opposition of an inductor to a change in current

rms current

root mean square of the current

rms voltage

root mean square of the voltage

RLC Series Circuits with AC

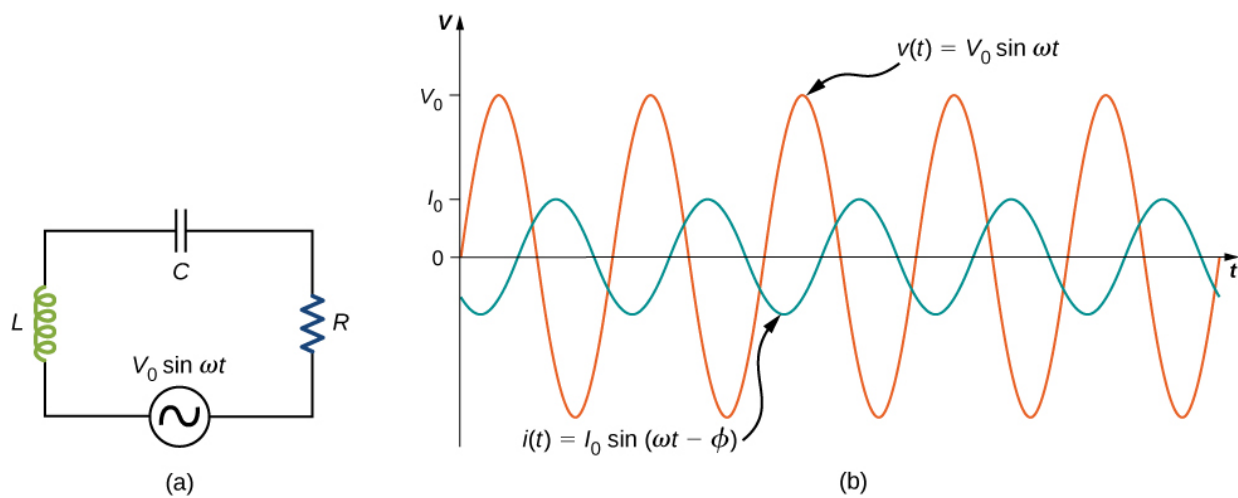
By the end of the section, you will be able to:

- Describe how the current varies in a resistor, a capacitor, and an inductor while in series with an ac power source
- Use phasors to understand the phase angle of a resistor, capacitor, and inductor ac circuit and to understand what that phase angle means
- Calculate the impedance of a circuit

The ac circuit shown in [\[link\]](#), called an *RLC* series circuit, is a series combination of a resistor, capacitor, and inductor connected across an ac source. It produces an emf of

Equation:

$$v(t) = V_0 \sin \omega t.$$



(a) An *RLC* series circuit. (b) A comparison of the generator output voltage and the current. The value of the phase difference ϕ depends on the values of R , C , and L .

Since the elements are in series, the same current flows through each element at all points in time. The relative phase between the current and the emf is not obvious when all three elements are present.

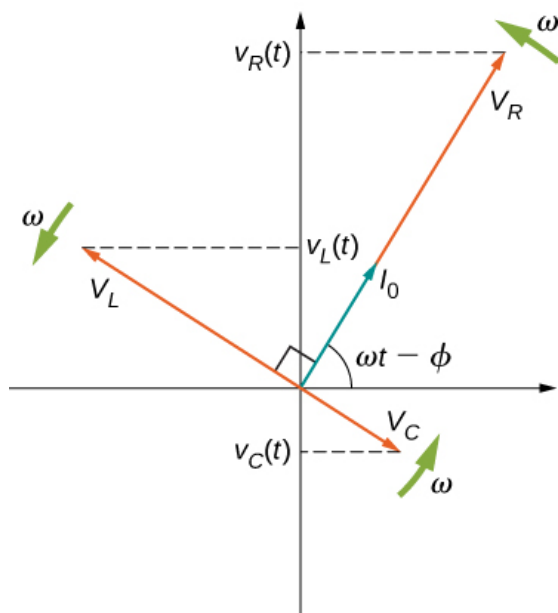
Consequently, we represent the current by the general expression

Equation:

$$i(t) = I_0 \sin(\omega t - \phi),$$

where I_0 is the current amplitude and ϕ is the **phase angle** between the current and the applied voltage. The phase angle is thus the amount by which the voltage and current are out of phase with each other in a circuit. Our task is to find I_0 and ϕ .

A phasor diagram involving $i(t)$, $v_R(t)$, $v_C(t)$, and $v_L(t)$ is helpful for analyzing the circuit. As shown in [\[link\]](#), the phasor representing $v_R(t)$ points in the same direction as the phasor for $i(t)$; its amplitude is $V_R = I_0 R$. The $v_C(t)$ phasor lags the $i(t)$ phasor by $\pi/2$ rad and has the amplitude $V_C = I_0 X_C$. The phasor for $v_L(t)$ leads the $i(t)$ phasor by $\pi/2$ rad and has the amplitude $V_L = I_0 X_L$.



The phasor diagram for the *RLC* series circuit of [\[link\]](#).

At any instant, the voltage across the *RLC* combination is $v_R(t) + v_L(t) + v_C(t) = v(t)$, the emf of the source. Since a component of a sum of vectors is the sum of the components of the individual vectors—for example, $(A + B)_y = A_y + B_y$ —the projection of the vector sum of phasors onto the vertical axis is the sum of the vertical projections of the individual phasors. Hence, if we add vectorially the phasors representing $v_R(t)$, $v_L(t)$, and $v_C(t)$ and then find the projection of the resultant onto the vertical axis, we obtain

Equation:

$$v_R(t) + v_L(t) + v_C(t) = v(t) = V_0 \sin \omega t.$$

The vector sum of the phasors is shown in [\[link\]](#). The resultant phasor has an amplitude V_0 and is directed at an angle ϕ with respect to the $v_R(t)$, or $i(t)$, phasor. The projection of this resultant phasor onto the vertical axis is $v(t) = V_0 \sin \omega t$. We can easily determine the unknown quantities I_0 and ϕ from the geometry of the phasor diagram. For the phase angle,

Equation:

$$\phi = \tan^{-1} \frac{V_L - V_C}{V_R} = \tan^{-1} \frac{I_0 X_L - I_0 X_C}{I_0 R},$$

and after cancellation of I_0 , this becomes

Note:

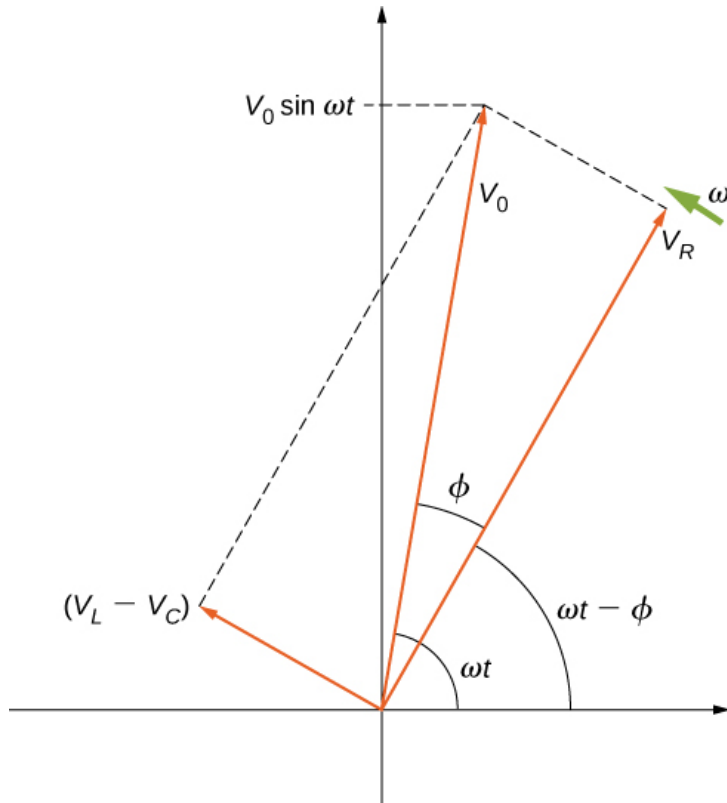
Equation:

$$\phi = \tan^{-1} \frac{X_L - X_C}{R}.$$

Furthermore, from the Pythagorean theorem,

Equation:

$$V_0 = \sqrt{V_R^2 + (V_L - V_C)^2} = \sqrt{(I_0 R)^2 + (I_0 X_L - I_0 X_C)^2} = I_0 \sqrt{R^2 + (X_L - X_C)^2}.$$



The resultant of the phasors for $v_L(t)$, $v_C(t)$, and $v_R(t)$ is equal to the phasor for $v(t) = V_0 \sin \omega t$. The $i(t)$ phasor (not shown) is aligned with the $v_R(t)$ phasor.

The current amplitude is therefore the ac version of Ohm's law:

Note:

Equation:

$$I_0 = \frac{V_0}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{V_0}{Z},$$

where

Note:

Equation:

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

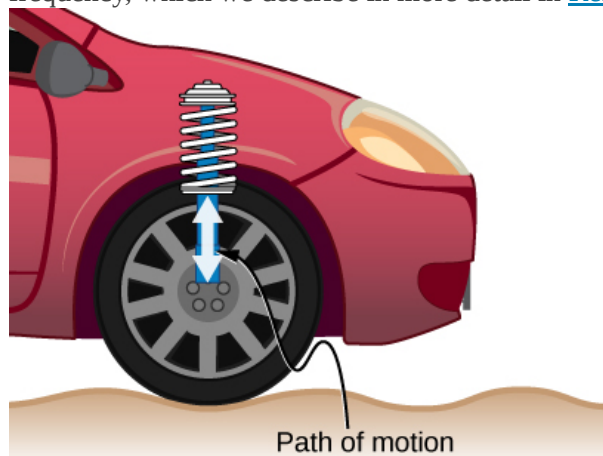
is known as the **impedance** of the circuit. Its unit is the ohm, and it is the ac analog to resistance in a dc circuit, which measures the combined effect of resistance, capacitive reactance, and inductive reactance ([link](#)).



Power capacitors are used to balance the impedance of

the effective inductance in transmission lines.

The RLC circuit is analogous to the wheel of a car driven over a corrugated road ([link](#)). The regularly spaced bumps in the road drive the wheel up and down; in the same way, a voltage source increases and decreases. The shock absorber acts like the resistance of the RLC circuit, damping and limiting the amplitude of the oscillation. Energy within the wheel system goes back and forth between kinetic and potential energy stored in the car spring, analogous to the shift between a maximum current, with energy stored in an inductor, and no current, with energy stored in the electric field of a capacitor. The amplitude of the wheel's motion is at a maximum if the bumps in the road are hit at the resonant frequency, which we describe in more detail in [Resonance in an AC Circuit](#).



On a car, the shock absorber damps motion and dissipates energy. This is much like the resistance in an RLC circuit. The mass and spring determine the resonant frequency.

Note:

AC Circuits

To analyze an ac circuit containing resistors, capacitors, and inductors, it is helpful to think of each device's reactance and find the equivalent reactance using the rules we used for equivalent resistance in the past. Phasors are a great method to determine whether the emf of the circuit has positive or negative phase (namely, leads or lags other values). A mnemonic device of "ELI the ICE man" is sometimes used to remember that the emf (E) leads the current (I) in an inductor (L) and the current (I) leads the emf (E) in a capacitor (C).

Use the following steps to determine the emf of the circuit by phasors:

1. Draw the phasors for voltage across each device: resistor, capacitor, and inductor, including the phase angle in the circuit.
2. If there is both a capacitor and an inductor, find the net voltage from these two phasors, since they are antiparallel.

3. Find the equivalent phasor from the phasor in step 2 and the resistor's phasor using trigonometry or components of the phasors. The equivalent phasor found is the emf of the circuit.

Example:**An RLC Series Circuit**

The output of an ac generator connected to an *RLC* series combination has a frequency of 200 Hz and an amplitude of 0.100 V. If $R = 4.00\ \Omega$, $L = 3.00 \times 10^{-3}\ \text{H}$, and $C = 8.00 \times 10^{-4}\ \text{F}$, what are (a) the capacitive reactance, (b) the inductive reactance, (c) the impedance, (d) the current amplitude, and (e) the phase difference between the current and the emf of the generator?

Strategy

The reactances and impedance in (a)–(c) are found by substitutions into [\[link\]](#), [\[link\]](#), and [\[link\]](#), respectively. The current amplitude is calculated from the peak voltage and the impedance. The phase difference between the current and the emf is calculated by the inverse tangent of the difference between the reactances divided by the resistance.

Solution

- a. From [\[link\]](#), the capacitive reactance is

Equation:

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi (200\ \text{Hz}) (8.00 \times 10^{-4}\ \text{F})} = 0.995\ \Omega.$$

- b. From [\[link\]](#), the inductive reactance is

Equation:

$$X_L = \omega L = 2\pi (200\ \text{Hz}) (3.00 \times 10^{-3}\ \text{H}) = 3.77\ \Omega.$$

- c. Substituting the values of R , X_C , and X_L into [\[link\]](#), we obtain for the impedance

Equation:

$$Z = \sqrt{(4.00\ \Omega)^2 + (3.77\ \Omega - 0.995\ \Omega)^2} = 4.87\ \Omega.$$

- d. The current amplitude is

Equation:

$$I_0 = \frac{V_0}{Z} = \frac{0.100\ \text{V}}{4.87\ \Omega} = 2.05 \times 10^{-2}\ \text{A}.$$

- e. From [\[link\]](#), the phase difference between the current and the emf is

Equation:

$$\phi = \tan^{-1} \frac{X_L - X_C}{R} = \tan^{-1} \frac{2.77\ \Omega}{4.00\ \Omega} = 0.607\ \text{rad}.$$

Significance

The phase angle is positive because the reactance of the inductor is larger than the reactance of the capacitor.

Note:

Exercise:

Problem:

Check Your Understanding Find the voltages across the resistor, the capacitor, and the inductor in the circuit of [\[link\]](#) using $v(t) = V_0 \sin \omega t$ as the output of the ac generator.

Solution:

$$v_R = (V_0 R/Z) \sin(\omega t - \phi); v_C = (V_0 X_C/Z) \sin(\omega t - \phi + \pi/2) = -(V_0 X_C/Z) \cos(\omega t - \phi); \\ v_L = (V_0 X_L/Z) \sin(\omega t - \phi - \pi/2) = (V_0 X_L/Z) \cos(\omega t - \phi)$$

Summary

- An *RLC* series circuit is a resistor, capacitor, and inductor series combination across an ac source.
- The same current flows through each element of an *RLC* series circuit at all points in time.
- The counterpart of resistance in a dc circuit is impedance, which measures the combined effect of resistors, capacitors, and inductors. The maximum current is defined by the ac version of Ohm's law.
- Impedance has units of ohms and is found using the resistance, the capacitive reactance, and the inductive reactance.

Conceptual Questions

Exercise:

Problem:

In an *RLC* series circuit, can the voltage measured across the capacitor be greater than the voltage of the source? Answer the same question for the voltage across the inductor.

Solution:

yes for both

Problems

Exercise:

Problem:

What is the impedance of a series combination of a $50\text{-}\Omega$ resistor, a $5.0\text{-}\mu\text{F}$ capacitor, and a $10\text{-}\mu\text{F}$ capacitor at a frequency of 2.0 kHz ?

Exercise:

Problem:

A resistor and capacitor are connected in series across an ac generator. The emf of the generator is given by $v(t) = V_0 \cos \omega t$, where $V_0 = 120 \text{ V}$, $\omega = 120\pi \text{ rad/s}$, $R = 400 \Omega$, and $C = 4.0 \mu\text{F}$.

(a) What is the impedance of the circuit? (b) What is the amplitude of the current through the resistor? (c) Write an expression for the current through the resistor. (d) Write expressions representing the voltages across the resistor and across the capacitor.

Solution:

a. 770Ω ; b. 0.16 A ; c. $I = (0.16 \text{ A}) \cos (120\pi t - 0.33\pi)$; d. $v_R = 62 \cos (120\pi t)$;
 $v_C = 103 \cos (120\pi t - \pi/2)$

Exercise:**Problem:**

A resistor and inductor are connected in series across an ac generator. The emf of the generator is given by $v(t) = V_0 \cos \omega t$, where $V_0 = 120 \text{ V}$ and $\omega = 120\pi \text{ rad/s}$; also, $R = 400 \Omega$ and $L = 1.5 \text{ H}$. (a) What is the impedance of the circuit? (b) What is the amplitude of the current through the resistor? (c) Write an expression for the current through the resistor. (d) Write expressions representing the voltages across the resistor and across the inductor.

Exercise:**Problem:**

In an RLC series circuit, the voltage amplitude and frequency of the source are 100 V and 500 Hz , respectively, an $R = 500 \Omega$, $L = 0.20 \text{ H}$, and $C = 2.0 \mu\text{F}$. (a) What is the impedance of the circuit? (b) What is the amplitude of the current from the source? (c) If the emf of the source is given by $v(t) = (100 \text{ V}) \sin 1000\pi t$, how does the current vary with time? (d) Repeat the calculations with C changed to $0.20 \mu\text{F}$.

Solution:

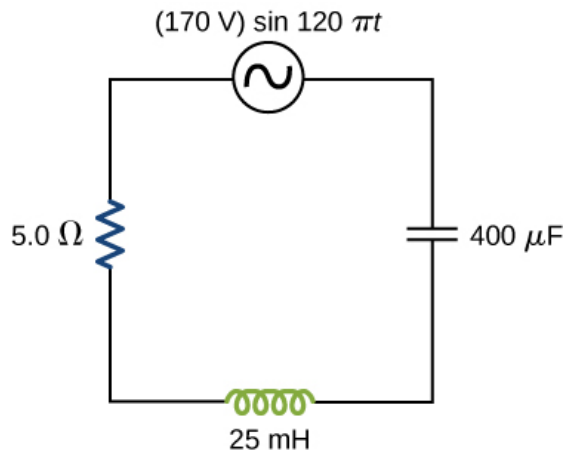
a. 690Ω ; b. 0.15 A ; c. $I = (0.15 \text{ A}) \sin (1000\pi t - 0.753)$; d. 1100Ω , 0.092 A ,
 $I = (0.092 \text{ A}) \sin (1000\pi t + 1.09)$

Exercise:**Problem:**

An RLC series circuit with $R = 600 \Omega$, $L = 30 \text{ mH}$, and $C = 0.050 \mu\text{F}$ is driven by an ac source whose frequency and voltage amplitude are 500 Hz and 50 V , respectively. (a) What is the impedance of the circuit? (b) What is the amplitude of the current in the circuit? (c) What is the phase angle between the emf of the source and the current?

Exercise:**Problem:**

For the circuit shown below, what are (a) the total impedance and (b) the phase angle between the current and the emf? (c) Write an expression for $i(t)$.



Solution:

a. $5.7 \, \Omega$; b. 29° ; c. $I = (30. \, \text{A})\cos(120\pi t + 0.51)$

Glossary

impedance

ac analog to resistance in a dc circuit, which measures the combined effect of resistance, capacitive reactance, and inductive reactance

phase angle

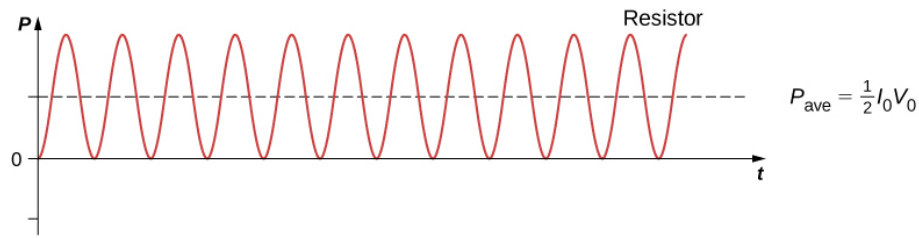
amount by which the voltage and current are out of phase with each other in a circuit

Power in an AC Circuit

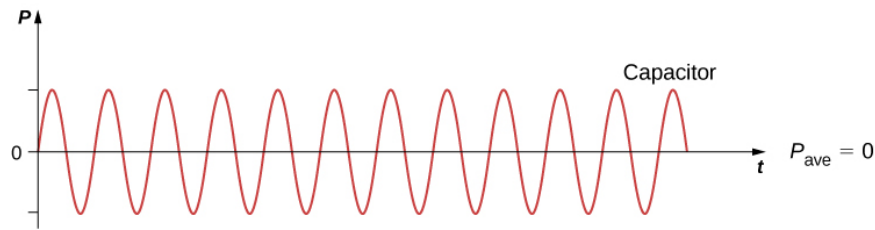
By the end of the section, you will be able to:

- Describe how average power from an ac circuit can be written in terms of peak current and voltage and of rms current and voltage
- Determine the relationship between the phase angle of the current and voltage and the average power, known as the power factor

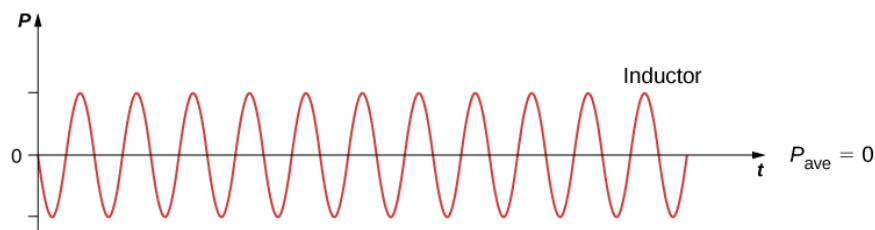
A circuit element dissipates or produces power according to $P = IV$, where I is the current through the element and V is the voltage across it. Since the current and the voltage both depend on time in an ac circuit, the instantaneous power $p(t) = i(t)v(t)$ is also time dependent. A plot of $p(t)$ for various circuit elements is shown in [\[link\]](#). For a resistor, $i(t)$ and $v(t)$ are in phase and therefore always have the same sign (see [\[link\]](#)). For a capacitor or inductor, the relative signs of $i(t)$ and $v(t)$ vary over a cycle due to their phase differences (see [\[link\]](#) and [\[link\]](#)). Consequently, $p(t)$ is positive at some times and negative at others, indicating that capacitive and inductive elements produce power at some instants and absorb it at others.



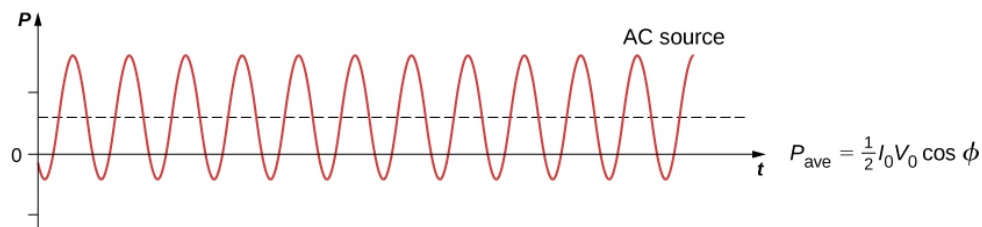
(a)



(b)



(c)



(d)

Graph of instantaneous power for various circuit elements. (a) For the resistor, $P_{\text{ave}} = I_0 V_0 / 2$, whereas for (b) the capacitor and (c) the inductor, $P_{\text{ave}} = 0$.

(d) For the source, $P_{\text{ave}} = I_0 V_0 (\cos \phi) / 2$, which may be positive, negative, or zero, depending on ϕ .

Because instantaneous power varies in both magnitude and sign over a cycle, it seldom has any practical importance. What we're almost always concerned with is the power averaged over time, which we refer to as the **average power**. It is defined by the time average of the instantaneous power over one cycle:

Equation:

$$P_{\text{ave}} = \frac{1}{T} \int_0^T p(t) dt,$$

where $T = 2\pi/\omega$ is the period of the oscillations. With the substitutions $v(t) = V_0 \sin \omega t$ and $i(t) = I_0 \sin (\omega t - \phi)$, this integral becomes

Equation:

$$P_{\text{ave}} = \frac{I_0 V_0}{T} \int_0^T \sin (\omega t - \phi) \sin \omega t dt.$$

Using the trigonometric relation $\sin (A - B) = \sin A \cos B - \sin B \cos A$, we obtain

Equation:

$$P_{\text{ave}} = \frac{I_0 V_0 \cos \phi}{T} \int_0^T \sin^2 \omega t dt - \frac{I_0 V_0 \sin \phi}{T} \int_0^T \sin \omega t \cos \omega t dt.$$

Evaluation of these two integrals yields

Equation:

$$\frac{1}{T} \int_0^T \sin^2 \omega t dt = \frac{1}{2}$$

and

Equation:

$$\frac{1}{T} \int_0^T \sin \omega t \cos \omega t dt = 0.$$

Hence, the average power associated with a circuit element is given by

Note:

Equation:

$$P_{\text{ave}} = \frac{1}{2} I_0 V_0 \cos \phi.$$

In engineering applications, $\cos \phi$ is known as the **power factor**, which is the amount by which the power delivered in the circuit is less than the theoretical maximum of the circuit due to voltage and current being out of phase. For a resistor, $\phi = 0$, so the average power dissipated is

Equation:

$$P_{\text{ave}} = \frac{1}{2} I_0 V_0.$$

A comparison of $p(t)$ and P_{ave} is shown in [\[link\]](#)(d). To make $P_{\text{ave}} = (1/2)I_0V_0$ look like its dc counterpart, we use the rms values I_{rms} and V_{rms} of the current and the voltage. By definition, these are **Equation:**

$$I_{\text{rms}} = \sqrt{i_{\text{ave}}^2} \text{ and } V_{\text{rms}} = \sqrt{v_{\text{ave}}^2},$$

where

Equation:

$$i_{\text{ave}}^2 = \frac{1}{T} \int_0^T i^2(t) dt \text{ and } v_{\text{ave}}^2 = \frac{1}{T} \int_0^T v^2(t) dt.$$

With $i(t) = I_0 \sin(\omega t - \phi)$ and $v(t) = V_0 \sin \omega t$, we obtain

Equation:

$$I_{\text{rms}} = \frac{1}{\sqrt{2}} I_0 \text{ and } V_{\text{rms}} = \frac{1}{\sqrt{2}} V_0.$$

We may then write for the average power dissipated by a resistor,

Note:

Equation:

$$P_{\text{ave}} = \frac{1}{2} I_0 V_0 = I_{\text{rms}} V_{\text{rms}} = I_{\text{rms}}^2 R.$$

This equation further emphasizes why the rms value is chosen in discussion rather than peak values. Both equations for average power are correct for [\[link\]](#), but the rms values in the formula give a cleaner representation, so the extra factor of 1/2 is not necessary.

Alternating voltages and currents are usually described in terms of their rms values. For example, the 110 V from a household outlet is an rms value. The amplitude of this source is $110\sqrt{2} \text{ V} = 156 \text{ V}$. Because most ac meters are calibrated in terms of rms values, a typical ac voltmeter placed across a household outlet will read 110 V.

For a capacitor and an inductor, $\phi = \pi/2$ and $-\pi/2$ rad, respectively. Since $\cos \pi/2 = \cos(-\pi/2) = 0$, we find from [\[link\]](#) that the average power dissipated by either of these elements is $P_{\text{ave}} = 0$. Capacitors and inductors absorb energy from the circuit during one half-cycle and then discharge it back to the circuit during the other half-cycle. This behavior is illustrated in the plots of [\[link\]](#), (b) and (c), which show $p(t)$ oscillating sinusoidally about zero.

The phase angle for an ac generator may have any value. If $\cos \phi > 0$, the generator produces power; if $\cos \phi < 0$, it absorbs power. In terms of rms values, the average power of an ac generator is written as

Equation:

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}} \cos \phi.$$

For the generator in an *RLC* circuit,

Equation:

$$\tan \phi = \frac{X_L - X_C}{R}$$

and

Equation:

$$\cos \phi = \frac{R}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{R}{Z}.$$

Hence the average power of the generator is

Equation:

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}} \cos \phi = \frac{V_{\text{rms}}}{Z} V_{\text{rms}} \frac{R}{Z} = \frac{V_{\text{rms}}^2 R}{Z^2}.$$

This can also be written as

Equation:

$$P_{\text{ave}} = I_{\text{rms}}^2 R,$$

which designates that the power produced by the generator is dissipated in the resistor. As we can see, Ohm's law for the rms ac is found by dividing the rms voltage by the impedance.

Example:

Power Output of a Generator

An ac generator whose emf is given by

Equation:

$$v(t) = (4.00 \text{ V}) \sin [(1.00 \times 10^4 \text{ rad/s})t]$$

is connected to an *RLC* circuit for which $L = 2.00 \times 10^{-3} \text{ H}$, $C = 4.00 \times 10^{-6} \text{ F}$, and $R = 5.00 \Omega$.

(a) What is the rms voltage across the generator? (b) What is the impedance of the circuit? (c) What is the average power output of the generator?

Strategy

The rms voltage is the amplitude of the voltage times $1/\sqrt{2}$. The impedance of the circuit involves the resistance and the reactances of the capacitor and the inductor. The average power is calculated by [\[link\]](#), or more specifically, the last part of the equation, because we have the impedance of the circuit Z , the rms voltage V_{rms} , and the resistance R .

Solution

a. Since $V_0 = 4.00 \text{ V}$, the rms voltage across the generator is

Equation:

$$V_{\text{rms}} = \frac{1}{\sqrt{2}} (4.00 \text{ V}) = 2.83 \text{ V}.$$

b. The impedance of the circuit is

Equation:

$$\begin{aligned} Z &= \sqrt{R^2 + (X_L - X_C)^2} \\ &= \left\{ (5.00 \, \Omega)^2 + \left[(1.00 \times 10^4 \text{ rad/s}) (2.00 \times 10^{-3} \text{ H}) - \frac{1}{(1.00 \times 10^4 \text{ rad/s}) (4.00 \times 10^{-6} \text{ F})} \right]^2 \right\}^{1/2} \\ &= 7.07 \, \Omega. \end{aligned}$$

c. From [\[link\]](#), the average power transferred to the circuit is

Equation:

$$P_{\text{ave}} = \frac{V_{\text{rms}}^2 R}{Z^2} = \frac{(2.83 \text{ V})^2 (5.00 \, \Omega)}{(7.07 \, \Omega)^2} = 0.801 \text{ W}.$$

Significance

If the resistance is much larger than the reactance of the capacitor or inductor, the average power is a dc circuit equation of $P = V^2/R$, where V replaces the rms voltage.

Note:

Exercise:

Problem:

Check Your Understanding An ac voltmeter attached across the terminals of a 45-Hz ac generator reads 7.07 V. Write an expression for the emf of the generator.

Solution:

$$v(t) = (10.0 \text{ V}) \sin 90\pi t$$

Note:

Exercise:

Problem:

Check Your Understanding Show that the rms voltages across a resistor, a capacitor, and an inductor in an ac circuit where the rms current is I_{rms} are given by $I_{\text{rms}}R$, $I_{\text{rms}}X_C$, and $I_{\text{rms}}X_L$, respectively. Determine these values for the components of the RLC circuit of [\[link\]](#).

Solution:

2.00 V; 10.01 V; 8.01 V

Summary

- The average ac power is found by multiplying the rms values of current and voltage.
- Ohm's law for the rms ac is found by dividing the rms voltage by the impedance.
- In an ac circuit, there is a phase angle between the source voltage and the current, which can be found by dividing the resistance by the impedance.
- The average power delivered to an *RLC* circuit is affected by the phase angle.
- The power factor ranges from -1 to 1 .

Conceptual Questions

Exercise:

Problem:

For what value of the phase angle ϕ between the voltage output of an ac source and the current is the average power output of the source a maximum?

Exercise:

Problem: Discuss the differences between average power and instantaneous power.

Solution:

The instantaneous power is the power at a given instant. The average power is the power averaged over a cycle or number of cycles.

Exercise:

Problem:

The average ac current delivered to a circuit is zero. Despite this, power is dissipated in the circuit. Explain.

Exercise:

Problem:

Can the instantaneous power output of an ac source ever be negative? Can the average power output be negative?

Solution:

The instantaneous power can be negative, but the power output can't be negative.

Exercise:

Problem:

The power rating of a resistor used in ac circuits refers to the maximum average power dissipated in the resistor. How does this compare with the maximum instantaneous power dissipated in the resistor?

Problems

Exercise:

Problem:

The emf of an ac source is given by $v(t) = V_0 \sin \omega t$, where $V_0 = 100 \text{ V}$ and $\omega = 200\pi \text{ rad/s}$. Calculate the average power output of the source if it is connected across (a) a $20\text{-}\mu\text{F}$ capacitor, (b) a 20-mH inductor, and (c) a $50\text{-}\Omega$ resistor.

Exercise:

Problem:

Calculate the rms currents for an ac source is given by $v(t) = V_0 \sin \omega t$, where $V_0 = 100 \text{ V}$ and $\omega = 200\pi \text{ rad/s}$ when connected across (a) a $20\text{-}\mu\text{F}$ capacitor, (b) a 20-mH inductor, and (c) a $50\text{-}\Omega$ resistor.

Solution:

a. 0.89 A ; b. 5.6A ; c. 1.4 A

Exercise:

Problem:

A 40-mH inductor is connected to a 60-Hz AC source whose voltage amplitude is 50 V . If an AC voltmeter is placed across the inductor, what does it read?

Exercise:

Problem:

For an RLC series circuit, the voltage amplitude and frequency of the source are 100 V and 500 Hz , respectively; $R = 500 \text{ }\Omega$; and $L = 0.20 \text{ H}$. Find the average power dissipated in the resistor for the following values for the capacitance: (a) $C = 2.0\text{ }\mu\text{F}$ and (b) $C = 0.20 \text{ }\mu\text{F}$.

Solution:

a. 5.3 W ; b. 2.1 W

Exercise:

Problem:

An ac source of voltage amplitude 10 V delivers electric energy at a rate of 0.80 W when its current output is 2.5 A . What is the phase angle ϕ between the emf and the current?

Exercise:

Problem:

An RLC series circuit has an impedance of $60 \text{ }\Omega$ and a power factor of 0.50 , with the voltage lagging the current. (a) Should a capacitor or an inductor be placed in series with the elements to raise the power factor of the circuit? (b) What is the value of the reactance across the inductor that will raise the power factor to unity?

Solution:

a. inductor; b. $X_L = 52 \Omega$

Glossary

average power

time average of the instantaneous power over one cycle

power factor

amount by which the power delivered in the circuit is less than the theoretical maximum of the circuit due to voltage and current being out of phase

Resonance in an AC Circuit

By the end of the section, you will be able to:

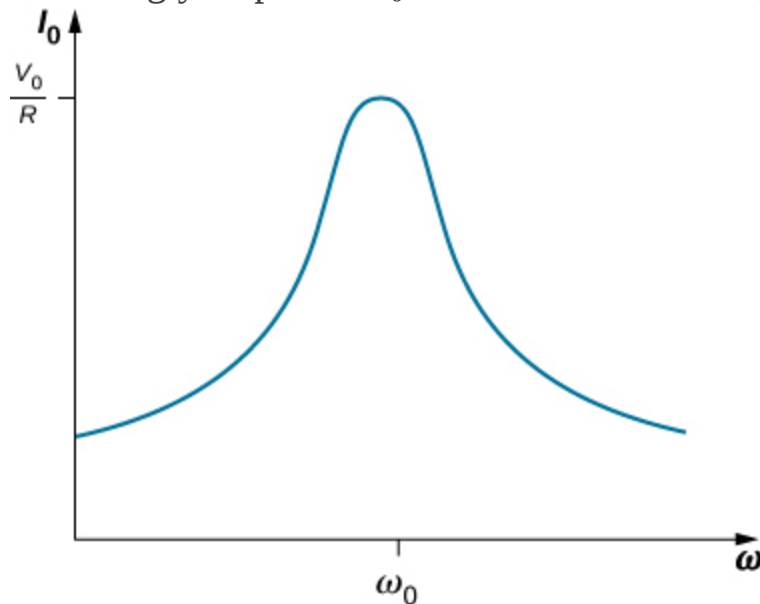
- Determine the peak ac resonant angular frequency for a RLC circuit
- Explain the width of the average power versus angular frequency curve and its significance using terms like bandwidth and quality factor

In the *RLC* series circuit of [\[link\]](#), the current amplitude is, from [\[link\]](#),

Equation:

$$I_0 = \frac{V_0}{\sqrt{R^2 + (\omega L - 1/\omega C)^2}}.$$

If we can vary the frequency of the ac generator while keeping the amplitude of its output voltage constant, then the current changes accordingly. A plot of I_0 versus ω is shown in [\[link\]](#).



At an *RLC* circuit's resonant frequency, $\omega_0 = \sqrt{1/LC}$, the current amplitude is at its maximum value.

In [Oscillations](#), we encountered a similar graph where the amplitude of a damped harmonic oscillator was plotted against the angular frequency of a sinusoidal driving force (see [Forced Oscillations](#)). This similarity is more than just a coincidence, as shown earlier by the application of Kirchhoff's loop rule to the circuit of [\[link\]](#). This yields

Equation:

$$L \frac{di}{dt} + iR + \frac{q}{C} = V_0 \sin \omega t,$$

or

Equation:

$$L \frac{d^2 q}{dt^2} + R \frac{dq}{dt} + \frac{1}{C} q = V_0 \sin \omega t,$$

where we substituted $dq(t)/dt$ for $i(t)$. A comparison of [\[link\]](#) and, from [Oscillations](#), [Damped Oscillations](#) for damped harmonic motion clearly demonstrates that the driven RLC series circuit is the electrical analog of the driven damped harmonic oscillator.

The **resonant frequency** f_0 of the RLC circuit is the frequency at which the amplitude of the current is a maximum and the circuit would oscillate if not driven by a voltage source. By inspection, this corresponds to the angular frequency $\omega_0 = 2\pi f_0$ at which the impedance Z in [\[link\]](#) is a minimum, or when

Equation:

$$\omega_0 L = \frac{1}{\omega_0 C}$$

and

Note:

Equation:

$$\omega_0 = \sqrt{\frac{1}{LC}}.$$

This is the resonant angular frequency of the circuit. Substituting ω_0 into [\[link\]](#), [\[link\]](#), and [\[link\]](#), we find that at resonance,

Equation:

$$\phi = \tan^{-1}(0) = 0, \quad I_0 = V_0/R, \quad \text{and} \quad Z = R.$$

Therefore, at resonance, an *RLC* circuit is purely resistive, with the applied emf and current in phase.

What happens to the power at resonance? [\[link\]](#) tells us how the average power transferred from an ac generator to the *RLC* combination varies with frequency. In addition, P_{ave} reaches a maximum when Z , which depends on the frequency, is a minimum, that is, when $X_L = X_C$ and $Z = R$. Thus, at resonance, the average power output of the source in an *RLC* series circuit is a maximum. From [\[link\]](#), this maximum is V_{rms}^2/R .

[\[link\]](#) is a typical plot of P_{ave} versus ω in the region of maximum power output. The **bandwidth** $\Delta\omega$ of the resonance peak is defined as the range of angular frequencies ω over which the average power P_{ave} is greater than one-half the maximum value of P_{ave} . The sharpness of the peak is described by a dimensionless quantity known as the **quality factor** Q of the circuit. By definition,

Note:

Equation:

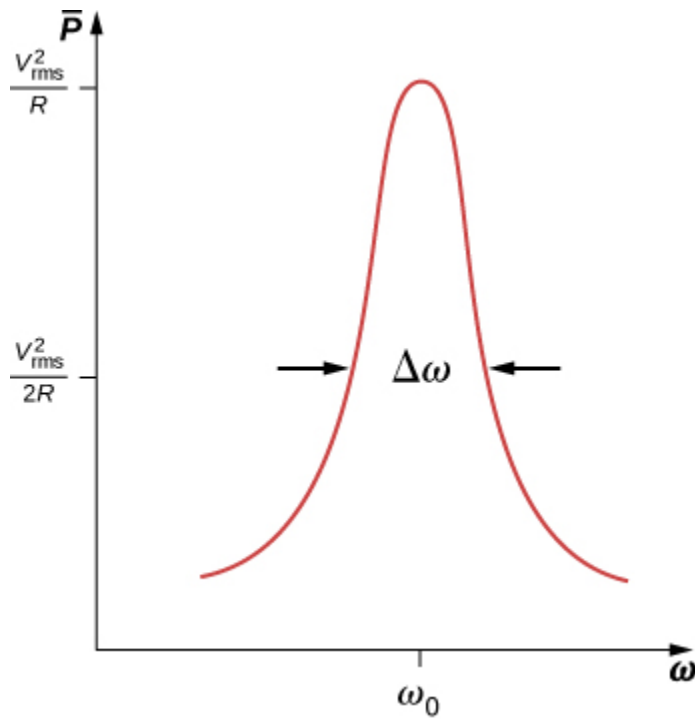
$$Q = \frac{\omega_0}{\Delta\omega},$$

where ω_0 is the resonant angular frequency. A high Q indicates a sharp resonance peak. We can give Q in terms of the circuit parameters as

Note:

Equation:

$$Q = \frac{\omega_0 L}{R}.$$



Like the current, the average power

transferred from an ac generator to an *RLC* circuit peaks at the resonant frequency.

Resonant circuits are commonly used to pass or reject selected frequency ranges. This is done by adjusting the value of one of the elements and hence “tuning” the circuit to a particular resonant frequency. For example, in radios, the receiver is tuned to the desired station by adjusting the resonant frequency of its circuitry to match the frequency of the station. If the tuning circuit has a high Q , it will have a small bandwidth, so signals from other stations at frequencies even slightly different from the resonant frequency encounter a high impedance and are not passed by the circuit. Cell phones work in a similar fashion, communicating with signals of around 1 GHz that are tuned by an inductor-capacitor circuit. One of the most common applications of capacitors is their use in ac-timing circuits, based on attaining a resonant frequency. A metal detector also uses a shift in resonance frequency in detecting metals ([\[link\]](#)).



When a metal detector comes near a piece of metal, the self-inductance of one of its coils changes. This causes a shift in the resonant frequency of a circuit containing the coil. That shift is detected by the circuitry and transmitted to the diver by means of the headphones. (credit: modification of work by Eric Lippmann, U.S. Navy)

Example:
Resonance in an RLC Series Circuit

(a) What is the resonant frequency of a circuit using the voltage and LRC values all wired in series from [\[link\]](#)? (b) If the ac generator is set to this frequency without changing the amplitude of the output voltage, what is the amplitude of the current?

Strategy

The resonant frequency for a RLC circuit is calculated from [\[link\]](#), which comes from a balance between the reactances of the capacitor and the inductor. Since the circuit is at resonance, the impedance is equal to the resistor. Then, the peak current is calculated by the voltage divided by the resistance.

Solution

a. The resonant frequency is found from [\[link\]](#):

Equation:

$$\begin{aligned} f_0 &= \frac{1}{2\pi} \sqrt{\frac{1}{LC}} = \frac{1}{2\pi} \sqrt{\frac{1}{(3.00 \times 10^{-3} \text{ H})(8.00 \times 10^{-4} \text{ F})}} \\ &= 1.03 \times 10^2 \text{ Hz.} \end{aligned}$$

b. At resonance, the impedance of the circuit is purely resistive, and the current amplitude is

Equation:

$$I_0 = \frac{0.100 \text{ V}}{4.00 \Omega} = 2.50 \times 10^{-2} \text{ A.}$$

Significance

If the circuit were not set to the resonant frequency, we would need the impedance of the entire circuit to calculate the current.

Example:

Power Transfer in an RLC Series Circuit at Resonance

(a) What is the resonant angular frequency of an RLC circuit with $R = 0.200 \Omega$, $L = 4.00 \times 10^{-3} \text{ H}$, and $C = 2.00 \times 10^{-6} \text{ F}$? (b) If an

ac source of constant amplitude 4.00 V is set to this frequency, what is the average power transferred to the circuit? (c) Determine Q and the bandwidth of this circuit.

Strategy

The resonant angular frequency is calculated from [\[link\]](#). The average power is calculated from the rms voltage and the resistance in the circuit. The quality factor is calculated from [\[link\]](#) and by knowing the resonant frequency. The bandwidth is calculated from [\[link\]](#) and by knowing the quality factor.

Solution

- a. The resonant angular frequency is

Equation:

$$\begin{aligned}\omega_0 &= \sqrt{\frac{1}{LC}} = \sqrt{\frac{1}{(4.00 \times 10^{-3} \text{ H})(2.00 \times 10^{-6} \text{ F})}} \\ &= 1.12 \times 10^4 \text{ rad/s.}\end{aligned}$$

- b. At this frequency, the average power transferred to the circuit is a maximum. It is

Equation:

$$P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R} = \frac{\left[\left(1/\sqrt{2}\right) (4.00 \text{ V}) \right]^2}{0.200 \, \Omega} = 40.0 \text{ W.}$$

- c. The quality factor of the circuit is

Equation:

$$Q = \frac{\omega_0 L}{R} = \frac{(1.12 \times 10^4 \text{ rad/s}) (4.00 \times 10^{-3} \text{ H})}{0.200 \, \Omega} = 224.$$

We then find for the bandwidth

Equation:

$$\Delta\omega = \frac{\omega_0}{Q} = \frac{1.12 \times 10^4 \text{ rad/s}}{224} = 50.0 \text{ rad/s}.$$

Significance

If a narrower bandwidth is desired, a lower resistance or higher inductance would help. However, a lower resistance increases the power transferred to the circuit, which may not be desirable, depending on the maximum power that could possibly be transferred.

Note:**Exercise:****Problem:**

Check Your Understanding In the circuit of [\[link\]](#), $L = 2.0 \times 10^{-3} \text{ H}$, $C = 5.0 \times 10^{-4} \text{ F}$, and $R = 40 \Omega$. (a) What is the resonant frequency? (b) What is the impedance of the circuit at resonance? (c) If the voltage amplitude is 10 V, what is $i(t)$ at resonance? (d) The frequency of the AC generator is now changed to 200 Hz. Calculate the phase difference between the current and the emf of the generator.

Solution:

a. 160 Hz; b. 40Ω ; c. $(0.25 \text{ A}) \sin 10^3 t$; d. 0.023 rad

Note:**Exercise:**

Problem:

Check Your Understanding What happens to the resonant frequency of an *RLC* series circuit when the following quantities are increased by a factor of 4: (a) the capacitance, (b) the self-inductance, and (c) the resistance?

Solution:

a. halved; b. halved; c. same

Note:**Exercise:****Problem:**

Check Your Understanding The resonant angular frequency of an *RLC* series circuit is 4.0×10^2 rad/s. An ac source operating at this frequency transfers an average power of 2.0×10^{-2} W to the circuit. The resistance of the circuit is 0.50Ω . Write an expression for the emf of the source.

Solution:

$$v(t) = (0.14 \text{ V}) \sin(4.0 \times 10^2 t)$$

Summary

- At the resonant frequency, inductive reactance equals capacitive reactance.
- The average power versus angular frequency plot for a *RLC* circuit has a peak located at the resonant frequency; the sharpness or width of the peak is known as the bandwidth.

- The bandwidth is related to a dimensionless quantity called the quality factor. A high quality factor value is a sharp or narrow peak.

Problems

Exercise:

Problem:

(a) Calculate the resonant angular frequency of an RLC series circuit for which $R = 20\ \Omega$, $L = 75\ \text{mH}$, and $C = 4.0\ \mu\text{F}$. (b) If R is changed to $300\ \Omega$, what happens to the resonant angular frequency?

Exercise:

Problem:

The resonant frequency of an RLC series circuit is $2.0 \times 10^3\ \text{Hz}$. If the self-inductance in the circuit is $5.0\ \text{mH}$, what is the capacitance in the circuit?

Solution:

$$1.3 \times 10^{-6}\ \text{F}$$

Exercise:

Problem:

(a) What is the resonant frequency of an RLC series circuit with $R = 20\ \Omega$, $L = 2.0\ \text{mH}$, and $C = 4.0\ \mu\text{F}$? (b) What is the impedance of the circuit at resonance?

Exercise:

Problem:

For an RLC series circuit, $R = 100\ \Omega$, $L = 150\ \text{mH}$, and $C = 0.25\ \mu\text{F}$. (a) If an ac source of variable frequency is connected to the circuit, at what frequency is maximum power dissipated in the resistor? (b) What is the quality factor of the circuit?

Solution:

a. 820 Hz; b. 7.8

Exercise:**Problem:**

An ac source of voltage amplitude 100 V and variable frequency f drives an RLC series circuit with $R = 10\ \Omega$, $L = 2.0\text{ mH}$, and $C = 25\ \mu\text{F}$. (a) Plot the current through the resistor as a function of the frequency f . (b) Use the plot to determine the resonant frequency of the circuit.

Exercise:**Problem:**

(a) What is the resonant frequency of a resistor, capacitor, and inductor connected in series if $R = 100\ \Omega$, $L = 2.0\text{ H}$, and $C = 5.0\ \mu\text{F}$? (b) If this combination is connected to a 100-V source operating at the resonant frequency, what is the power output of the source? (c) What is the Q of the circuit? (d) What is the bandwidth of the circuit?

Solution:

a. 50 Hz; b. 50 W; c. 6.32; d. 50 rad/s

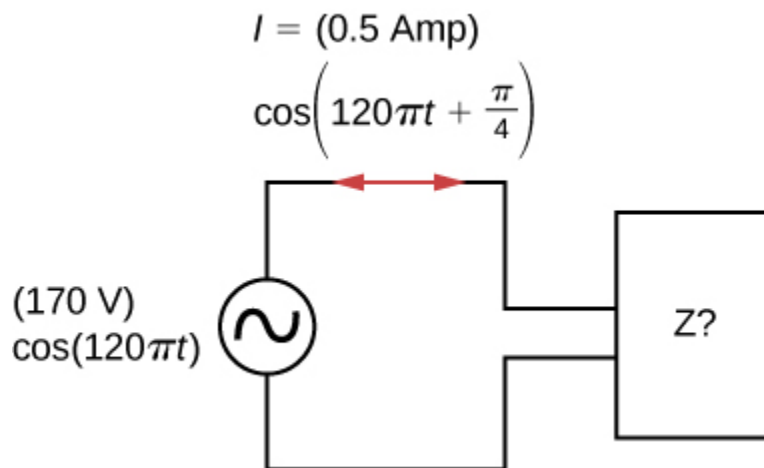
Exercise:**Problem:**

Suppose a coil has a self-inductance of 20.0 H and a resistance of $200\ \Omega$. What (a) capacitance and (b) resistance must be connected in series with the coil to produce a circuit that has a resonant frequency of 100 Hz and a Q of 10?

Exercise:

Problem:

An ac generator is connected to a device whose internal circuits are not known. We only know current and voltage outside the device, as shown below. Based on the information given, what can you infer about the electrical nature of the device and its power usage?



Solution:

The reactance of the capacitor is larger than the reactance of the inductor because the current leads the voltage. The power usage is 30 W.

Glossary**bandwidth**

range of angular frequencies over which the average power is greater than one-half the maximum value of the average power

quality factor

dimensionless quantity that describes the sharpness of the peak of the bandwidth; a high quality factor is a sharp or narrow resonance peak

resonant frequency

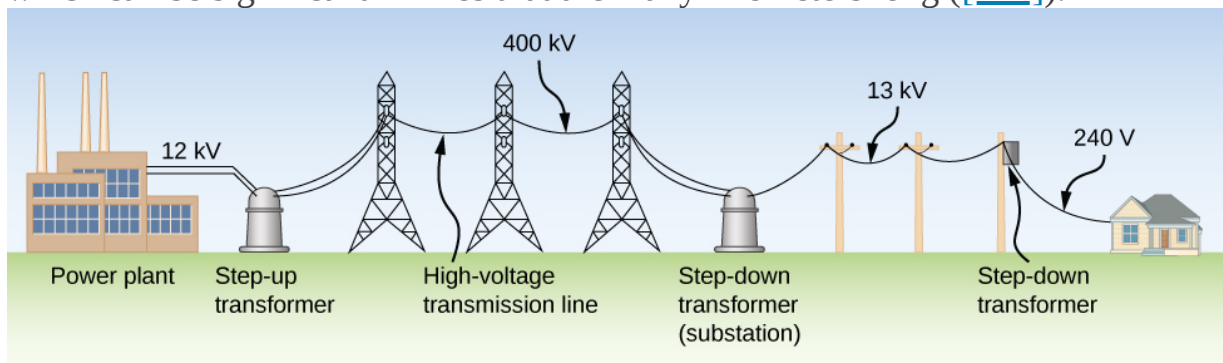
frequency at which the amplitude of the current is a maximum and the circuit would oscillate if not driven by a voltage source

Transformers

By the end of the section, you will be able to:

- Explain why power plants transmit electricity at high voltages and low currents and how they do this
- Develop relationships among current, voltage, and the number of windings in step-up and step-down transformers

Although ac electric power is produced at relatively low voltages, it is sent through transmission lines at very high voltages (as high as 500 kV). The same power can be transmitted at different voltages because power is the product $I_{\text{rms}} V_{\text{rms}}$. (For simplicity, we ignore the phase factor $\cos \phi$.) A particular power requirement can therefore be met with a low voltage and a high current or with a high voltage and a low current. The advantage of the high-voltage/low-current choice is that it results in lower $I_{\text{rms}}^2 R$ ohmic losses in the transmission lines, which can be significant in lines that are many kilometers long ([\[link\]](#)).



The rms voltage from a power plant eventually needs to be stepped down from 12 kV to 240 V so that it can be safely introduced into a home. A high-voltage transmission line allows a low current to be transmitted via a substation over long distances.

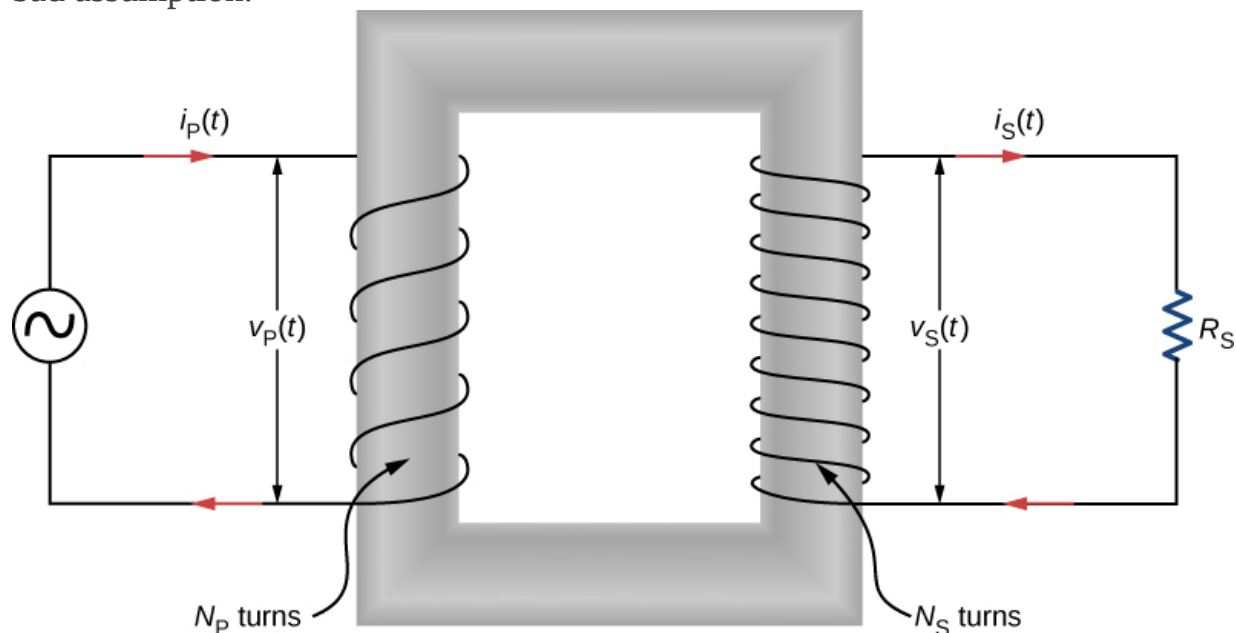
Typically, the alternating emfs produced at power plants are “stepped up” to very high voltages before being transmitted through power lines; then, they must be “stepped down” to relatively safe values (110 or 220 V rms) before they are introduced into homes. The device that transforms voltages from one value to another using induction is the **transformer** ([\[link\]](#)).



Transformers are used to step down the high voltages in transmission lines to the 110 to 220 V used in homes.
(credit: modification of work by “Fortyseven”/Flickr)

As [\[link\]](#) illustrates, a transformer basically consists of two separated coils, or windings, wrapped around a soft iron core. The primary winding has N_P loops, or turns, and is connected to an alternating voltage $v_P(t)$. The secondary winding has N_S turns and is connected to a load resistor R_S . We assume the ideal case for

which all magnetic field lines are confined to the core so that the same magnetic flux permeates each turn of both the primary and the secondary windings. We also neglect energy losses to magnetic hysteresis, to ohmic heating in the windings, and to ohmic heating of the induced eddy currents in the core. A good transformer can have losses as low as 1% of the transmitted power, so this is not a bad assumption.



A step-up transformer (more turns in the secondary winding than in the primary winding). The two windings are wrapped around a soft iron core.

To analyze the transformer circuit, we first consider the primary winding. The input voltage $v_P(t)$ is equal to the potential difference induced across the primary winding. From Faraday's law, the induced potential difference is $-N_P (d\Phi/dt)$, where Φ is the flux through one turn of the primary winding. Thus,

Equation:

$$v_P(t) = -N_P \frac{d\Phi}{dt}.$$

Similarly, the output voltage $v_S(t)$ delivered to the load resistor must equal the potential difference induced across the secondary winding. Since the transformer is ideal, the flux through every turn of the secondary winding is also Φ , and

Equation:

$$v_S(t) = -N_S \frac{d\Phi}{dt}.$$

Combining the last two equations, we have

Equation:

$$v_S(t) = \frac{N_S}{N_P} v_P(t).$$

Hence, with appropriate values for N_S and N_P , the input voltage $v_P(t)$ may be “stepped up” ($N_S > N_P$) or “stepped down” ($N_S < N_P$) to $v_S(t)$, the output voltage. This is often abbreviated as the **transformer equation**,

Note:

Equation:

$$\frac{V_S}{V_P} = \frac{N_S}{N_P},$$

which shows that the ratio of the secondary to primary voltages in a transformer equals the ratio of the number of turns in their windings. For a **step-up transformer**, which increases voltage and decreases current, this ratio is greater than one; for a **step-down transformer**, which decreases voltage and increases current, this ratio is less than one.

From the law of energy conservation, the power introduced at any instant by $v_P(t)$ to the primary winding must be equal to the power dissipated in the resistor of the secondary circuit; thus,

Equation:

$$i_P(t)v_P(t) = i_S(t)v_S(t).$$

When combined with [\[link\]](#), this gives

Equation:

$$i_S(t) = \frac{N_P}{N_S} i_P(t).$$

If the voltage is stepped up, the current is stepped down, and vice versa.

Finally, we can use $i_S(t) = v_S(t)/R_S$, along with [\[link\]](#) and [\[link\]](#), to obtain

Equation:

$$v_P(t) = i_P \left[\left(\frac{N_P}{N_S} \right)^2 R_S \right],$$

which tells us that the input voltage $v_P(t)$ “sees” not a resistance R_S but rather a resistance

Equation:

$$R_P = \left(\frac{N_P}{N_S} \right)^2 R_S.$$

Our analysis has been based on instantaneous values of voltage and current. However, the resulting equations are not limited to instantaneous values; they hold also for maximum and rms values.

Example:

A Step-Down Transformer

A transformer on a utility pole steps the rms voltage down from 12 kV to 240 V.

(a) What is the ratio of the number of secondary turns to the number of primary turns? (b) If the input current to the transformer is 2.0 A, what is the output current? (c) Determine the power loss in the transmission line.

Strategy

The number of turns related to the voltages is found from [\[link\]](#). The output current is calculated using [\[link\]](#).

Solution

- a. Using [\[link\]](#) with rms values V_P and V_S , we have

Equation:

$$\frac{N_S}{N_P} = \frac{240 \text{ V}}{12 \times 10^3 \text{ V}} = \frac{1}{50},$$

so the primary winding has 50 times the number of turns in the secondary winding.

- b. From [\[link\]](#), the output rms current I_S is found using the transformer equation with current

Note:

Equation:

$$I_S = \frac{N_P}{N_S} I_P$$

such that

Equation:

$$I_S = \frac{N_P}{N_S} I_P = (50) (2.0 \text{ A}) = 100 \text{ A}.$$

- c. The power loss in the transmission line is calculated to be

Equation:

$$P_{\text{loss}} = I_P^2 R = (2.0 \text{ A})^2 (6000 \Omega) = 24.000 \text{ W}.$$

- d. If there were no transformer, the power would have to be sent at 240 V to work for these houses, and the power loss would be

Equation:

$$P_{\text{loss}} = I_S^2 R = (100 \text{ A})^2 (200 \Omega) = 2 \times 10^6 \text{ W}.$$

Therefore, when power needs to be transmitted, we want to avoid power loss. Thus, lines are sent with high voltages and low currents and adjusted with a transformer before power is sent into homes.

Significance

This application of a step-down transformer allows a home that uses 240-V outlets to have 100 A available to draw upon. This can power many devices in the home.

Note:

Exercise:

Problem:

Check Your Understanding A transformer steps the line voltage down from 110 to 9.0 V so that a current of 0.50 A can be delivered to a doorbell.

(a) What is the ratio of the number of turns in the primary and secondary windings? (b) What is the current in the primary winding? (c) What is the resistance seen by the 110-V source?

Solution:

a. 12:1; b. 0.042 A; c. $2.6 \times 10^3 \Omega$

Summary

- Power plants transmit high voltages at low currents to achieve lower ohmic losses in their many kilometers of transmission lines.
- Transformers use induction to transform voltages from one value to another.
- For a transformer, the voltages across the primary and secondary coils, or windings, are related by the transformer equation.
- The currents in the primary and secondary windings are related by the number of primary and secondary loops, or turns, in the windings of the transformer.
- A step-up transformer increases voltage and decreases current, whereas a step-down transformer decreases voltage and increases current.

Key Equations

AC voltage	$v = V_0 \sin \omega t$
AC current	$i = I_0 \sin \omega t$
capacitive reactance	$\frac{V_0}{I_0} = \frac{1}{\omega C} = X_C$
rms voltage	$V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$
rms current	$I_{\text{rms}} = \frac{I_0}{\sqrt{2}}$
inductive reactance	$\frac{V_0}{I_0} = \omega L = X_L$
Phase angle of an RLC series circuit	$\phi = \tan^{-1} \frac{X_L - X_C}{R}$
AC version of Ohm's law	$I_0 = \frac{V_0}{Z}$
Impedance of an RLC series circuit	$Z = \sqrt{R^2 + (X_L - X_C)^2}$
Average power associated with a circuit element	$P_{\text{ave}} = \frac{1}{2} I_0 V_0 \cos \phi$
Average power dissipated by a resistor	$P_{\text{ave}} = \frac{1}{2} I_0 V_0 = I_{\text{rms}} V_{\text{rms}} = I_{\text{rms}}^2 R$
Resonant angular frequency of a circuit	$\omega_0 = \sqrt{\frac{1}{LC}}$
Quality factor of a circuit	$Q = \frac{\omega_0}{\Delta\omega}$

Quality factor of a circuit in terms of the circuit parameters	$Q = \frac{\omega_0 L}{R}$
Transformer equation with voltage	$\frac{V_S}{V_P} = \frac{N_S}{N_P}$
Transformer equation with current	$I_S = \frac{N_P}{N_S} I_P$

Conceptual Questions

Exercise:

Problem:

Why do transmission lines operate at very high voltages while household circuits operate at fairly small voltages?

Solution:

There is less thermal loss if the transmission lines operate at low currents and high voltages.

Exercise:

Problem:

How can you distinguish the primary winding from the secondary winding in a step-up transformer?

Exercise:

Problem:

Battery packs in some electronic devices are charged using an adapter connected to a wall socket. Speculate as to the purpose of the adapter.

Solution:

The adapter has a step-down transformer to have a lower voltage and possibly higher current at which the device can operate.

Exercise:

Problem: Will a transformer work if the input is a dc voltage?

Exercise:

Problem:

Why are the primary and secondary coils of a transformer wrapped around the same closed loop of iron?

Solution:

so each loop can experience the same changing magnetic flux

Problems

Exercise:

Problem:

A step-up transformer is designed so that the output of its secondary winding is 2000 V (rms) when the primary winding is connected to a 110-V (rms) line voltage. (a) If there are 100 turns in the primary winding, how many turns are there in the secondary winding? (b) If a resistor connected across the secondary winding draws an rms current of 0.75 A, what is the current in the primary winding?

Exercise:

Problem:

A step-up transformer connected to a 110-V line is used to supply a hydrogen-gas discharge tube with 5.0 kV (rms). The tube dissipates 75 W of power. (a) What is the ratio of the number of turns in the secondary winding to the number of turns in the primary winding? (b) What are the rms currents in the primary and secondary windings? (c) What is the effective resistance seen by the 110-V source?

Solution:

a. 45:1; b. 0.68 A, 0.015 A; c. 160 Ω

Exercise:

Problem:

An ac source of emf delivers 5.0 mW of power at an rms current of 2.0 mA when it is connected to the primary coil of a transformer. The rms voltage across the secondary coil is 20 V. (a) What are the voltage across the primary coil and the current through the secondary coil? (b) What is the ratio of secondary to primary turns for the transformer?

Exercise:**Problem:**

A transformer is used to step down 110 V from a wall socket to 9.0 V for a radio. (a) If the primary winding has 500 turns, how many turns does the secondary winding have? (b) If the radio operates at a current of 500 mA, what is the current through the primary winding?

Solution:

a. 41 turns; b. 40.9 mA

Exercise:**Problem:**

A transformer is used to supply a 12-V model train with power from a 110-V wall plug. The train operates at 50 W of power. (a) What is the rms current in the secondary coil of the transformer? (b) What is the rms current in the primary coil? (c) What is the ratio of the number of primary to secondary turns? (d) What is the resistance of the train? (e) What is the resistance seen by the 110-V source?

Additional Problems**Exercise:****Problem:**

The emf of an ac source is given by $v(t) = V_0 \sin \omega t$, where $V_0 = 100$ V and $\omega = 200\pi$ rad/s. Find an expression that represents the output current of the source if it is connected across (a) a $20\text{-}\mu\text{F}$ capacitor, (b) a 20-mH inductor, and (c) a $50\text{-}\Omega$ resistor.

Solution:

a. $i(t) = (1.26\text{A}) \sin(200\pi t + \pi/2)$; b.
 $i(t) = (7.96\text{A}) \sin(200\pi t - \pi/2)$; c. $i(t) = (2\text{A}) \sin(200\pi t)$

Exercise:**Problem:**

A 700-pF capacitor is connected across an ac source with a voltage amplitude of 160 V and a frequency of 20 kHz. (a) Determine the capacitive reactance of the capacitor and the amplitude of the output current of the source. (b) If the frequency is changed to 60 Hz while keeping the voltage amplitude at 160 V, what are the capacitive reactance and the current amplitude?

Exercise:**Problem:**

A 20-mH inductor is connected across an AC source with a variable frequency and a constant-voltage amplitude of 9.0 V. (a) Determine the reactance of the circuit and the maximum current through the inductor when the frequency is set at 20 kHz. (b) Do the same calculations for a frequency of 60 Hz.

Solution:

a. $2.5 \times 10^3 \Omega$, $3.6 \times 10^{-3} \text{ A}$; b. 7.5Ω , 1.2A

Exercise:**Problem:**

A 30- μF capacitor is connected across a 60-Hz ac source whose voltage amplitude is 50 V. (a) What is the maximum charge on the capacitor? (b) What is the maximum current into the capacitor? (c) What is the phase relationship between the capacitor charge and the current in the circuit?

Exercise:

Problem:

A 7.0-mH inductor is connected across a 60-Hz ac source whose voltage amplitude is 50 V. (a) What is the maximum current through the inductor? (b) What is the phase relationship between the current through and the potential difference across the inductor?

Solution:

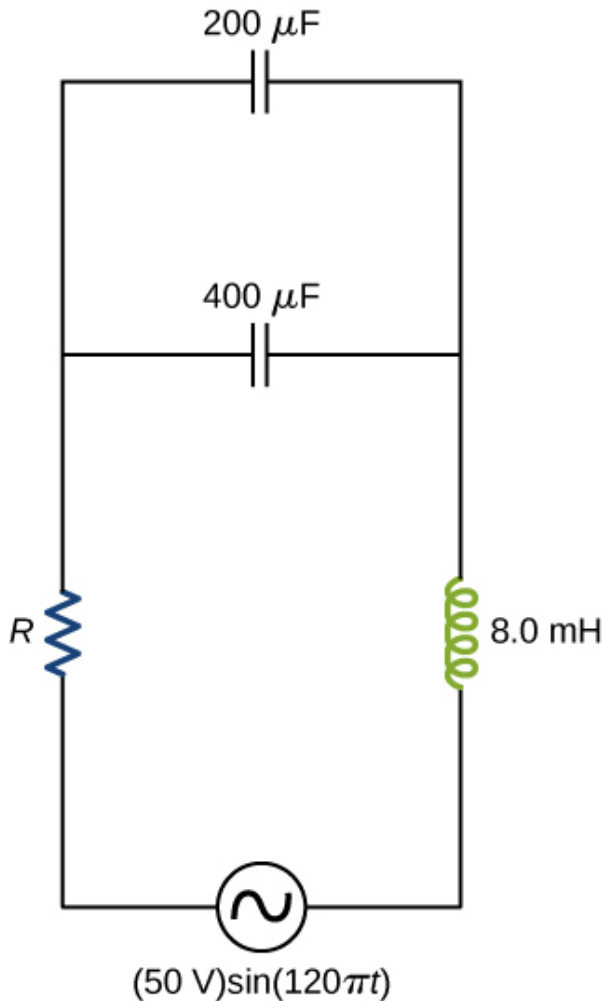
a. 19 A; b. inductor leads by 90°

Exercise:**Problem:**

What is the impedance of an *RLC* series circuit at the resonant frequency?

Exercise:**Problem:**

What is the resistance R in the circuit shown below if the amplitude of the ac through the inductor is 4.24 A?



Solution:

$11.7 \, \Omega$

Exercise:

Problem:

An ac source of voltage amplitude 100 V and frequency 1.0 kHz drives an RLC series circuit with $R = 20 \, \Omega$, $L = 4.0 \, \text{mH}$, and $C = 50 \, \mu\text{F}$. (a) Determine the rms current through the circuit. (b) What are the rms voltages across the three elements? (c) What is the phase angle between the emf and the current? (d) What is the power output of the source? (e) What is the power dissipated in the resistor?

Exercise:

Problem:

In an RLC series circuit, $R = 200\ \Omega$, $L = 1.0\ \text{H}$, $C = 50\ \mu\text{F}$, $V_0 = 120\ \text{V}$, and $f = 50\ \text{Hz}$. What is the power output of the source?

Solution:

14 W

Exercise:**Problem:**

A power plant generator produces 100 A at 15 kV (rms). A transformer is used to step up the transmission line voltage to 150 kV (rms). (a) What is rms current in the transmission line? (b) If the resistance per unit length of the line is $8.6 \times 10^{-8}\ \Omega/\text{m}$, what is the power loss per meter in the line? (c) What would the power loss per meter be if the line voltage were 15 kV (rms)?

Exercise:**Problem:**

Consider a power plant located 25 km outside a town delivering 50 MW of power to the town. The transmission lines are made of aluminum cables with a $7\ \text{cm}^2$ cross-sectional area. Find the loss of power in the transmission lines if it is transmitted at (a) 200 kV (rms) and (b) 120 V (rms).

Solution:

a. $5.9 \times 10^4\ \text{W}$; b. $1.64 \times 10^{11}\ \text{W}$

Exercise:

Problem:

Neon signs require 12-kV for their operation. A transformer is to be used to change the voltage from 220-V (rms) ac to 12-kV (rms) ac. What must the ratio be of turns in the secondary winding to the turns in the primary winding? (b) What is the maximum rms current the neon lamps can draw if the fuse in the primary winding goes off at 0.5 A? (c) How much power is used by the neon sign when it is drawing the maximum current allowed by the fuse in the primary winding?

Challenge Problems**Exercise:****Problem:**

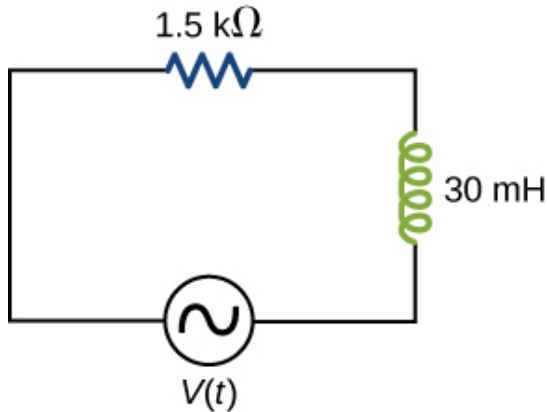
The 335-kV ac electricity from a power transmission line is fed into the primary winding of a transformer. The ratio of the number of turns in the secondary winding to the number in the primary winding is $N_s/N_p = 1000$. (a) What voltage is induced in the secondary winding? (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

a. 335 MV; b. the result is way too high, well beyond the breakdown voltage of air over reasonable distances; c. the input voltage is too high

Exercise:**Problem:**

A $1.5\text{-k}\Omega$ resistor and 30-mH inductor are connected in series, as shown below, across a 120-V (rms) ac power source oscillating at 60-Hz frequency. (a) Find the current in the circuit. (b) Find the voltage drops across the resistor and inductor. (c) Find the impedance of the circuit. (d) Find the power dissipated in the resistor. (e) Find the power dissipated in the inductor. (f) Find the power produced by the source.



Exercise:

Problem:

A $20\text{-}\Omega$ resistor, $50\text{-}\mu\text{F}$ capacitor, and 30-mH inductor are connected in series with an ac source of amplitude 10 V and frequency 125 Hz . (a) What is the impedance of the circuit? (b) What is the amplitude of the current in the circuit? (c) What is the phase constant of the current? Is it leading or lagging the source voltage? (d) Write voltage drops across the resistor, capacitor, and inductor and the source voltage as a function of time. (e) What is the power factor of the circuit? (f) How much energy is used by the resistor in 2.5 s ?

Solution:

a. $20\text{ }\Omega$; b. 0.5 A ; c. 5.4° , lagging;

d.

$$V_R = (9.96\text{ V})\cos(250\pi t + 5.4^\circ), V_C = (12.7\text{ V})\cos(250\pi t + 5.4^\circ - 90^\circ),$$

$$V_L = (11.8\text{ V})\cos(250\pi t + 5.4^\circ + 90^\circ), V_{\text{source}} = (10.0\text{ V})\cos(250\pi t);$$

e. 0.995 ; f. 6.25 J

Exercise:

Problem:

A $200\text{-}\Omega$ resistor, $150\text{-}\mu\text{F}$ capacitor, and 2.5-H inductor are connected in series with an ac source of amplitude 10 V and variable angular frequency ω .

(a) What is the value of the resonance frequency ω_R ? (b) What is the amplitude of the current if $\omega = \omega_R$? (c) What is the phase constant of the current when $\omega = \omega_R$? Is it leading or lagging the source voltage, or is it in phase? (d) Write an equation for the voltage drop across the resistor as a function of time when $\omega = \omega_R$. (e) What is the power factor of the circuit when $\omega = \omega_R$? (f) How much energy is used up by the resistor in 2.5 s when $\omega = \omega_R$?

Exercise:**Problem:**

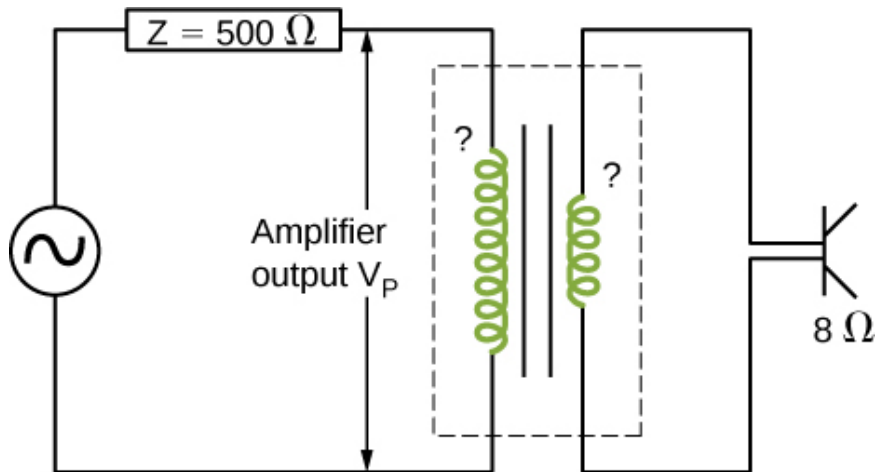
Find the reactances of the following capacitors and inductors in ac circuits with the given frequencies in each case: (a) 2-mH inductor with a frequency 60-Hz of the ac circuit; (b) 2-mH inductor with a frequency 600-Hz of the ac circuit; (c) 20-mH inductor with a frequency 6-Hz of the ac circuit; (d) 20-mH inductor with a frequency 60-Hz of the ac circuit; (e) 2-mF capacitor with a frequency 60-Hz of the ac circuit; and (f) 2-mF capacitor with a frequency 600-Hz of the AC circuit.

Solution:

a. $0.75\text{ }\Omega$; b. $7.5\text{ }\Omega$; c. $0.75\text{ }\Omega$; d. $7.5\text{ }\Omega$; e. $1.3\text{ }\Omega$; f. $0.13\text{ }\Omega$

Exercise:**Problem:**

An output impedance of an audio amplifier has an impedance of $500\text{ }\Omega$ and has a mismatch with a low-impedance $8\text{-}\Omega$ loudspeaker. You are asked to insert an appropriate transformer to match the impedances. What turns ratio will you use, and why? Use the simplified circuit shown below.



Exercise:

Problem:

Show that the SI unit for capacitive reactance is the ohm. Show that the SI unit for inductive reactance is also the ohm.

Solution:

The units as written for inductive reactance [\[link\]](#) are $\frac{\text{rad}}{\text{s}} \text{H}$. Radians can be ignored in unit analysis. The Henry can be defined as $\text{H} = \frac{\text{V} \cdot \text{s}}{\text{A}} = \Omega \cdot \text{s}$. Combining these together results in a unit of Ω for reactance.

Exercise:

Problem:

A coil with a self-inductance of 16 mH and a resistance of $6.0 \, \Omega$ is connected to an ac source whose frequency can be varied. At what frequency will the voltage across the coil lead the current through the coil by 45° ?

Exercise:

Problem:

An *RLC* series circuit consists of a $50\text{-}\Omega$ resistor, a $200\text{-}\mu\text{F}$ capacitor, and a 120-mH inductor whose coil has a resistance of $20 \, \Omega$. The source for the circuit has an rms emf of 240 V at a frequency of 60 Hz. Calculate the rms voltages across the (a) resistor, (b) capacitor, and (c) inductor.

Solution:

a. 156 V; b. 42 V; c. 154 V

Exercise:

Problem:

An RLC series circuit consists of a $10\text{-}\Omega$ resistor, an $8.0\text{-}\mu\text{F}$ capacitor, and a 50-mH inductor. A 110-V (rms) source of variable frequency is connected across the combination. What is the power output of the source when its frequency is set to one-half the resonant frequency of the circuit?

Exercise:

Problem:

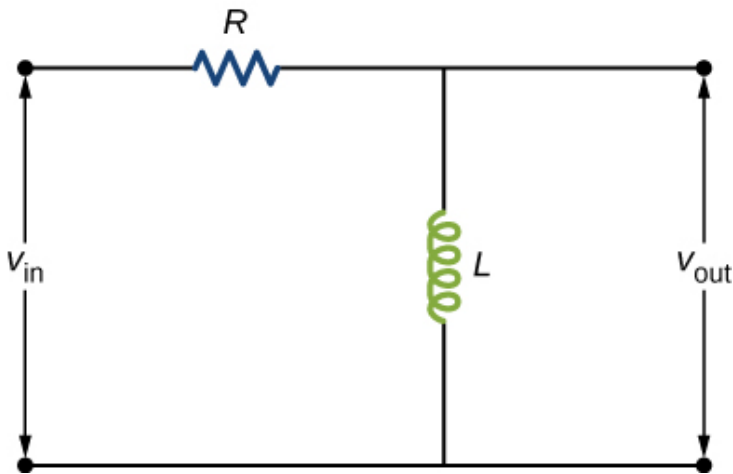
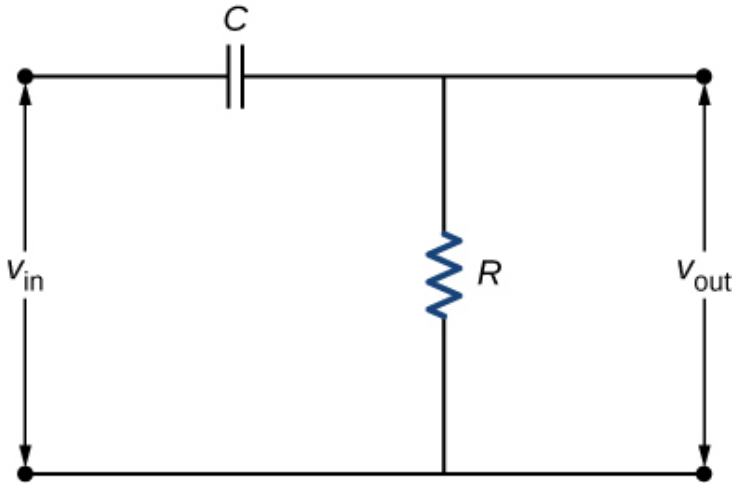
Shown below are two circuits that act as crude high-pass filters. The input voltage to the circuits is v_{in} , and the output voltage is v_{out} . (a) Show that for the capacitor circuit,

$$\frac{v_{\text{out}}}{v_{\text{in}}} = \frac{1}{\sqrt{1 + 1/\omega^2 R^2 C^2}},$$

and for the inductor circuit,

$$\frac{v_{\text{out}}}{v_{\text{in}}} = \frac{\omega L}{\sqrt{R^2 + \omega^2 L^2}}.$$

(b) Show that for high frequencies, $v_{\text{out}} \approx v_{\text{in}}$, but for low frequencies, $v_{\text{out}} \approx 0$.



Solution:

a. $\frac{v_{out}}{v_{in}} = \frac{1}{\sqrt{1 + \omega^2 R^2 C^2}}$ and $\frac{v_{out}}{v_{in}} = \frac{\omega L}{\sqrt{R^2 + \omega^2 L^2}}$; b. $v_{out} \approx v_{in}$ and $v_{out} \approx 0$

Exercise:

Problem:

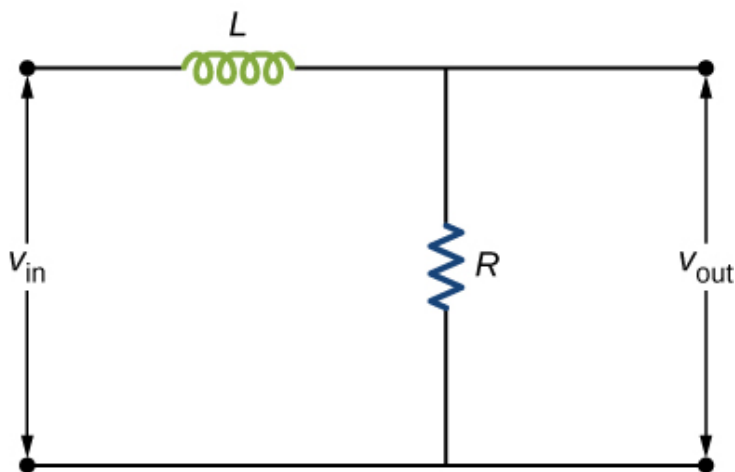
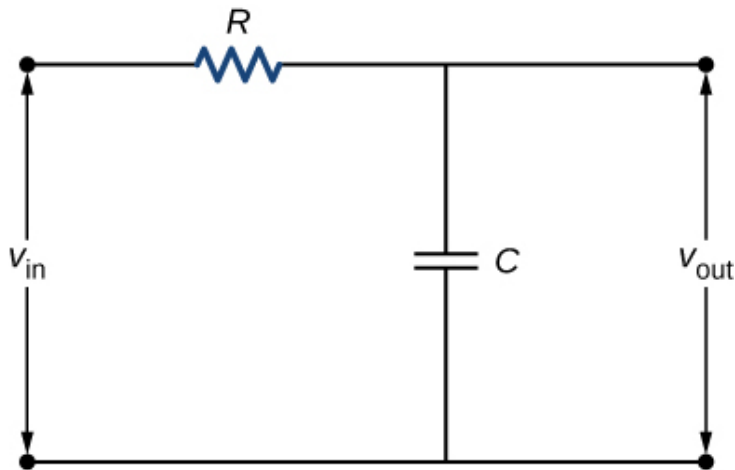
The two circuits shown below act as crude low-pass filters. The input voltage to the circuits is v_{in} , and the output voltage is v_{out} . (a) Show that for the capacitor circuit,

$$\frac{v_{out}}{v_{in}} = \frac{1}{\sqrt{1 + \omega^2 R^2 C^2}},$$

and for the inductor circuit,

$$\frac{v_{\text{out}}}{v_{\text{in}}} = \frac{R}{\sqrt{R^2 + \omega^2 L^2}}.$$

(b) Show that for low frequencies, $v_{\text{out}} \approx v_{\text{in}}$, but for high frequencies, $v_{\text{out}} \approx 0$.



Glossary

step-down transformer

transformer that decreases voltage and increases current

step-up transformer

transformer that increases voltage and decreases current

transformer

device that transforms voltages from one value to another using induction

transformer equation

equation showing that the ratio of the secondary to primary voltages in a transformer equals the ratio of the number of turns in their windings

Introduction

class="introduction"

The
pressure
from
sunlight
predicted by
Maxwell's
equations
helped
produce the
tail of
Comet
McNaught.
(credit:
modificatio
n of work
by Sebastian
Deiries—
ESO)



Our view of objects in the sky at night, the warm radiance of sunshine, the sting of sunburn, our cell phone conversations, and the X-rays revealing a broken bone—all are brought to us by electromagnetic waves. It would be hard to overstate the practical importance of electromagnetic waves, through their role in vision, through countless technological applications, and through their ability to transport the energy from the Sun through space to sustain life and almost all of its activities on Earth.

Theory predicted the general phenomenon of electromagnetic waves before anyone realized that light is a form of an electromagnetic wave. In the mid-nineteenth century, James Clerk Maxwell formulated a single theory combining all the electric and magnetic effects known at that time. Maxwell's equations, summarizing this theory, predicted the existence of electromagnetic waves that travel at the speed of light. His theory also predicted how these waves behave, and how they carry both energy and momentum. The tails of comets, such as Comet McNaught in [\[link\]](#), provide a spectacular example. Energy carried by light from the Sun warms the comet to release dust and gas. The momentum carried by the light exerts a weak force that shapes the dust into a tail of the kind seen here. The flux of particles emitted by the Sun, called the solar wind, typically produces an additional, second tail, as described in detail in this chapter.

In this chapter, we explain Maxwell's theory and show how it leads to his prediction of electromagnetic waves. We use his theory to examine what electromagnetic waves are, how they are produced, and how they transport energy and momentum. We conclude by summarizing some of the many practical applications of electromagnetic waves.

Maxwell's Equations and Electromagnetic Waves

By the end of this section, you will be able to:

- Explain Maxwell's correction of Ampère's law by including the displacement current
- State and apply Maxwell's equations in integral form
- Describe how the symmetry between changing electric and changing magnetic fields explains Maxwell's prediction of electromagnetic waves
- Describe how Hertz confirmed Maxwell's prediction of electromagnetic waves

James Clerk Maxwell (1831–1879) was one of the major contributors to physics in the nineteenth century ([\[link\]](#)). Although he died young, he made major contributions to the development of the kinetic theory of gases, to the understanding of color vision, and to the nature of Saturn's rings. He is probably best known for having combined existing knowledge of the laws of electricity and of magnetism with insights of his own into a complete overarching electromagnetic theory, represented by **Maxwell's equations**.



James Clerk Maxwell, a nineteenth-century physicist, developed a theory that explained the relationship between electricity and magnetism, and correctly predicted that visible light consists of electromagnetic waves.

Maxwell's Correction to the Laws of Electricity and Magnetism

The four basic laws of electricity and magnetism had been discovered experimentally through the work of physicists such as Oersted, Coulomb, Gauss, and Faraday. Maxwell discovered logical inconsistencies in these

earlier results and identified the incompleteness of Ampère's law as their cause.

Recall that according to Ampère's law, the integral of the magnetic field around a closed loop C is proportional to the current I passing through any surface whose boundary is loop C itself:

Equation:

$$\oint_C \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 I.$$

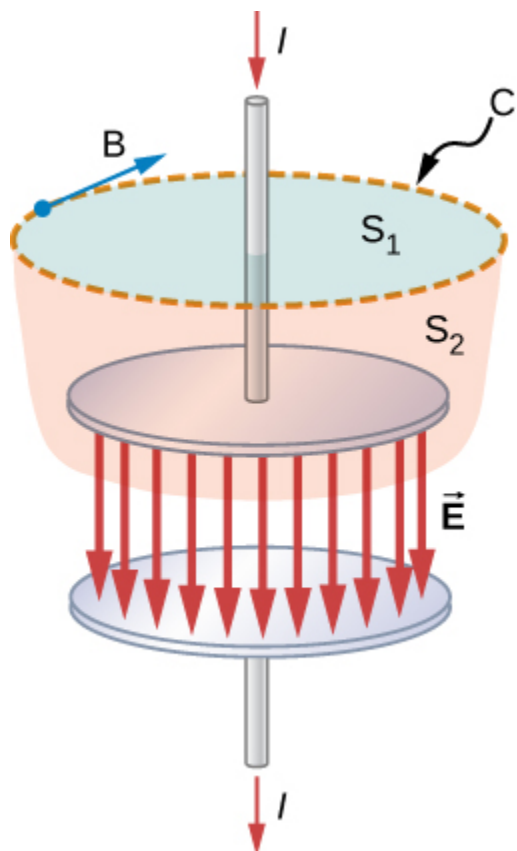
There are infinitely many surfaces that can be attached to any loop, and Ampère's law stated in [\[link\]](#) is independent of the choice of surface.

Consider the set-up in [\[link\]](#). A source of emf is abruptly connected across a parallel-plate capacitor so that a time-dependent current I develops in the wire. Suppose we apply Ampère's law to loop C shown at a time before the capacitor is fully charged, so that $I \neq 0$. Surface S_1 gives a nonzero value for the enclosed current I , whereas surface S_2 gives zero for the enclosed current because no current passes through it:

Equation:

$$\oint_C \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \begin{cases} \mu_0 I & \text{if surface } S_1 \text{ is used} \\ 0 & \text{if surface } S_2 \text{ is used} \end{cases}.$$

Clearly, Ampère's law in its usual form does not work here. This may not be surprising, because Ampère's law as applied in earlier chapters required a steady current, whereas the current in this experiment is changing with time and is not steady at all.



The currents through surface S_1 and surface S_2 are unequal, despite having the same boundary loop C .

How can Ampère's law be modified so that it works in all situations? Maxwell suggested including an additional contribution, called the displacement current I_d , to the real current I ,

Equation:

$$\oint_C \vec{B} \cdot d\vec{s} = \mu_0 (I + I_d)$$

where the displacement current is defined to be

Note:

Equation:

$$I_d = \varepsilon_0 \frac{d\Phi_E}{dt}.$$

Here ε_0 is the permittivity of free space and Φ_E is the electric flux, defined as

Equation:

$$\Phi_E = \iint_{\text{Surface } S} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}.$$

The **displacement current** is analogous to a real current in Ampère's law, entering into Ampère's law in the same way. It is produced, however, by a changing electric field. It accounts for a changing electric field producing a magnetic field, just as a real current does, but the displacement current can produce a magnetic field even where no real current is present. When this extra term is included, the modified Ampère's law equation becomes

Equation:

$$\oint_C \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 I + \varepsilon_0 \mu_0 \frac{d\Phi_E}{dt}$$

and is independent of the surface S through which the current I is measured.

We can now examine this modified version of Ampère's law to confirm that it holds independent of whether the surface S_1 or the surface S_2 in [\[link\]](#) is chosen. The electric field $\vec{\mathbf{E}}$ corresponding to the flux Φ_E in [\[link\]](#) is between the capacitor plates. Therefore, the $\vec{\mathbf{E}}$ field and the displacement current through the surface S_1 are both zero, and [\[link\]](#) takes the form

Equation:

$$\oint_C \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 I.$$

We must now show that for surface S_2 , through which no actual current flows, the displacement current leads to the same value $\mu_0 I$ for the right side of the Ampère's law equation. For surface S_2 , the equation becomes
Equation:

$$\oint_C \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 \frac{d}{dt} \left[\varepsilon_0 \iint_{\text{Surface } S_2} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} \right].$$

Gauss's law for electric charge requires a closed surface and cannot ordinarily be applied to a surface like S_1 alone or S_2 alone. But the two surfaces S_1 and S_2 form a closed surface in [\[link\]](#) and can be used in Gauss's law. Because the electric field is zero on S_1 , the flux contribution through S_1 is zero. This gives us

Equation:

$$\begin{aligned} \oint_{\text{Surface } S_1 + S_2} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} &= \iint_{\text{Surface } S_1} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} + \iint_{\text{Surface } S_2} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} \\ &= 0 + \iint_{\text{Surface } S_2} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} \\ &= \iint_{\text{Surface } S_2} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}. \end{aligned}$$

Therefore, we can replace the integral over S_2 in [\[link\]](#) with the closed Gaussian surface $S_1 + S_2$ and apply Gauss's law to obtain

Equation:

$$\oint_{S_1} \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 \frac{dQ_{\text{in}}}{dt} = \mu_0 I.$$

Thus, the modified Ampère's law equation is the same using surface S_2 , where the right-hand side results from the displacement current, as it is for the surface S_1 , where the contribution comes from the actual flow of electric charge.

Example:**Displacement current in a charging capacitor**

A parallel-plate capacitor with capacitance C whose plates have area A and separation distance d is connected to a resistor R and a battery of voltage V . The current starts to flow at $t = 0$. (a) Find the displacement current between the capacitor plates at time t . (b) From the properties of the capacitor, find the corresponding real current $I = \frac{dQ}{dt}$, and compare the answer to the expected current in the wires of the corresponding RC circuit.

Strategy

We can use the equations from the analysis of an RC circuit ([Alternating-Current Circuits](#)) plus Maxwell's version of Ampère's law.

Solution

- a. The voltage between the plates at time t is given by

Equation:

$$V_C = \frac{1}{C} Q(t) = V_0 \left(1 - e^{-t/RC}\right).$$

Let the z -axis point from the positive plate to the negative plate. Then the z -component of the electric field between the plates as a function of time t is

Equation:

$$E_z(t) = \frac{V_0}{d} \left(1 - e^{-t/RC}\right).$$

Therefore, the z -component of the displacement current I_d between the plates is

Equation:

$$I_d(t) = \varepsilon_0 A \frac{\partial E_z(t)}{\partial t} = \varepsilon_0 A \frac{V_0}{d} \times \frac{1}{RC} e^{-t/RC} = \frac{V_0}{R} e^{-t/RC},$$

where we have used $C = \varepsilon_0 \frac{A}{d}$ for the capacitance.

b. From the expression for V_C , the charge on the capacitor is

Equation:

$$Q(t) = CV_C = CV_0 (1 - e^{-t/RC}).$$

The current into the capacitor after the circuit is closed, is therefore

Equation:

$$I = \frac{dQ}{dt} = \frac{V_0}{R} e^{-t/RC}.$$

This current is the same as I_d found in (a).

Maxwell's Equations

With the correction for the displacement current, Maxwell's equations take the form

Note:

Equation:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q_{\text{in}}}{\varepsilon_0} \quad \left(\text{Gauss's law} \right)$$

Equation:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0 \quad \left(\text{Gauss's law for magnetism} \right)$$

Equation:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -\frac{d\Phi_{\text{m}}}{dt} \quad \left(\text{Faraday's law} \right)$$

Equation:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 I + \varepsilon_0 \mu_0 \frac{d\Phi_{\text{E}}}{dt} \quad \left(\text{Ampère-Maxwell law} \right).$$

Once the fields have been calculated using these four equations, the Lorentz force equation

Equation:

$$\vec{\mathbf{F}} = q\vec{\mathbf{E}} + q\vec{\mathbf{v}} \times \vec{\mathbf{B}}$$

gives the force that the fields exert on a particle with charge q moving with velocity $\vec{\mathbf{v}}$. The Lorentz force equation combines the force of the electric field and of the magnetic field on the moving charge. The magnetic and electric forces have been examined in earlier modules. These four Maxwell's equations are, respectively,

Note:

Maxwell's Equations

1. Gauss's law

The electric flux through any closed surface is equal to the electric charge Q_{in} enclosed by the surface. Gauss's law [\[link\]](#) describes the relation between an electric charge and the electric field it produces. This is often pictured in terms of electric field lines originating from positive charges and terminating on negative charges, and indicating the direction of the electric field at each point in space.

2. Gauss's law for magnetism

The magnetic field flux through any closed surface is zero [\[link\]](#). This is equivalent to the statement that magnetic field lines are continuous, having no beginning or end. Any magnetic field line entering the region enclosed by the surface must also leave it. No magnetic monopoles, where magnetic field lines would terminate, are known to exist (see [Magnetic Fields and Lines](#)).

3. Faraday's law

A changing magnetic field induces an electromotive force (emf) and, hence, an electric field. The direction of the emf opposes the change. This third of Maxwell's equations, [\[link\]](#), is Faraday's law of induction and includes Lenz's law. The electric field from a changing magnetic field has field lines that form closed loops, without any beginning or end.

4. Ampère-Maxwell law

Magnetic fields are generated by moving charges or by changing electric fields. This fourth of Maxwell's equations, [\[link\]](#), encompasses Ampère's law and adds another source of magnetic fields, namely changing electric fields.

Maxwell's equations and the Lorentz force law together encompass all the laws of electricity and magnetism. The symmetry that Maxwell introduced into his mathematical framework may not be immediately apparent.

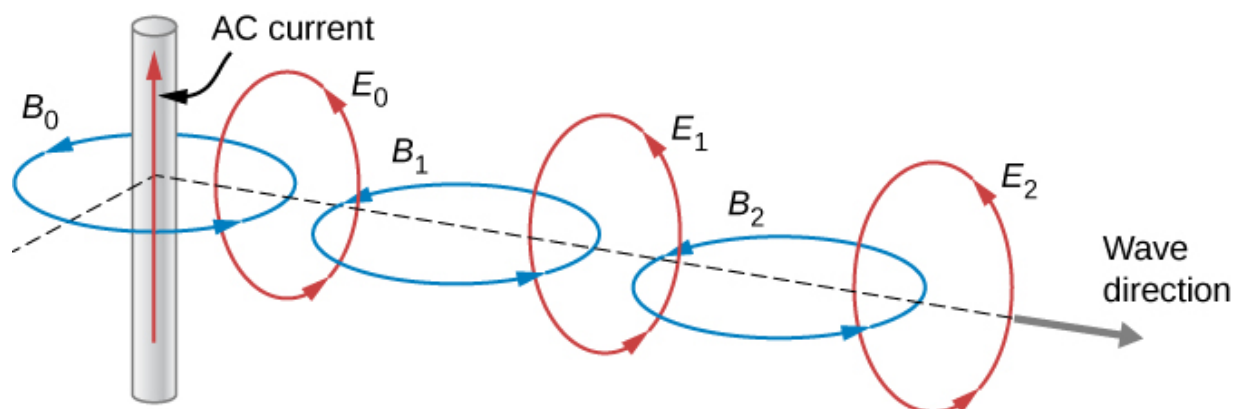
Faraday's law describes how changing magnetic fields produce electric fields. The displacement current introduced by Maxwell results instead from a changing electric field and accounts for a changing electric field producing a magnetic field. The equations for the effects of both changing electric fields and changing magnetic fields differ in form only where the absence of magnetic monopoles leads to missing terms. This symmetry between the effects of changing magnetic and electric fields is essential in explaining the nature of electromagnetic waves.

Later application of Einstein's theory of relativity to Maxwell's complete and symmetric theory showed that electric and magnetic forces are not separate but are different manifestations of the same thing—the electromagnetic force. The electromagnetic force and weak nuclear force are similarly unified as the electroweak force. This unification of forces has

been one motivation for attempts to unify all of the four basic forces in nature—the gravitational, electrical, strong, and weak nuclear forces (see [Particle Physics and Cosmology](#)).

The Mechanism of Electromagnetic Wave Propagation

To see how the symmetry introduced by Maxwell accounts for the existence of combined electric and magnetic waves that propagate through space, imagine a time-varying magnetic field $\vec{B}_0(t)$ produced by the high-frequency alternating current seen in [\[link\]](#). We represent $\vec{B}_0(t)$ in the diagram by one of its field lines. From Faraday's law, the changing magnetic field through a surface induces a time-varying electric field $\vec{E}_0(t)$ at the boundary of that surface. The displacement current source for the electric field, like the Faraday's law source for the magnetic field, produces only closed loops of field lines, because of the mathematical symmetry involved in the equations for the induced electric and induced magnetic fields. A field line representation of $\vec{E}_0(t)$ is shown. In turn, the changing electric field $\vec{E}_0(t)$ creates a magnetic field $\vec{B}_1(t)$ according to the modified Ampère's law. This changing field induces $\vec{E}_1(t)$, which induces $\vec{B}_2(t)$, and so on. We then have a self-continuing process that leads to the creation of time-varying electric and magnetic fields in regions farther and farther away from O . This process may be visualized as the propagation of an electromagnetic wave through space.



How changing \vec{E} and \vec{B} fields propagate through space.

In the next section, we show in more precise mathematical terms how Maxwell's equations lead to the prediction of electromagnetic waves that can travel through space without a material medium, implying a speed of electromagnetic waves equal to the speed of light.

Prior to Maxwell's work, experiments had already indicated that light was a wave phenomenon, although the nature of the waves was yet unknown. In 1801, Thomas Young (1773–1829) showed that when a light beam was separated by two narrow slits and then recombined, a pattern made up of bright and dark fringes was formed on a screen. Young explained this behavior by assuming that light was composed of waves that added constructively at some points and destructively at others (see [Interference](#)). Subsequently, Jean Foucault (1819–1868), with measurements of the speed of light in various media, and Augustin Fresnel (1788–1827), with detailed experiments involving interference and diffraction of light, provided further conclusive evidence that light was a wave. So, light was known to be a wave, and Maxwell had predicted the existence of electromagnetic waves that traveled at the speed of light. The conclusion seemed inescapable: Light must be a form of electromagnetic radiation. But Maxwell's theory showed that other wavelengths and frequencies than those of light were possible for electromagnetic waves. He showed that electromagnetic radiation with the same fundamental properties as visible light should exist at any frequency. It remained for others to test, and confirm, this prediction.

Note:

Exercise:

Problem:

Check Your Understanding When the emf across a capacitor is turned on and the capacitor is allowed to charge, when does the magnetic field induced by the displacement current have the greatest magnitude?

Solution:

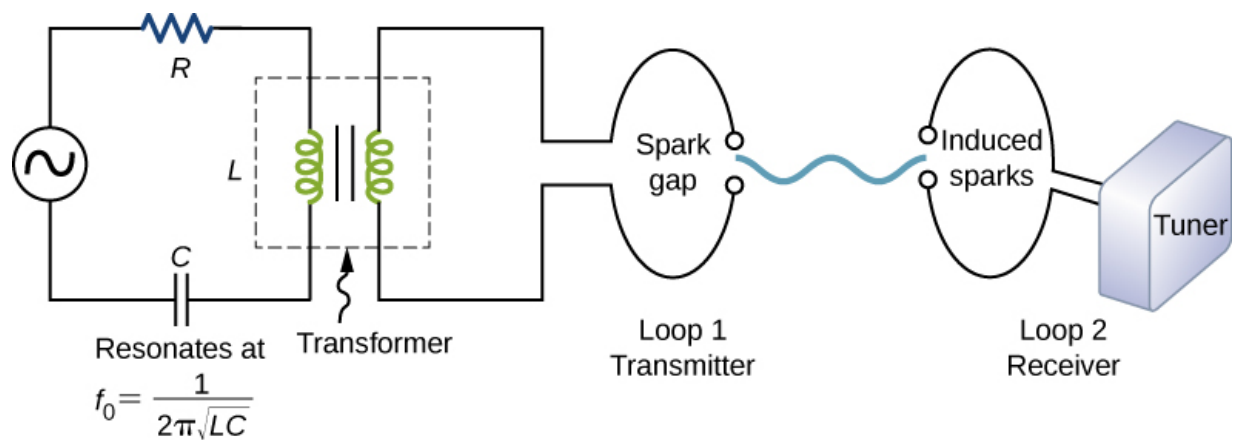
It is greatest immediately after the current is switched on. The displacement current and the magnetic field from it are proportional to the rate of change of electric field between the plates, which is greatest when the plates first begin to charge.

Hertz's Observations

The German physicist Heinrich Hertz (1857–1894) was the first to generate and detect certain types of electromagnetic waves in the laboratory. Starting in 1887, he performed a series of experiments that not only confirmed the existence of electromagnetic waves but also verified that they travel at the speed of light.

Hertz used an alternating-current RLC (resistor-inductor-capacitor) circuit that resonates at a known frequency $f_0 = \frac{1}{2\pi\sqrt{LC}}$ and connected it to a loop of wire, as shown in [\[link\]](#). High voltages induced across the gap in the loop produced sparks that were visible evidence of the current in the circuit and helped generate electromagnetic waves.

Across the laboratory, Hertz placed another loop attached to another RLC circuit, which could be tuned (as the dial on a radio) to the same resonant frequency as the first and could thus be made to receive electromagnetic waves. This loop also had a gap across which sparks were generated, giving solid evidence that electromagnetic waves had been received.



The apparatus used by Hertz in 1887 to generate and detect electromagnetic waves.

Hertz also studied the reflection, refraction, and interference patterns of the electromagnetic waves he generated, confirming their wave character. He was able to determine the wavelengths from the interference patterns, and knowing their frequencies, he could calculate the propagation speed using the equation $v = f\lambda$, where v is the speed of a wave, f is its frequency, and λ is its wavelength. Hertz was thus able to prove that electromagnetic waves travel at the speed of light. The SI unit for frequency, the hertz (1 Hz = 1 cycle/s), is named in his honor.

Note:

Exercise:

Problem:

Check Your Understanding Could a purely electric field propagate as a wave through a vacuum without a magnetic field? Justify your answer.

Solution:

No. The changing electric field according to the modified version of Ampère's law would necessarily induce a changing magnetic field.

Summary

- Maxwell's prediction of electromagnetic waves resulted from his formulation of a complete and symmetric theory of electricity and magnetism, known as Maxwell's equations.
- The four Maxwell's equations together with the Lorentz force law encompass the major laws of electricity and magnetism. The first of these is Gauss's law for electricity; the second is Gauss's law for magnetism; the third is Faraday's law of induction (including Lenz's law); and the fourth is Ampère's law in a symmetric formulation that adds another source of magnetism, namely changing electric fields.
- The symmetry introduced between electric and magnetic fields through Maxwell's displacement current explains the mechanism of electromagnetic wave propagation, in which changing magnetic fields produce changing electric fields and vice versa.
- Although light was already known to be a wave, the nature of the wave was not understood before Maxwell. Maxwell's equations also predicted electromagnetic waves with wavelengths and frequencies outside the range of light. These theoretical predictions were first confirmed experimentally by Heinrich Hertz.

Conceptual Questions

Exercise:

Problem:

Explain how the displacement current maintains the continuity of current in a circuit containing a capacitor.

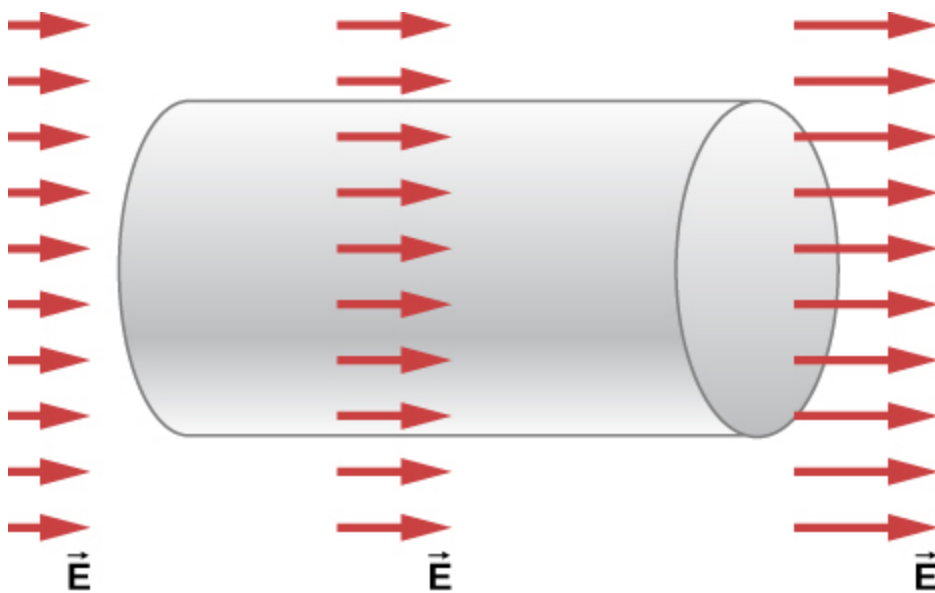
Solution:

The current into the capacitor to change the electric field between the plates is equal to the displacement current between the plates.

Exercise:

Problem:

Describe the field lines of the induced magnetic field along the edge of the imaginary horizontal cylinder shown below if the cylinder is in a spatially uniform electric field that is horizontal, pointing to the right, and increasing in magnitude.



Exercise:

Problem:

Why is it much easier to demonstrate in a student lab that a changing magnetic field induces an electric field than it is to demonstrate that a changing electric field produces a magnetic field?

Solution:

The first demonstration requires simply observing the current produced in a wire that experiences a changing magnetic field. The second demonstration requires moving electric charge from one

location to another, and therefore involves electric currents that generate a changing electric field. The magnetic fields from these currents are not easily separated from the magnetic field that the displacement current produces.

Problems

Exercise:

Problem:

Show that the magnetic field at a distance r from the axis of two circular parallel plates, produced by placing charge $Q(t)$ on the plates is

$$B_{\text{ind}} = \frac{\mu_0}{2\pi r} \frac{dQ(t)}{dt}.$$

Solution:

$$\begin{aligned} B_{\text{ind}} &= \frac{\mu_0}{2\pi r} I_{\text{ind}} = \frac{\mu_0}{2\pi r} \varepsilon_0 \frac{\partial \Phi_E}{\partial t} = \frac{\mu_0}{2\pi r} \varepsilon_0 \left(A \frac{\partial E}{\partial t} \right) = \frac{\mu_0}{2\pi r} \varepsilon_0 A \left(\frac{1}{d} \frac{dV(t)}{dt} \right) \\ &= \frac{\mu_0}{2\pi r} \left[\frac{\varepsilon_0 A}{d} \right] \left[\frac{1}{C} \frac{dQ(t)}{dt} \right] = \frac{\mu_0}{2\pi r} \frac{dQ(t)}{dt} \quad \text{because } C = \frac{\varepsilon_0 A}{d} \end{aligned}$$

Exercise:

Problem:

Express the displacement current in a capacitor in terms of the capacitance and the rate of change of the voltage across the capacitor.

Exercise:

Problem:

A potential difference $V(t) = V_0 \sin \omega t$ is maintained across a parallel-plate capacitor with capacitance C consisting of two circular parallel plates. A thin wire with resistance R connects the centers of the two plates, allowing charge to leak between plates while they are charging.

- (a) Obtain expressions for the leakage current $I_{\text{res}}(t)$ in the thin wire. Use these results to obtain an expression for the current $I_{\text{real}}(t)$ in the wires connected to the capacitor.
 - (b) Find the displacement current in the space between the plates from the changing electric field between the plates.
 - (c) Compare $I_{\text{real}}(t)$ with the sum of the displacement current $I_d(t)$ and resistor current $I_{\text{res}}(t)$ between the plates, and explain why the relationship you observe would be expected.
-

Solution:

a. $I_{\text{res}} = \frac{V_0 \sin \omega t}{R}$; b. $I_d = CV_0 \omega \cos \omega t$;

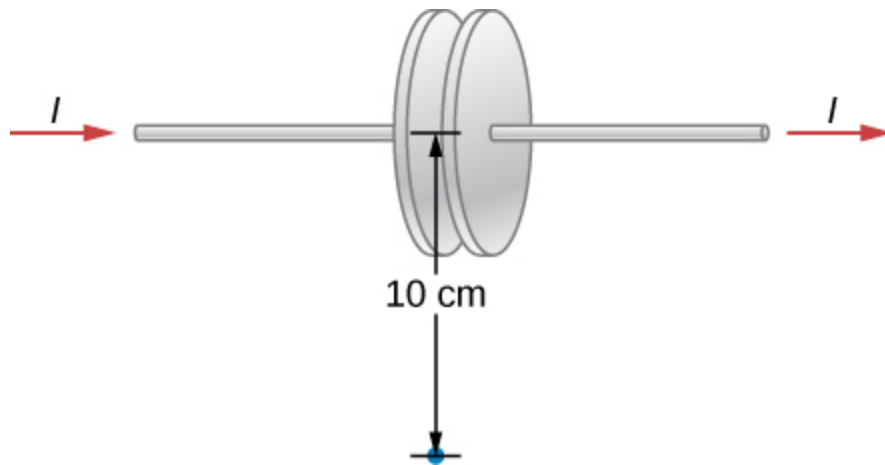
c.

$$I_{\text{real}} = I_{\text{res}} + \frac{dQ}{dt} = \frac{V_0 \sin \omega t}{R} + CV_0 \frac{d}{dt} \sin \omega t = \frac{V_0 \sin \omega t}{R} + CV_0 \omega \cos \omega t$$

; which is the sum of I_{res} and I_{real} , consistent with how the displacement current maintaining the continuity of current.

Exercise:**Problem:**

Suppose the parallel-plate capacitor shown below is accumulating charge at a rate of 0.010 C/s. What is the induced magnetic field at a distance of 10 cm from the capacitor?



Exercise:

Problem:

The potential difference $V(t)$ between parallel plates shown above is instantaneously increasing at a rate of 10^7 V/s . What is the displacement current between the plates if the separation of the plates is 1.00 cm and they have an area of 0.200 m^2 ?

Solution:

$$1.77 \times 10^{-3} \text{ A}$$

Exercise:

Problem:

A parallel-plate capacitor has a plate area of $A = 0.250 \text{ m}^2$ and a separation of 0.0100 m . What must be the angular frequency ω for a voltage $V(t) = V_0 \sin \omega t$ with $V_0 = 100 \text{ V}$ to produce a maximum displacement induced current of 1.00 A between the plates?

Exercise:

Problem:

The voltage across a parallel-plate capacitor with area $A = 800 \text{ cm}^2$ and separation $d = 2 \text{ mm}$ varies sinusoidally as $V = (15 \text{ mV}) \cos (150t)$, where t is in seconds. Find the displacement current between the plates.

Solution:

$$I_d = (7.97 \times 10^{-10} \text{ A}) \sin (150 t)$$

Exercise:**Problem:**

The voltage across a parallel-plate capacitor with area A and separation d varies with time t as $V = at^2$, where a is a constant. Find the displacement current between the plates.

Glossary

displacement current

extra term in Maxwell's equations that is analogous to a real current but accounts for a changing electric field producing a magnetic field, even when the real current is present

Maxwell's equations

set of four equations that comprise a complete, overarching theory of electromagnetism

Plane Electromagnetic Waves

By the end of this section, you will be able to:

- Describe how Maxwell's equations predict the relative directions of the electric fields and magnetic fields, and the direction of propagation of plane electromagnetic waves
- Explain how Maxwell's equations predict that the speed of propagation of electromagnetic waves in free space is exactly the speed of light
- Calculate the relative magnitude of the electric and magnetic fields in an electromagnetic plane wave
- Describe how electromagnetic waves are produced and detected

Mechanical waves travel through a medium such as a string, water, or air. Perhaps the most significant prediction of Maxwell's equations is the existence of combined electric and magnetic (or electromagnetic) fields that propagate through space as electromagnetic waves. Because Maxwell's equations hold in free space, the predicted electromagnetic waves, unlike mechanical waves, do not require a medium for their propagation.

A general treatment of the physics of electromagnetic waves is beyond the scope of this textbook. We can, however, investigate the special case of an electromagnetic wave that propagates through free space along the x -axis of a given coordinate system.

Electromagnetic Waves in One Direction

An electromagnetic wave consists of an electric field, defined as usual in terms of the force per charge on a stationary charge, and a magnetic field, defined in terms of the force per charge on a moving charge. The electromagnetic field is assumed to be a function of only the x -coordinate and time. The y -component of the electric field is then written as $E_y(x, t)$, the z -component of the magnetic field as $B_z(x, t)$, etc. Because we are assuming free space, there are no free charges or currents, so we can set $Q_{\text{in}} = 0$ and $I = 0$ in Maxwell's equations.

The transverse nature of electromagnetic waves

We examine first what Gauss's law for electric fields implies about the relative directions of the electric field and the propagation direction in an electromagnetic wave. Assume the Gaussian surface to be the surface of a rectangular box whose cross-section is a square of side l and whose third side has length Δx , as shown in [\[link\]](#). Because the electric field is a function only of x and t , the y -component of the electric field is the same on both the top (labeled Side 2) and bottom (labeled Side 1) of the box, so that these two contributions to the flux cancel. The corresponding argument also holds for the net flux from the z -component of the electric field through Sides 3 and 4. Any net flux through the surface therefore comes entirely from the x -component of the electric field. Because the electric field has no y - or z -dependence, $E_x(x, t)$ is constant over the face of the box with area A and has a possibly different value $E_x(x + \Delta x, t)$ that is constant over the opposite face of the box. Applying Gauss's law gives

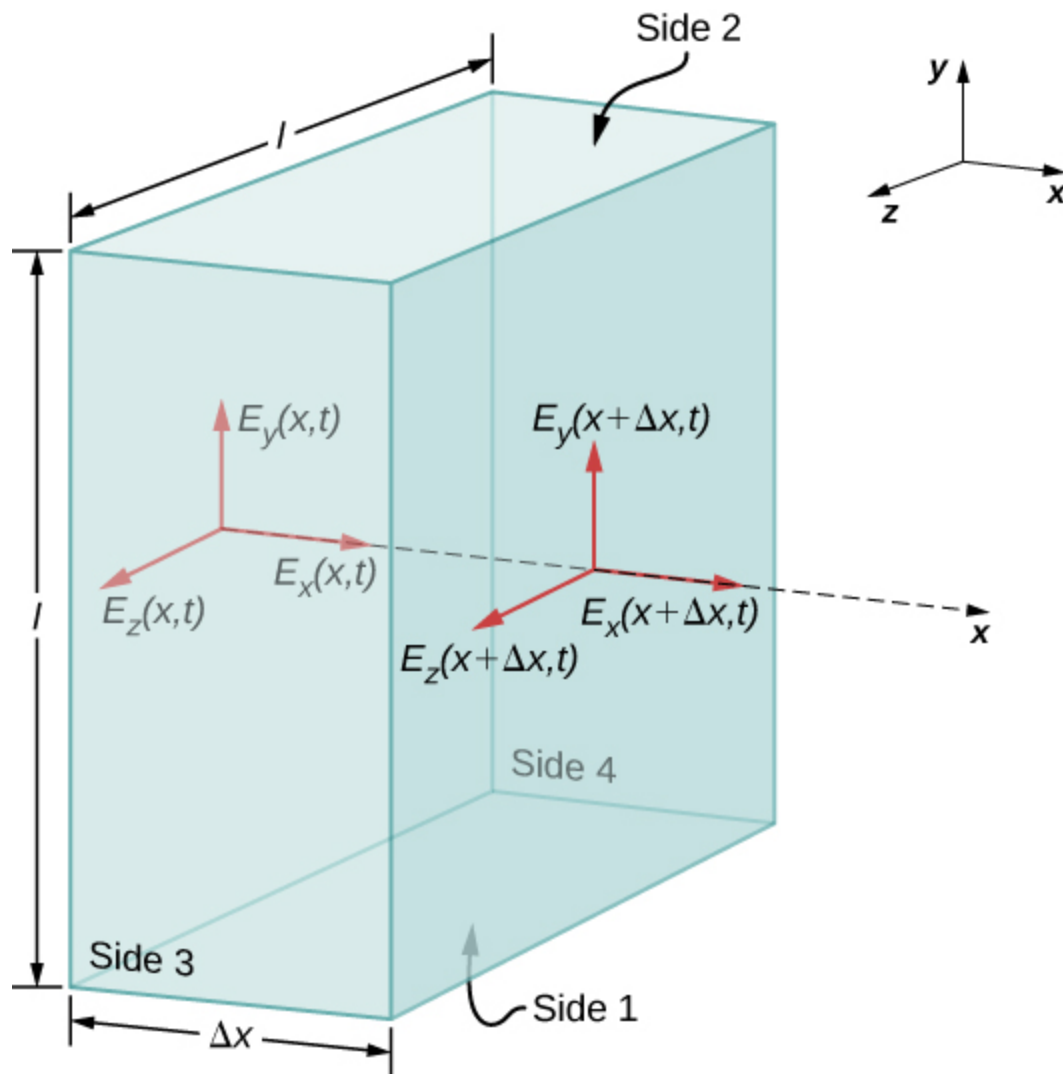
Equation:

$$\text{Net flux} = -E_x(x, t)A + E_x(x + \Delta x, t)A = \frac{Q_{\text{in}}}{\epsilon_0}$$

where $A = l \times l$ is the area of the front and back faces of the rectangular surface. But the charge enclosed is $Q_{\text{in}} = 0$, so this component's net flux is also zero, and [\[link\]](#) implies $E_x(x, t) = E_x(x + \Delta x, t)$ for any Δx .

Therefore, if there is an x -component of the electric field, it cannot vary with x . A uniform field of that kind would merely be superposed artificially on the traveling wave, for example, by having a pair of parallel-charged plates. Such a component $E_x(x, t)$ would not be part of an electromagnetic wave propagating along the x -axis; so $E_x(x, t) = 0$ for this wave.

Therefore, the only nonzero components of the electric field are $E_y(x, t)$ and $E_z(x, t)$, perpendicular to the direction of propagation of the wave.



The surface of a rectangular box of dimensions $l \times l \times \Delta x$ is our Gaussian surface. The electric field shown is from an electromagnetic wave propagating along the x-axis.

A similar argument holds by substituting E for B and using Gauss's law for magnetism instead of Gauss's law for electric fields. This shows that the B field is also perpendicular to the direction of propagation of the wave. The electromagnetic wave is therefore a transverse wave, with its oscillating electric and magnetic fields perpendicular to its direction of propagation.

The speed of propagation of electromagnetic waves

We can next apply Maxwell's equations to the description given in connection with [\[link\]](#) in the previous section to obtain an equation for the E field from the changing B field, and for the B field from a changing E field. We then combine the two equations to show how the changing E and B fields propagate through space at a speed precisely equal to the speed of light.

First, we apply Faraday's law over Side 3 of the Gaussian surface, using the path shown in [\[link\]](#). Because $E_x(x, t) = 0$, we have

Equation:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -E_y(x, t)l + E_y(x + \Delta x, t)l.$$

Assuming Δx is small and approximating $E_y(x + \Delta x, t)$ by

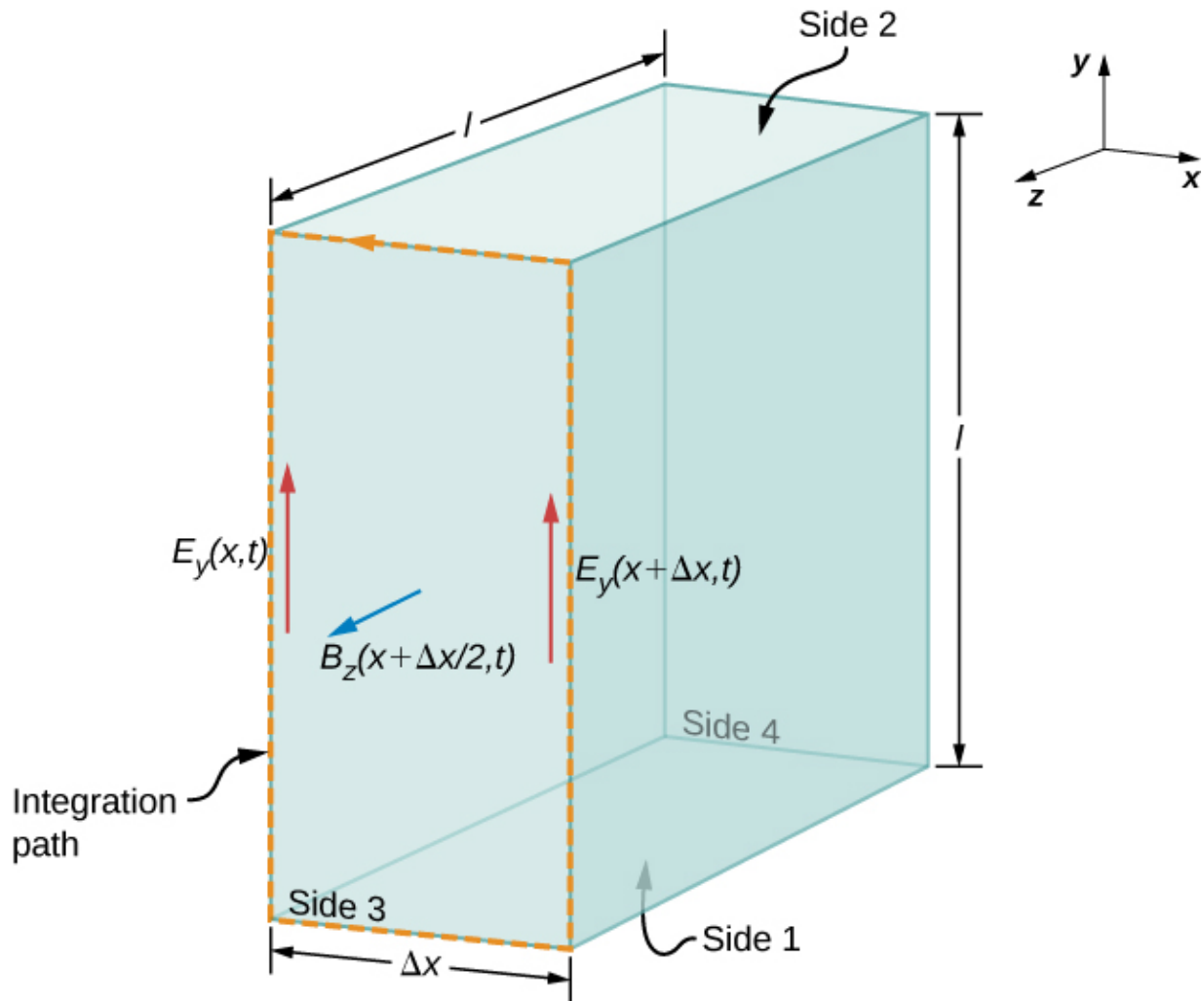
Equation:

$$E_y(x + \Delta x, t) = E_y(x, t) + \frac{\partial E_y(x, t)}{\partial x} \Delta x,$$

we obtain

Equation:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = \frac{\partial E_y(x, t)}{\partial x} (l\Delta x).$$



We apply Faraday's law to the front of the rectangle by evaluating $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}}$ along the rectangular edge of Side 3 in the direction indicated, taking the B field crossing the face to be approximately its value in the middle of the area traversed.

Because Δx is small, the magnetic flux through the face can be approximated by its value in the center of the area traversed, namely $B_z \left(x + \frac{\Delta x}{2}, t \right)$. The flux of the B field through Face 3 is then the B field times the area,

Equation:

$$\oint_S \vec{\mathbf{B}} \cdot \vec{\mathbf{n}} dA = B_z \left(x + \frac{\Delta x}{2}, t \right) (l\Delta x).$$

From Faraday's law,

Equation:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -\frac{d}{dt} \int_S \vec{\mathbf{B}} \cdot \vec{\mathbf{n}} dA.$$

Therefore, from [\[link\]](#) and [\[link\]](#),

Equation:

$$\frac{\partial E_y(x, t)}{\partial x} (l\Delta x) = -\frac{\partial}{\partial t} \left[B_z \left(x + \frac{\Delta x}{2}, t \right) \right] (l\Delta x).$$

Canceling $l\Delta x$ and taking the limit as $\Delta x = 0$, we are left with

Equation:

$$\frac{\partial E_y(x, t)}{\partial x} = -\frac{\partial B_z(x, t)}{\partial t}.$$

We could have applied Faraday's law instead to the top surface (numbered 2) in [\[link\]](#), to obtain the resulting equation

Equation:

$$\frac{\partial E_z(x, t)}{\partial x} = \frac{\partial B_y(x, t)}{\partial t}.$$

This is the equation describing the spatially dependent E field produced by the time-dependent B field.

Next we apply the Ampère-Maxwell law (with $I = 0$) over the same two faces (Surface 3 and then Surface 2) of the rectangular box of [\[link\]](#).

Applying [\[link\]](#),

Equation:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 \varepsilon_0 (d/dt) \int_S \vec{\mathbf{E}} \cdot \mathbf{n} da$$

to Surface 3, and then to Surface 2, yields the two equations

Equation:

$$\frac{\partial B_y(x, t)}{\partial x} = -\varepsilon_0 \mu_0 \frac{\partial E_z(x, t)}{\partial t}, \text{ and}$$

Equation:

$$\frac{\partial B_z(x, t)}{\partial x} = -\varepsilon_0 \mu_0 \frac{\partial E_y(x, t)}{\partial t}.$$

These equations describe the spatially dependent B field produced by the time-dependent E field.

We next combine the equations showing the changing B field producing an E field with the equation showing the changing E field producing a B field. Taking the derivative of [\[link\]](#) with respect to x and using [\[link\]](#) gives

Equation:

$$\frac{\partial^2 E_y}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial E_y}{\partial x} \right) = -\frac{\partial}{\partial x} \left(\frac{\partial B_z}{\partial t} \right) = -\frac{\partial}{\partial t} \left(\frac{\partial B_z}{\partial x} \right) = \frac{\partial}{\partial t} \left(\varepsilon_0 \mu_0 \frac{\partial E_y}{\partial t} \right)$$

or

Note:

Equation:

$$\frac{\partial^2 E_y}{\partial x^2} = \varepsilon_0 \mu_0 \frac{\partial^2 E_y}{\partial t^2}.$$

This is the form taken by the general wave equation for our plane wave. Because the equations describe a wave traveling at some as-yet-unspecified speed c , we can assume the field components are each functions of $x - ct$ for the wave traveling in the $+x$ -direction, that is,

Equation:

$$E_y(x, t) = f(\xi) \quad \text{where } \xi = x - ct.$$

It is left as a mathematical exercise to show, using the chain rule for differentiation, that [\[link\]](#) and [\[link\]](#) imply

Equation:

$$1 = \varepsilon_0 \mu_0 c^2.$$

The speed of the electromagnetic wave in free space is therefore given in terms of the permeability and the permittivity of free space by

Note:

Equation:

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}.$$

We could just as easily have assumed an electromagnetic wave with field components $E_z(x, t)$ and $B_y(x, t)$. The same type of analysis with [\[link\]](#)

and [\[link\]](#) would also show that the speed of an electromagnetic wave is $c = 1/\sqrt{\epsilon_0\mu_0}$.

The physics of traveling electromagnetic fields was worked out by Maxwell in 1873. He showed in a more general way than our derivation that electromagnetic waves always travel in free space with a speed given by [\[link\]](#). If we evaluate the speed $c = \frac{1}{\sqrt{\epsilon_0\mu_0}}$, we find that

Equation:

$$c = \frac{1}{\sqrt{\left(8.85 \times 10^{-12} \frac{\text{C}^2}{\text{N}\cdot\text{m}^2}\right) \left(4\pi \times 10^{-7} \frac{\text{T}\cdot\text{m}}{\text{A}}\right)}} = 3.00 \times 10^8 \text{ m/s},$$

which is the speed of light. Imagine the excitement that Maxwell must have felt when he discovered this equation! He had found a fundamental connection between two seemingly unrelated phenomena: electromagnetic fields and light.

Note:

Exercise:

Problem:

Check Your Understanding The wave equation was obtained by (1) finding the E field produced by the changing B field, (2) finding the B field produced by the changing E field, and combining the two results. Which of Maxwell's equations was the basis of step (1) and which of step (2)?

Solution:

(1) Faraday's law, (2) the Ampère-Maxwell law

How the E and B Fields Are Related

So far, we have seen that the rates of change of different components of the E and B fields are related, that the electromagnetic wave is transverse, and that the wave propagates at speed c . We next show what Maxwell's equations imply about the ratio of the E and B field magnitudes and the relative directions of the E and B fields.

We now consider solutions to [\[link\]](#) in the form of plane waves for the electric field:

Equation:

$$E_y(x, t) = E_0 \cos(kx - \omega t).$$

We have arbitrarily taken the wave to be traveling in the $+x$ -direction and chosen its phase so that the maximum field strength occurs at the origin at time $t = 0$. We are justified in considering only sines and cosines in this way, and generalizing the results, because Fourier's theorem implies we can express any wave, including even square step functions, as a superposition of sines and cosines.

At any one specific point in space, the E field oscillates sinusoidally at angular frequency ω between $+E_0$ and $-E_0$, and similarly, the B field oscillates between $+B_0$ and $-B_0$. The amplitude of the wave is the maximum value of $E_y(x, t)$. The period of oscillation T is the time required for a complete oscillation. The frequency f is the number of complete oscillations per unit of time, and is related to the angular frequency ω by $\omega = 2\pi f$. The wavelength λ is the distance covered by one complete cycle of the wave, and the wavenumber k is the number of wavelengths that fit into a distance of 2π in the units being used. These quantities are related in the same way as for a mechanical wave:

Equation:

$$\omega = 2\pi f, \quad f = \frac{1}{T}, \quad k = \frac{2\pi}{\lambda}, \quad \text{and} \quad c = f\lambda = \omega/k.$$

Given that the solution of E_y has the form shown in [\[link\]](#), we need to determine the B field that accompanies it. From [\[link\]](#), the magnetic field component B_z must obey

Equation:

$$\begin{aligned}\frac{\partial B_z}{\partial t} &= -\frac{\partial E_y}{\partial x} \\ \frac{\partial B_z}{\partial t} &= -\frac{\partial}{\partial x} E_0 \cos(kx - \omega t) = kE_0 \sin(kx - \omega t).\end{aligned}$$

Because the solution for the B -field pattern of the wave propagates in the $+x$ -direction at the same speed c as the E -field pattern, it must be a function of $k(x - ct) = kx - \omega t$. Thus, we conclude from [\[link\]](#) that B_z is

Equation:

$$B_z(x, t) = \frac{k}{\omega} E_0 \cos(kx - \omega t) = \frac{1}{c} E_0 \cos(kx - \omega t).$$

These results may be written as

Equation:

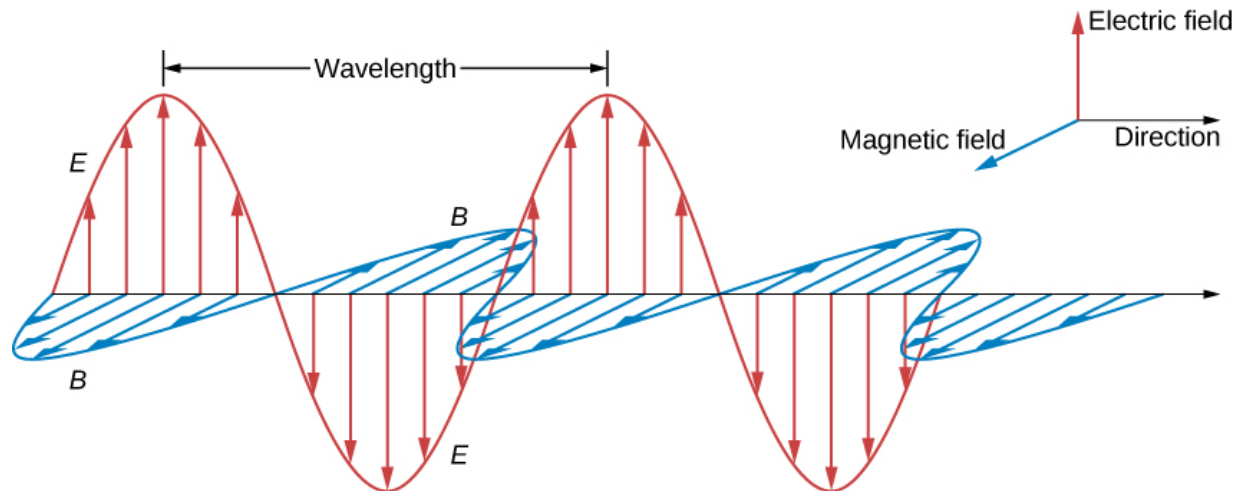
$$\begin{aligned}E_y(x, t) &= E_0 \cos(kx - \omega t) \\ B_z(x, t) &= B_0 \cos(kx - \omega t)\end{aligned}$$

Note:

Equation:

$$\frac{E_y}{B_z} = \frac{E_0}{B_0} = c.$$

Therefore, the peaks of the E and B fields coincide, as do the troughs of the wave, and at each point, the E and B fields are in the same ratio equal to the speed of light c . The plane wave has the form shown in [\[link\]](#).



The plane wave solution of Maxwell's equations has the B field directly proportional to the E field at each point, with the relative directions shown.

Example:

Calculating B -Field Strength in an Electromagnetic Wave

What is the maximum strength of the B field in an electromagnetic wave that has a maximum E -field strength of 1000 V/m?

Strategy

To find the B -field strength, we rearrange [\[link\]](#) to solve for B , yielding

Equation:

$$B = \frac{E}{c}.$$

Solution

We are given E , and c is the speed of light. Entering these into the expression for B yields

Equation:

$$B = \frac{1000 \text{ V/m}}{3.00 \times 10^8 \text{ m/s}} = 3.33 \times 10^{-6} \text{ T}.$$

Significance

The B -field strength is less than a tenth of Earth's admittedly weak magnetic field. This means that a relatively strong electric field of 1000 V/m is accompanied by a relatively weak magnetic field.

Changing electric fields create relatively weak magnetic fields. The combined electric and magnetic fields can be detected in electromagnetic waves, however, by taking advantage of the phenomenon of resonance, as Hertz did. A system with the same natural frequency as the electromagnetic wave can be made to oscillate. All radio and TV receivers use this principle to pick up and then amplify weak electromagnetic waves, while rejecting all others not at their resonant frequency.

Note:

Exercise:

Problem:

Check Your Understanding What conclusions did our analysis of Maxwell's equations lead to about these properties of a plane electromagnetic wave:

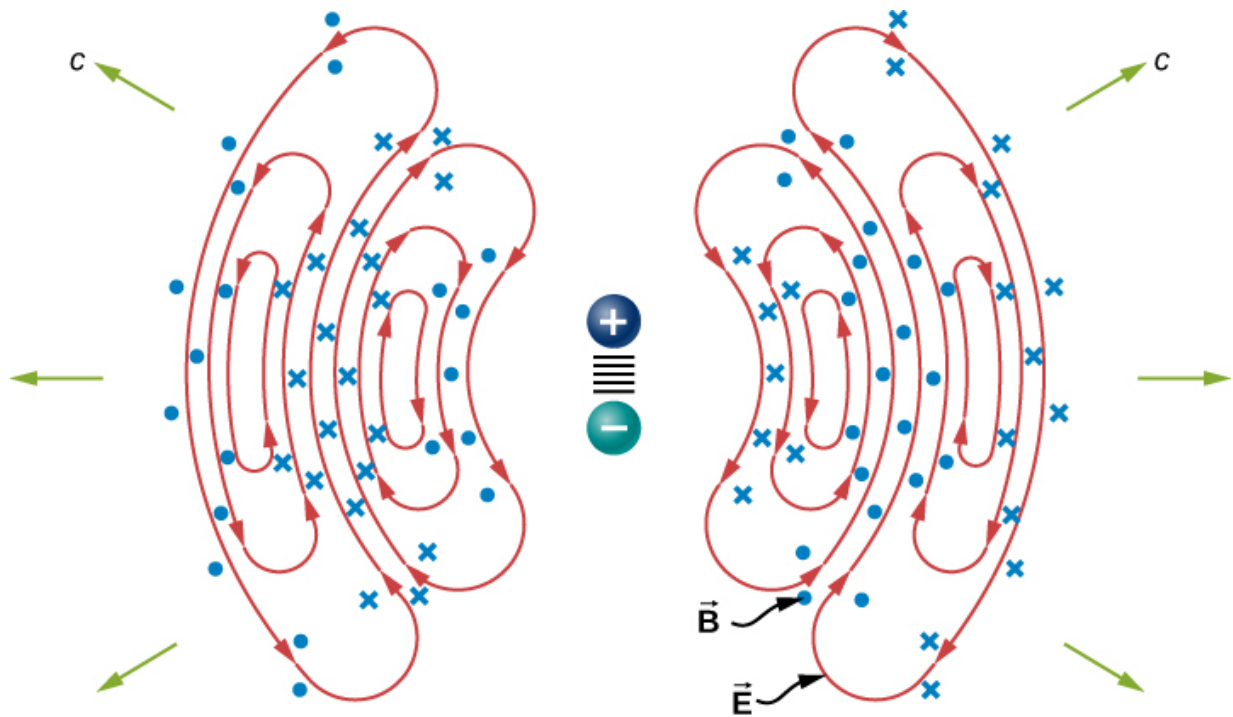
- (a) the relative directions of wave propagation, of the E field, and of B field,
- (b) the speed of travel of the wave and how the speed depends on frequency, and
- (c) the relative magnitudes of the E and B fields.

Solution:

a. The directions of wave propagation, of the E field, and of B field are all mutually perpendicular. b. The speed of the electromagnetic wave is the speed of light $c = 1/\sqrt{\epsilon_0\mu_0}$ independent of frequency. c. The ratio of electric and magnetic field amplitudes is $E/B = c$.

Production and Detection of Electromagnetic Waves

A steady electric current produces a magnetic field that is constant in time and which does not propagate as a wave. Accelerating charges, however, produce electromagnetic waves. An electric charge oscillating up and down, or an alternating current or flow of charge in a conductor, emit radiation at the frequencies of their oscillations. The electromagnetic field of a *dipole antenna* is shown in [\[link\]](#). The positive and negative charges on the two conductors are made to reverse at the desired frequency by the output of a transmitter as the power source. The continually changing current accelerates charge in the antenna, and this results in an oscillating electric field a distance away from the antenna. The changing electric fields produce changing magnetic fields that in turn produce changing electric fields, which thereby propagate as electromagnetic waves. The frequency of this radiation is the same as the frequency of the ac source that is accelerating the electrons in the antenna. The two conducting elements of the dipole antenna are commonly straight wires. The total length of the two wires is typically about one-half of the desired wavelength (hence, the alternative name *half-wave antenna*), because this allows standing waves to be set up and enhances the effectiveness of the radiation.



The oscillatory motion of the charges in a dipole antenna produces electromagnetic radiation.

The electric field lines in one plane are shown. The magnetic field is perpendicular to this plane. This radiation field has cylindrical symmetry around the axis of the dipole. Field lines near the dipole are not shown. The pattern is not at all uniform in all directions. The strongest signal is in directions perpendicular to the axis of the antenna, which would be horizontal if the antenna is mounted vertically. There is zero intensity along the axis of the antenna. The fields detected far from the antenna are from the changing electric and magnetic fields inducing each other and traveling as electromagnetic waves. Far from the antenna, the wave fronts, or surfaces of equal phase for the electromagnetic wave, are almost spherical. Even farther from the antenna, the radiation propagates like electromagnetic plane waves.

The electromagnetic waves carry energy away from their source, similar to a sound wave carrying energy away from a standing wave on a guitar string. An antenna for receiving electromagnetic signals works in reverse.

Incoming electromagnetic waves induce oscillating currents in the antenna, each at its own frequency. The radio receiver includes a tuner circuit, whose resonant frequency can be adjusted. The tuner responds strongly to the desired frequency but not others, allowing the user to tune to the desired broadcast. Electrical components amplify the signal formed by the moving electrons. The signal is then converted into an audio and/or video format.

Note:

Use this [simulation](#) to broadcast radio waves. Wiggle the transmitter electron manually or have it oscillate automatically. Display the field as a curve or vectors. The strip chart shows the electron positions at the transmitter and at the receiver.

Summary

- Maxwell's equations predict that the directions of the electric and magnetic fields of the wave, and the wave's direction of propagation, are all mutually perpendicular. The electromagnetic wave is a transverse wave.
- The strengths of the electric and magnetic parts of the wave are related by $c = E/B$, which implies that the magnetic field B is very weak relative to the electric field E .
- Accelerating charges create electromagnetic waves (for example, an oscillating current in a wire produces electromagnetic waves with the same frequency as the oscillation).

Conceptual Questions

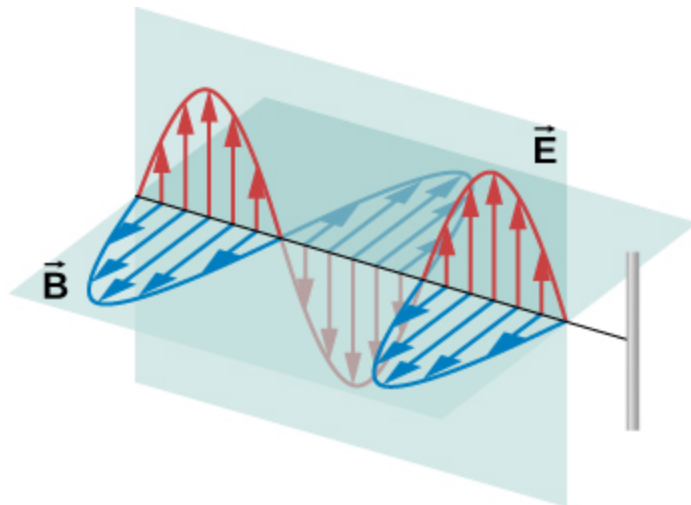
Exercise:

Problem:

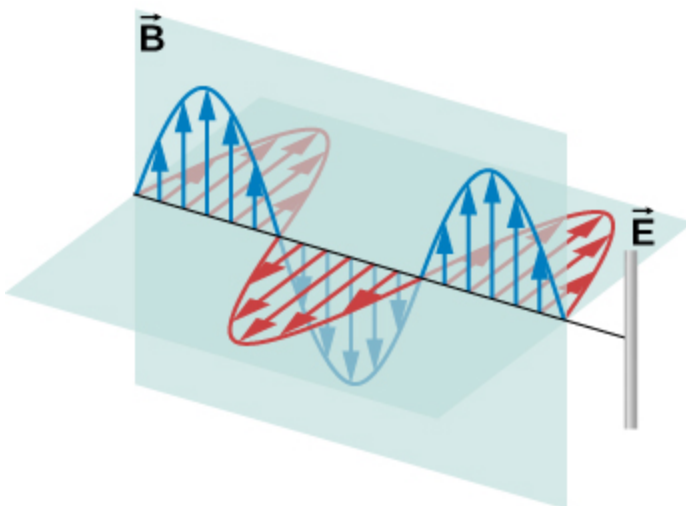
If the electric field of an electromagnetic wave is oscillating along the z -axis and the magnetic field is oscillating along the x -axis, in what possible direction is the wave traveling?

Exercise:**Problem:**

In which situation shown below will the electromagnetic wave be more successful in inducing a current in the wire? Explain.



(a)



(b)

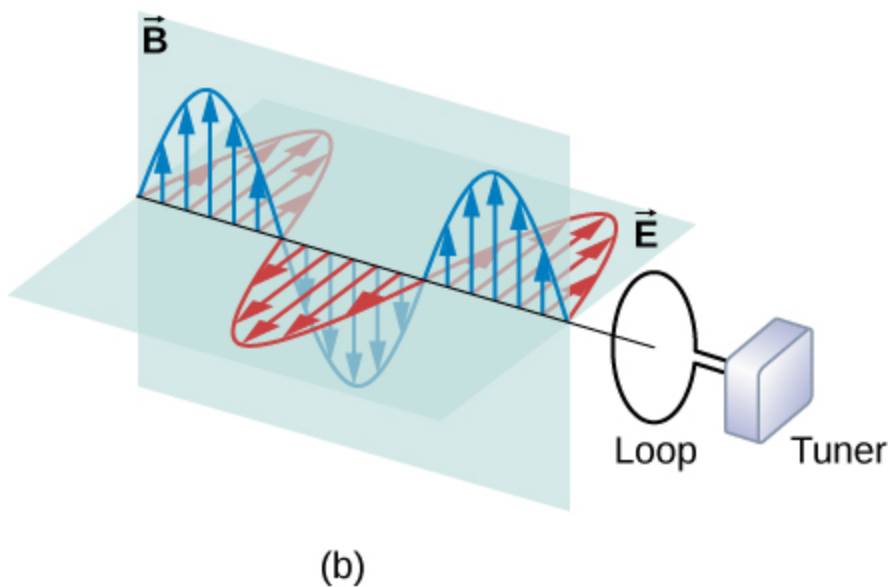
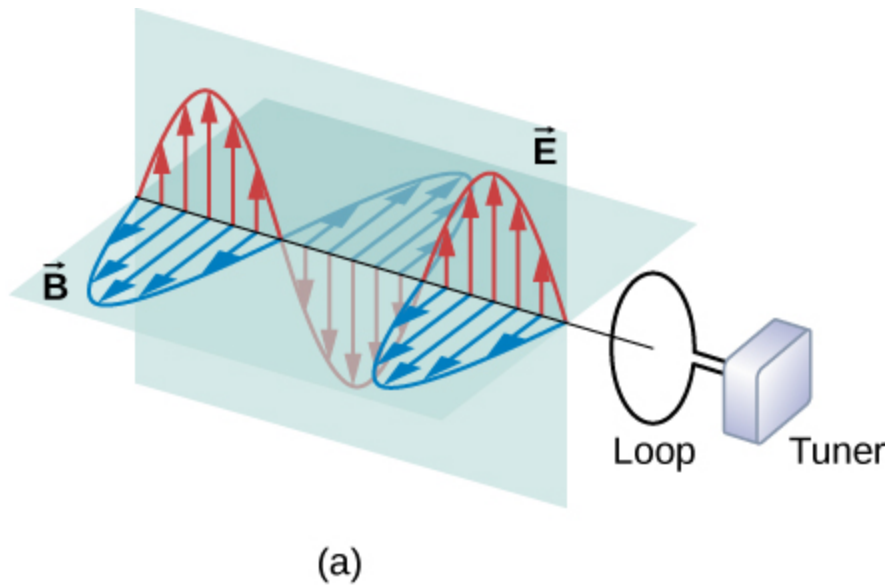
Solution:

in (a), because the electric field is parallel to the wire, accelerating the electrons

Exercise:

Problem:

In which situation shown below will the electromagnetic wave be more successful in inducing a current in the loop? Explain.



Exercise:

Problem:

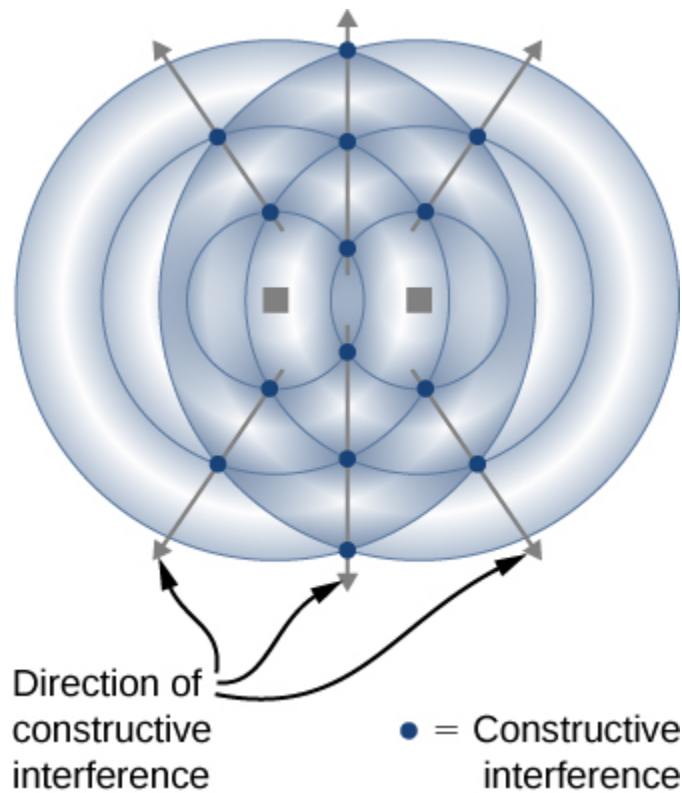
Under what conditions might wires in a circuit where the current flows in only one direction emit electromagnetic waves?

Solution:

A steady current in a dc circuit will not produce electromagnetic waves. If the magnitude of the current varies while remaining in the same direction, the wires will emit electromagnetic waves, for example, if the current is turned on or off.

Exercise:**Problem:**

Shown below is the interference pattern of two radio antennas broadcasting the same signal. Explain how this is analogous to the interference pattern for sound produced by two speakers. Could this be used to make a directional antenna system that broadcasts preferentially in certain directions? Explain.



Problems

Exercise:

Problem:

If the Sun suddenly turned off, we would not know it until its light stopped coming. How long would that be, given that the Sun is 1.496×10^{11} m away?

Solution:

499 s

Exercise:

Problem:

What is the maximum electric field strength in an electromagnetic wave that has a maximum magnetic field strength of $5.00 \times 10^{-4} \text{ T}$ (about 10 times Earth's magnetic field)?

Exercise:**Problem:**

An electromagnetic wave has a frequency of 12 MHz. What is its wavelength in vacuum?

Solution:

25 m

Exercise:**Problem:**

If electric and magnetic field strengths vary sinusoidally in time at frequency 1.00 GHz, being zero at $t = 0$, then $E = E_0 \sin 2\pi ft$ and $B = B_0 \sin 2\pi ft$. (a) When are the field strengths next equal to zero? (b) When do they reach their most negative value? (c) How much time is needed for them to complete one cycle?

Exercise:

Problem:

The electric field of an electromagnetic wave traveling in vacuum is described by the following wave function:

$$\vec{E} = (5.00 \text{ V/m}) \cos [kx - (6.00 \times 10^9 \text{ s}^{-1})t + 0.40] \hat{j}$$

where k is the wavenumber in rad/m, x is in m, t is in s.

Find the following quantities:

- (a) amplitude
 - (b) frequency
 - (c) wavelength
 - (d) the direction of the travel of the wave
 - (e) the associated magnetic field wave
-

Solution:

- a. 5.00 V/m; b. $9.55 \times 10^8 \text{ Hz}$; c. 31.4 cm; d. toward the +x-axis;
e. $B = (1.67 \times 10^{-8} \text{ T}) \cos [kx - (6 \times 10^9 \text{ s}^{-1})t + 0.40] \hat{k}$

Exercise:**Problem:**

A plane electromagnetic wave of frequency 20 GHz moves in the positive y-axis direction such that its electric field is pointed along the z-axis. The amplitude of the electric field is 10 V/m. The start of time is chosen so that at $t = 0$, the electric field has a value 10 V/m at the origin. (a) Write the wave function that will describe the electric field wave. (b) Find the wave function that will describe the associated magnetic field wave.

Exercise:

Problem:

The following represents an electromagnetic wave traveling in the direction of the positive y -axis:

$$E_x = 0; E_y = E_0 \cos (kx - \omega t); E_z = 0$$

$$B_x = 0; B_y = 0; B_z = B_0 \cos (kx - \omega t)$$

The wave is passing through a wide tube of circular cross-section of radius R whose axis is along the y -axis. Find the expression for the displacement current through the tube.

Solution:

$$I_d = \pi \epsilon_0 \omega R^2 E_0 \sin (kx - \omega t)$$

Energy Carried by Electromagnetic Waves

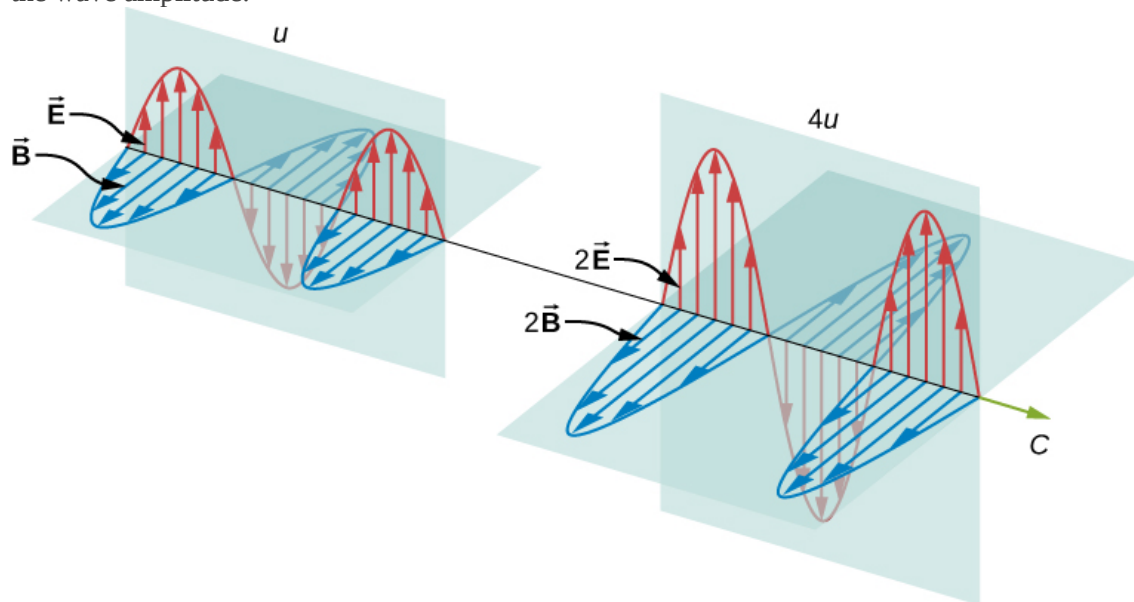
By the end of this section, you will be able to:

- Express the time-averaged energy density of electromagnetic waves in terms of their electric and magnetic field amplitudes
- Calculate the Poynting vector and the energy intensity of electromagnetic waves
- Explain how the energy of an electromagnetic wave depends on its amplitude, whereas the energy of a photon is proportional to its frequency

Anyone who has used a microwave oven knows there is energy in electromagnetic waves. Sometimes this energy is obvious, such as in the warmth of the summer Sun. Other times, it is subtle, such as the unfelt energy of gamma rays, which can destroy living cells.

Electromagnetic waves bring energy into a system by virtue of their electric and magnetic fields. These fields can exert forces and move charges in the system and, thus, do work on them. However, there is energy in an electromagnetic wave itself, whether it is absorbed or not. Once created, the fields carry energy away from a source. If some energy is later absorbed, the field strengths are diminished and anything left travels on.

Clearly, the larger the strength of the electric and magnetic fields, the more work they can do and the greater the energy the electromagnetic wave carries. In electromagnetic waves, the amplitude is the maximum field strength of the electric and magnetic fields ([link](#)). The wave energy is determined by the wave amplitude.



Energy carried by a wave depends on its amplitude. With electromagnetic waves, doubling the E fields and B fields quadruples the energy density u and the energy flux uc .

For a plane wave traveling in the direction of the positive x -axis with the phase of the wave chosen so that the wave maximum is at the origin at $t = 0$, the electric and magnetic fields obey the equations

Equation:

$$E_y(x, t) = E_0 \cos (kx - \omega t)$$

$$B_z(x, t) = B_0 \cos (kx - \omega t).$$

The energy in any part of the electromagnetic wave is the sum of the energies of the electric and magnetic fields. This energy per unit volume, or energy density u , is the sum of the energy density from the electric field and the energy density from the magnetic field. Expressions for both field energy densities were discussed earlier (u_E in [Capacitance](#) and u_B in [Inductance](#)). Combining these the contributions, we obtain

Equation:

$$u(x, t) = u_E + u_B = \frac{1}{2} \epsilon_0 E^2 + \frac{1}{2 \mu_0} B^2.$$

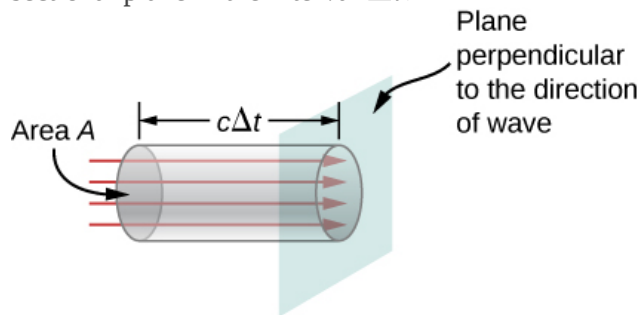
The expression $E = cB = \frac{1}{\sqrt{\epsilon_0 \mu_0}} B$ then shows that the magnetic energy density u_B and electric energy density u_E are equal, despite the fact that changing electric fields generally produce only small magnetic fields. The equality of the electric and magnetic energy densities leads to

Equation:

$$u(x, t) = \epsilon_0 E^2 = \frac{B^2}{\mu_0}.$$

The energy density moves with the electric and magnetic fields in a similar manner to the waves themselves.

We can find the rate of transport of energy by considering a small time interval Δt . As shown in [\[link\]](#), the energy contained in a cylinder of length $c\Delta t$ and cross-sectional area A passes through the cross-sectional plane in the interval Δt .



The energy $uAc\Delta t$ contained in the electric and magnetic fields of the electromagnetic wave in the volume $Ac\Delta t$ passes through the area A in time Δt .

The energy passing through area A in time Δt is

Equation:

$$u \times \text{volume} = uAc\Delta t.$$

The energy per unit area per unit time passing through a plane perpendicular to the wave, called the energy flux and denoted by S , can be calculated by dividing the energy by the area A and the time interval Δt .

Equation:

$$S = \frac{\text{Energy passing area } A \text{ in time } \Delta t}{A\Delta t} = uc = \varepsilon_0 c E^2 = \frac{1}{\mu_0} EB.$$

More generally, the flux of energy through any surface also depends on the orientation of the surface. To take the direction into account, we introduce a vector \vec{S} , called the **Poynting vector**, with the following definition:

Note:

Equation:

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}.$$

The cross-product of \vec{E} and \vec{B} points in the direction perpendicular to both vectors. To confirm that the direction of \vec{S} is that of wave propagation, and not its negative, return to [\[link\]](#). Note that Lenz's and Faraday's laws imply that when the magnetic field shown is increasing in time, the electric field is greater at x than at $x + \Delta x$. The electric field is decreasing with increasing x at the given time and location. The proportionality between electric and magnetic fields requires the electric field to increase in time along with the magnetic field. This is possible only if the wave is propagating to the right in the diagram, in which case, the relative orientations show that $\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}$ is specifically in the direction of propagation of the electromagnetic wave.

The energy flux at any place also varies in time, as can be seen by substituting u from [\[link\]](#) into [\[link\]](#).

Equation:

$$S(x, t) = c\varepsilon_0 E_0^2 \cos^2(kx - \omega t)$$

Because the frequency of visible light is very high, of the order of 10^{14} Hz, the energy flux for visible light through any area is an extremely rapidly varying quantity. Most measuring devices, including our eyes, detect only an average over many cycles. The time average of the energy flux is the intensity I of the electromagnetic wave and is the power per unit area. It can be expressed by averaging the cosine function in [\[link\]](#) over one complete cycle, which is the same as time-averaging over many cycles (here, T is one period):

Equation:

$$I = S_{\text{avg}} = c\varepsilon_0 E_0^2 \frac{1}{T} \int_0^T \cos^2 \left(2\pi \frac{t}{T} \right) dt.$$

We can either evaluate the integral, or else note that because the sine and cosine differ merely in phase, the average over a complete cycle for $\cos^2(\xi)$ is the same as for $\sin^2(\xi)$, to obtain

Equation:

$$\langle \cos^2 \xi \rangle = \frac{1}{2} [\langle \cos^2 \xi \rangle + \langle \sin^2 \xi \rangle] = \frac{1}{2} \langle 1 \rangle = \frac{1}{2}.$$

where the angle brackets $\langle \dots \rangle$ stand for the time-averaging operation. The intensity of light moving at speed c in vacuum is then found to be

Note:

Equation:

$$I = S_{\text{avg}} = \frac{1}{2} c\varepsilon_0 E_0^2$$

in terms of the maximum electric field strength E_0 , which is also the electric field amplitude. Algebraic manipulation produces the relationship

Note:

Equation:

$$I = \frac{cB_0^2}{2\mu_0}$$

where B_0 is the magnetic field amplitude, which is the same as the maximum magnetic field strength. One more expression for I_{avg} in terms of both electric and magnetic field strengths is useful. Substituting the fact that $cB_0 = E_0$, the previous expression becomes

Note:

Equation:

$$I = \frac{E_0 B_0}{2\mu_0}.$$

We can use whichever of the three preceding equations is most convenient, because the three equations are really just different versions of the same result: The energy in a wave is related to amplitude squared. Furthermore, because these equations are based on the assumption that the electromagnetic waves are sinusoidal, the peak intensity is twice the average intensity; that is, $I_0 = 2I$.

Example:

A Laser Beam

The beam from a small laboratory laser typically has an intensity of about $1.0 \times 10^{-3} \text{ W/m}^2$. Assuming that the beam is composed of plane waves, calculate the amplitudes of the electric and magnetic fields in the beam.

Strategy

Use the equation expressing intensity in terms of electric field to calculate the electric field from the intensity.

Solution

From [\[link\]](#), the intensity of the laser beam is

Equation:

$$I = \frac{1}{2} c \epsilon_0 E_0^2.$$

The amplitude of the electric field is therefore

Equation:

$$E_0 = \sqrt{\frac{2}{c \epsilon_0} I} = \sqrt{\frac{2}{(3.00 \times 10^8 \text{ m/s})(8.85 \times 10^{-12} \text{ F/m})} (1.0 \times 10^{-3} \text{ W/m}^2)} = 0.87 \text{ V/m}.$$

The amplitude of the magnetic field can be obtained from [\[link\]](#):

Equation:

$$B_0 = \frac{E_0}{c} = 2.9 \times 10^{-9} \text{ T}.$$

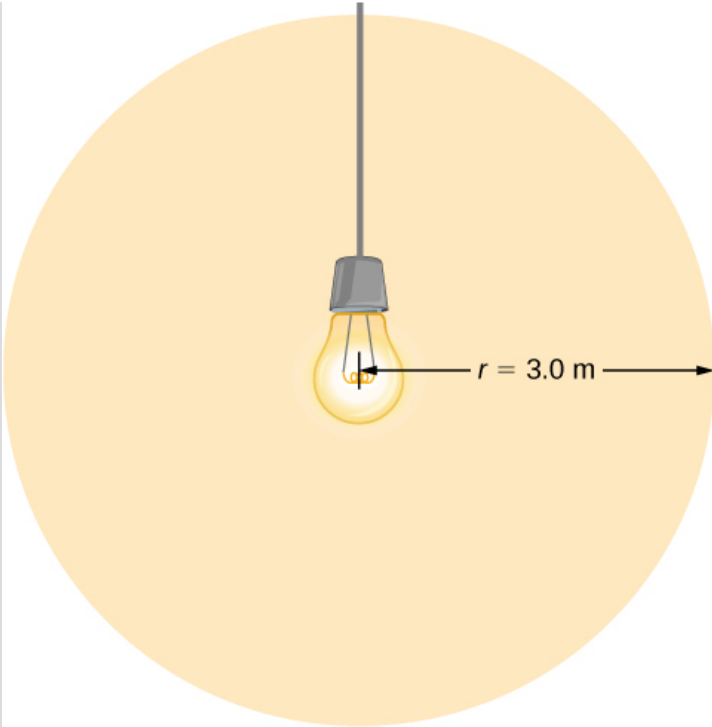
Example:

Light Bulb Fields

A light bulb emits 5.00 W of power as visible light. What are the average electric and magnetic fields from the light at a distance of 3.0 m?

Strategy

Assume the bulb's power output P is distributed uniformly over a sphere of radius 3.0 m to calculate the intensity, and from it, the electric field.



Solution

The power radiated as visible light is then

Equation:

$$I = \frac{P}{4\pi r^2} = \frac{c\epsilon_0 E_0^2}{2},$$

$$E_0 = \sqrt{2 \frac{P}{4\pi r^2 c \epsilon_0}} = \sqrt{2 \frac{5.00 \text{ W}}{4\pi (3.0 \text{ m})^2 (3.00 \times 10^8 \text{ m/s}) (8.85 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2)}} = 5.77 \text{ N/C},$$

$$B_0 = E_0/c = 1.92 \times 10^{-8} \text{ T}.$$

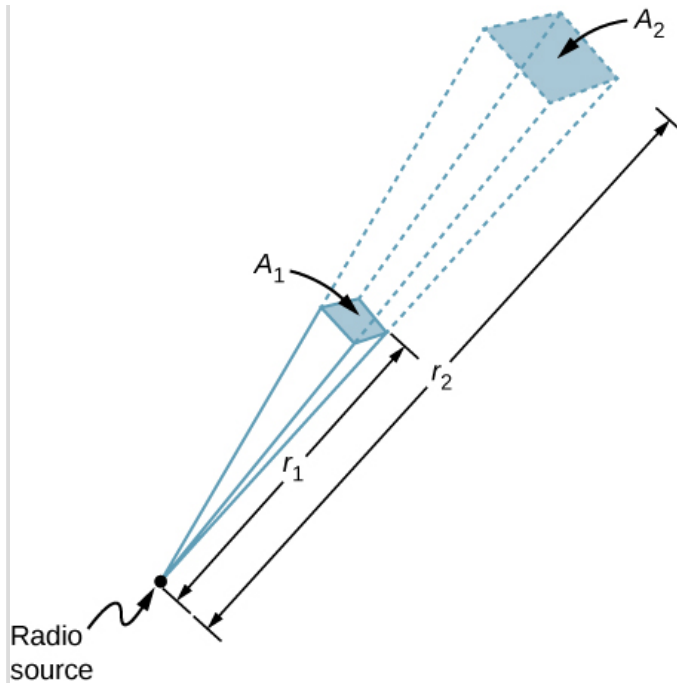
Significance

The intensity I falls off as the distance squared if the radiation is dispersed uniformly in all directions.

Example:

Radio Range

A 60-kW radio transmitter on Earth sends its signal to a satellite 100 km away ([link](#)). At what distance in the same direction would the signal have the same maximum field strength if the transmitter's output power were increased to 90 kW?



In three dimensions, a signal spreads over a solid angle as it travels outward from its source.

Strategy

The area over which the power in a particular direction is dispersed increases as distance squared, as illustrated in the figure. Change the power output P by a factor of $(90 \text{ kW}/60 \text{ kW})$ and change the area by the same factor to keep $I = \frac{P}{A} = \frac{c\epsilon_0 E_0^2}{2}$ the same. Then use the proportion of area A in the diagram to distance squared to find the distance that produces the calculated change in area.

Solution

Using the proportionality of the areas to the squares of the distances, and solving, we obtain from the diagram

Equation:

$$\frac{r_2^2}{r_1^2} = \frac{A_2}{A_1} = \frac{90 \text{ W}}{60 \text{ W}},$$

$$r_2 = \sqrt{\frac{90}{60}} (100 \text{ km}) = 122 \text{ km}.$$

Significance

The range of a radio signal is the maximum distance between the transmitter and receiver that allows for normal operation. In the absence of complications such as reflections from obstacles, the intensity follows an inverse square law, and doubling the range would require multiplying the power by four.

Summary

- The energy carried by any wave is proportional to its amplitude squared. For electromagnetic waves, this means intensity can be expressed as

Equation:

$$I = \frac{c\epsilon_0 E_0^2}{2}$$

where I is the average intensity in W/m^2 and E_0 is the maximum electric field strength of a continuous sinusoidal wave. This can also be expressed in terms of the maximum magnetic field strength B_0 as

Equation:

$$I = \frac{cB_0^2}{2\mu_0}$$

and in terms of both electric and magnetic fields as

Equation:

$$I = \frac{E_0 B_0}{2\mu_0}.$$

The three expressions for I_{avg} are all equivalent.

Conceptual Questions

Exercise:

Problem:

When you stand outdoors in the sunlight, why can you feel the energy that the sunlight carries, but not the momentum it carries?

Solution:

The amount of energy (about 100 W/m^2) is can quickly produce a considerable change in temperature, but the light pressure (about $3.00 \times 10^{-7} \text{ N/m}^2$) is much too small to notice.

Exercise:

Problem:

How does the intensity of an electromagnetic wave depend on its electric field? How does it depend on its magnetic field?

Exercise:

Problem: What is the physical significance of the Poynting vector?

Solution:

It has the magnitude of the energy flux and points in the direction of wave propagation. It gives the direction of energy flow and the amount of energy per area transported per second.

Exercise:**Problem:**

A 2.0-mW helium-neon laser transmits a continuous beam of red light of cross-sectional area 0.25 cm^2 . If the beam does not diverge appreciably, how would its rms electric field vary with distance from the laser? Explain.

Problems**Exercise:****Problem:**

While outdoors on a sunny day, a student holds a large convex lens of radius 4.0 cm above a sheet of paper to produce a bright spot on the paper that is 1.0 cm in radius, rather than a sharp focus. By what factor is the electric field in the bright spot of light related to the electric field of sunlight leaving the side of the lens facing the paper?

Exercise:**Problem:**

A plane electromagnetic wave travels northward. At one instant, its electric field has a magnitude of 6.0 V/m and points eastward. What are the magnitude and direction of the magnetic field at this instant?

Solution:

The magnetic field is downward, and it has magnitude $2.00 \times 10^{-8} \text{ T}$.

Exercise:

The electric field of an electromagnetic wave is given by $E = (6.0 \times 10^{-3} \text{ V/m}) \sin \left[2\pi \left(\frac{x}{18 \text{ m}} - \frac{t}{6.0 \times 10^{-8} \text{ s}} \right) \right] \hat{\mathbf{j}}$.

Problem: Write the equations for the associated magnetic field and Poynting vector.

Exercise:**Problem:**

A radio station broadcasts at a frequency of 760 kHz. At a receiver some distance from the antenna, the maximum magnetic field of the electromagnetic wave detected is $2.15 \times 10^{-11} \text{ T}$. (a) What is the maximum electric field? (b) What is the wavelength of the electromagnetic wave?

Solution:

a. $6.45 \times 10^{-3} \text{ V/m}$; b. 394 m

Exercise:**Problem:**

The filament in a clear incandescent light bulb radiates visible light at a power of 5.00 W. Model the glass part of the bulb as a sphere of radius $r_0 = 3.00$ cm and calculate the amount of electromagnetic energy from visible light inside the bulb.

Exercise:**Problem:**

At what distance does a 100-W lightbulb produce the same intensity of light as a 75-W lightbulb produces 10 m away? (Assume both have the same efficiency for converting electrical energy in the circuit into emitted electromagnetic energy.)

Solution:

11.5 m

Exercise:**Problem:**

An incandescent light bulb emits only 2.6 W of its power as visible light. What is the rms electric field of the emitted light at a distance of 3.0 m from the bulb?

Exercise:**Problem:**

A 150-W lightbulb emits 5% of its energy as electromagnetic radiation. What is the magnitude of the average Poynting vector 10 m from the bulb?

Solution:

$5.97 \times 10^{-3} \text{ W/m}^2$

Exercise:**Problem:**

A small helium-neon laser has a power output of 2.5 mW. What is the electromagnetic energy in a 1.0-m length of the beam?

Exercise:**Problem:**

At the top of Earth's atmosphere, the time-averaged Poynting vector associated with sunlight has a magnitude of about 1.4 kW/m^2 .

(a) What are the maximum values of the electric and magnetic fields for a wave of this intensity?

(b) What is the total power radiated by the sun? Assume that the Earth is 1.5×10^{11} m from the Sun and that sunlight is composed of electromagnetic plane waves.

Solution:

a. $E_0 = 1027 \text{ V/m}$, $B_0 = 3.42 \times 10^{-6} \text{ T}$; b. $3.96 \times 10^{26} \text{ W}$

Exercise:

Problem:

The magnetic field of a plane electromagnetic wave moving along the z axis is given by

$\vec{B} = B_0 (\cos kz + \omega t) \hat{j}$, where $B_0 = 5.00 \times 10^{-10} \text{ T}$ and $k = 3.14 \times 10^{-2} \text{ m}^{-1}$.

(a) Write an expression for the electric field associated with the wave. (b) What are the frequency and the wavelength of the wave? (c) What is its average Poynting vector?

Exercise:

Problem:

What is the intensity of an electromagnetic wave with a peak electric field strength of 125 V/m ?

Solution:

$$20.8 \text{ W/m}^2$$

Exercise:

Problem:

Assume the helium-neon lasers commonly used in student physics laboratories have power outputs of 0.500 mW . (a) If such a laser beam is projected onto a circular spot 1.00 mm in diameter, what is its intensity? (b) Find the peak magnetic field strength. (c) Find the peak electric field strength.

Exercise:

Problem:

An AM radio transmitter broadcasts 50.0 kW of power uniformly in all directions. (a) Assuming all of the radio waves that strike the ground are completely absorbed, and that there is no absorption by the atmosphere or other objects, what is the intensity 30.0 km away? (*Hint:* Half the power will be spread over the area of a hemisphere.) (b) What is the maximum electric field strength at this distance?

Solution:

$$\text{a. } 4.42 \times 10^{-6} \text{ W/m}^2; \text{ b. } 5.77 \times 10^{-2} \text{ V/m}$$

Exercise:

Problem:

Suppose the maximum safe intensity of microwaves for human exposure is taken to be 1.00 W/m^2 . (a) If a radar unit leaks 10.0 W of microwaves (other than those sent by its antenna) uniformly in all directions, how far away must you be to be exposed to an intensity considered to be safe? Assume that the power spreads uniformly over the area of a sphere with no complications from absorption or reflection. (b) What is the maximum electric field strength at the safe intensity? (Note that early radar units leaked more than modern ones do. This caused identifiable health problems, such as cataracts, for people who worked near them.)

Exercise:**Problem:**

A 2.50-m-diameter university communications satellite dish receives TV signals that have a maximum electric field strength (for one channel) of $7.50 \mu\text{V/m}$ (see below). (a) What is the intensity of this wave? (b) What is the power received by the antenna? (c) If the orbiting satellite broadcasts uniformly over an area of $1.50 \times 10^{13} \text{ m}^2$ (a large fraction of North America), how much power does it radiate?



Solution:

a. $7.47 \times 10^{-14} \text{ W/m}^2$; b. $3.66 \times 10^{-13} \text{ W}$; c. 1.12 W

Exercise:**Problem:**

Lasers can be constructed that produce an extremely high intensity electromagnetic wave for a brief time—called pulsed lasers. They are used to initiate nuclear fusion, for example. Such a laser may produce an electromagnetic wave with a maximum electric field strength of $1.00 \times 10^{11} \text{ V/m}$ for a time of 1.00 ns . (a) What is the maximum magnetic field strength in the wave? (b) What is the intensity of the beam? (c) What energy does it deliver on an 1.00-mm^2 area?

Glossary

Poynting vector

vector equal to the cross product of the electric-and magnetic fields, that describes the flow of electromagnetic energy through a surface

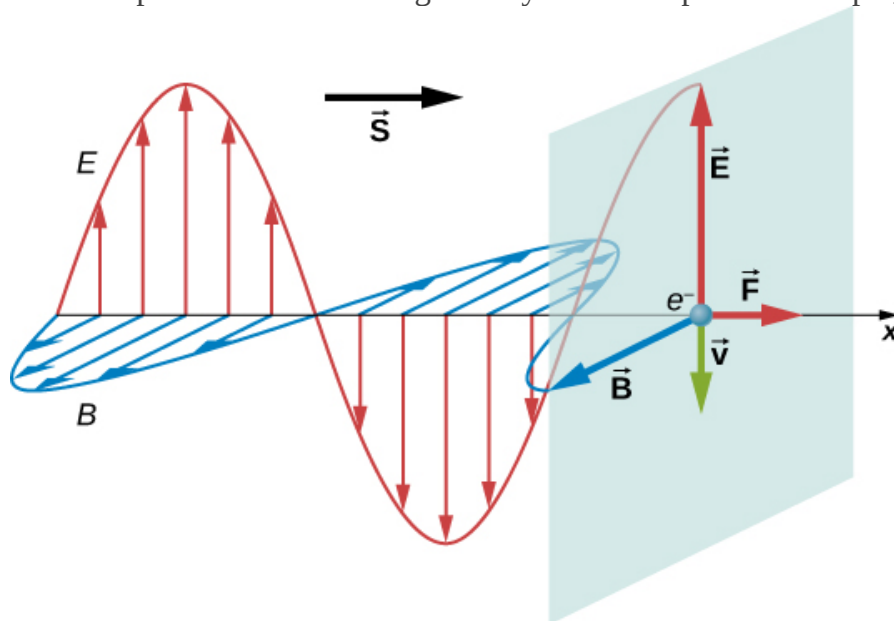
Momentum and Radiation Pressure

By the end of this section, you will be able to:

- Describe the relationship of the radiation pressure and the energy density of an electromagnetic wave
- Explain how the radiation pressure of light, while small, can produce observable astronomical effects

Material objects consist of charged particles. An electromagnetic wave incident on the object exerts forces on the charged particles, in accordance with the Lorentz force, [\[link\]](#). These forces do work on the particles of the object, increasing its energy, as discussed in the previous section. The energy that sunlight carries is a familiar part of every warm sunny day. A much less familiar feature of electromagnetic radiation is the extremely weak pressure that electromagnetic radiation produces by exerting a force in the direction of the wave. This force occurs because electromagnetic waves contain and transport momentum.

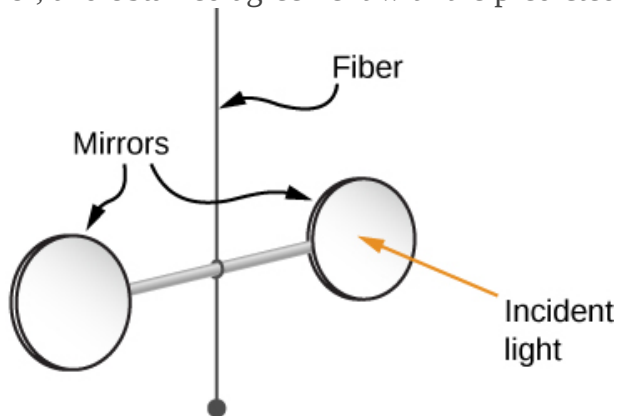
To understand the direction of the force for a very specific case, consider a plane electromagnetic wave incident on a metal in which electron motion, as part of a current, is damped by the resistance of the metal, so that the average electron motion is in phase with the force causing it. This is comparable to an object moving against friction and stopping as soon as the force pushing it stops ([\[link\]](#)). When the electric field is in the direction of the positive y-axis, electrons move in the negative y-direction, with the magnetic field in the direction of the positive z-axis. By applying the right-hand rule, and accounting for the negative charge of the electron, we can see that the force on the electron from the magnetic field is in the direction of the positive x-axis, which is the direction of wave propagation. When the E field reverses, the B field does too, and the force is again in the same direction. Maxwell's equations together with the Lorentz force equation imply the existence of radiation pressure much more generally than this specific example, however.



Electric and magnetic fields of an electromagnetic wave can combine to produce a force in the direction of propagation, as illustrated for the special case of electrons whose motion is highly damped by the resistance of a metal.

Maxwell predicted that an electromagnetic wave carries momentum. An object absorbing an electromagnetic wave would experience a force in the direction of propagation of the wave. The force corresponds to radiation pressure exerted on the object by the wave. The force would be twice as great if the radiation were reflected rather than absorbed.

Maxwell's prediction was confirmed in 1903 by Nichols and Hull by precisely measuring radiation pressures with a torsion balance. The schematic arrangement is shown in [\[link\]](#). The mirrors suspended from a fiber were housed inside a glass container. Nichols and Hull were able to obtain a small measurable deflection of the mirrors from shining light on one of them. From the measured deflection, they could calculate the unbalanced force on the mirror, and obtained agreement with the predicted value of the force.



Simplified diagram of the central part of the apparatus Nichols and Hull used to precisely measure radiation pressure and confirm Maxwell's prediction.

The **radiation pressure** p_{rad} applied by an electromagnetic wave on a perfectly absorbing surface turns out to be equal to the energy density of the wave:

Equation:

$$p_{\text{rad}} = u \text{ (Perfect absorber).}$$

If the material is perfectly reflecting, such as a metal surface, and if the incidence is along the normal to the surface, then the pressure exerted is twice as much because the momentum direction reverses upon reflection:

Equation:

$$p_{\text{rad}} = 2u \text{ (Perfect reflector).}$$

We can confirm that the units are right:

Equation:

$$[u] = \frac{\text{J}}{\text{m}^3} = \frac{\text{N} \cdot \text{m}}{\text{m}^3} = \frac{\text{N}}{\text{m}^2} = \text{units of pressure.}$$

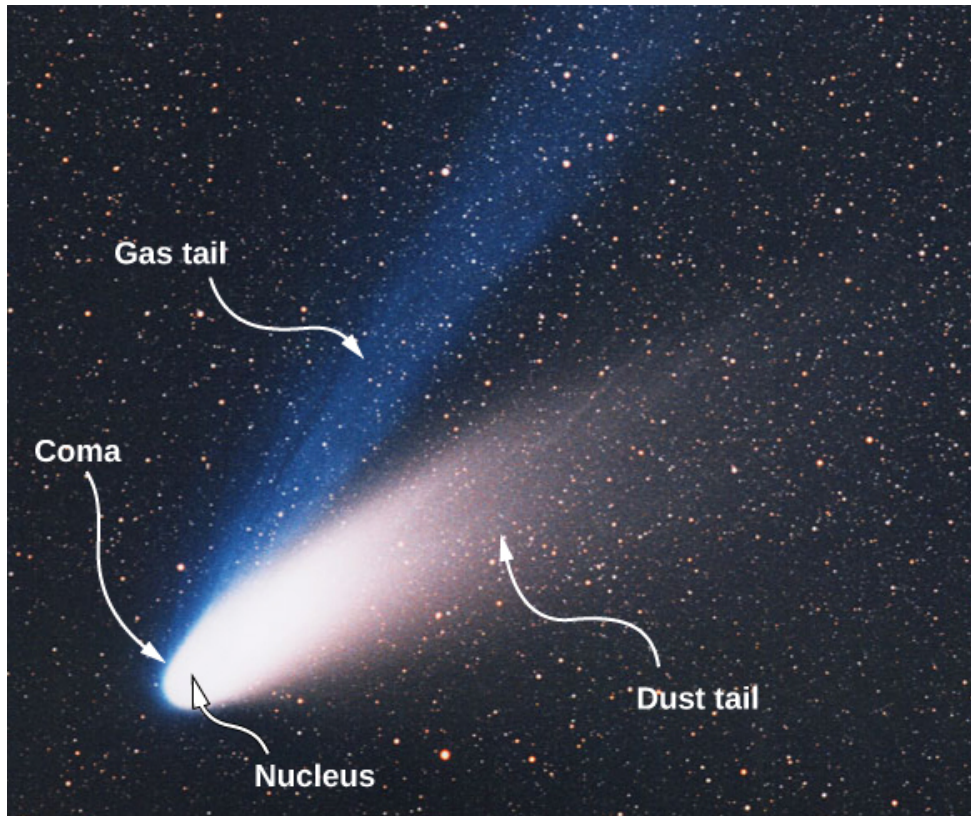
[\[link\]](#) and [\[link\]](#) give the instantaneous pressure, but because the energy density oscillates rapidly, we are usually interested in the time-averaged radiation pressure, which can be written in terms of intensity:

Note:

Equation:

$$p = \langle p_{\text{rad}} \rangle = \begin{cases} I/c & \text{Perfect absorber} \\ 2I/c & \text{Perfect reflector.} \end{cases}$$

Radiation pressure plays a role in explaining many observed astronomical phenomena, including the appearance of comets. Comets are basically chunks of icy material in which frozen gases and particles of rock and dust are embedded. When a comet approaches the Sun, it warms up and its surface begins to evaporate. The *coma* of the comet is the hazy area around it from the gases and dust. Some of the gases and dust form tails when they leave the comet. Notice in [\[link\]](#) that a comet has *two* tails. The *ion tail* (or *gas tail* in [\[link\]](#)) is composed mainly of ionized gases. These ions interact electromagnetically with the solar wind, which is a continuous stream of charged particles emitted by the Sun. The force of the solar wind on the ionized gases is strong enough that the ion tail almost always points directly away from the Sun. The second tail is composed of dust particles. Because the *dust tail* is electrically neutral, it does not interact with the solar wind. However, this tail is affected by the radiation pressure produced by the light from the Sun. Although quite small, this pressure is strong enough to cause the dust tail to be displaced from the path of the comet.



Evaporation of material being warmed by the Sun forms two tails, as shown in this photo of Comet Ison. (credit: modification of work by E. Slawik—ESO)

Example:

Halley's Comet

On February 9, 1986, Comet Halley was at its closest point to the Sun, about 9.0×10^{10} m from the center of the Sun. The average power output of the Sun is 3.8×10^{26} W.

- Calculate the radiation pressure on the comet at this point in its orbit. Assume that the comet reflects all the incident light.
- Suppose that a 10-kg chunk of material of cross-sectional area 4.0×10^{-2} m² breaks loose from the comet. Calculate the force on this chunk due to the solar radiation. Compare this force with the gravitational force of the Sun.

Strategy

Calculate the intensity of solar radiation at the given distance from the Sun and use that to calculate the radiation pressure. From the pressure and area, calculate the force.

Solution

- a. The intensity of the solar radiation is the average solar power per unit area. Hence, at 9.0×10^{10} m from the center of the Sun, we have

Equation:

$$I = S_{\text{avg}} = \frac{3.8 \times 10^{26} \text{ W}}{4\pi(9.0 \times 10^{10} \text{ m})^2} = 3.7 \times 10^3 \text{ W/m}^2.$$

Assuming the comet reflects all the incident radiation, we obtain from [\[link\]](#)

Equation:

$$p = \frac{2I}{c} = \frac{2(3.7 \times 10^3 \text{ W/m}^2)}{3.00 \times 10^8 \text{ m/s}} = 2.5 \times 10^{-5} \text{ N/m}^2.$$

- b. The force on the chunk due to the radiation is

Equation:

$$\begin{aligned} F &= pA = (2.5 \times 10^{-5} \text{ N/m}^2)(4.0 \times 10^{-2} \text{ m}^2) \\ &= 1.0 \times 10^{-6} \text{ N}, \end{aligned}$$

whereas the gravitational force of the Sun is

Equation:

$$F_g = \frac{GMm}{r^2} = \frac{(6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2)(2.0 \times 10^{30} \text{ kg})(10 \text{ kg})}{(9.0 \times 10^{10} \text{ m})^2} = 0.16 \text{ N}.$$

Significance

The gravitational force of the Sun on the chunk is therefore much greater than the force of the radiation.

After Maxwell showed that light carried momentum as well as energy, a novel idea eventually emerged, initially only as science fiction. Perhaps a spacecraft with a large reflecting light sail could use radiation pressure for propulsion. Such a vehicle would not have to carry fuel. It would experience a constant but small force from solar radiation, instead of the short bursts from rocket propulsion. It would accelerate slowly, but by being accelerated continuously, it would eventually reach great speeds. A spacecraft with small total mass and a sail with a large area would be necessary to obtain a usable acceleration.

When the space program began in the 1960s, the idea started to receive serious attention from NASA. The most recent development in light propelled spacecraft has come from a

citizen-funded group, the Planetary Society. It is currently testing the use of light sails to propel a small vehicle built from *CubeSats*, tiny satellites that NASA places in orbit for various research projects during space launches intended mainly for other purposes.

The *LightSail* spacecraft shown below ([link](#)) consists of three *CubeSats* bundled together. It has a total mass of only about 5 kg and is about the size as a loaf of bread. Its sails are made of very thin Mylar and open after launch to have a surface area of 32 m².



Two small *CubeSat* satellites deployed from the International Space Station in May, 2016. The solar sails open out when the CubeSats are far enough away from the Station. (credit: modification of work by NASA)

Note:

The first *LightSail* spacecraft was launched in 2015 to test the sail deployment system. It was placed in low-earth orbit in 2015 by hitching a ride on an Atlas 5 rocket launched for an unrelated mission. The test was successful, but the low-earth orbit allowed too much drag on the spacecraft to accelerate it by sunlight. Eventually, it burned in the atmosphere, as expected. The next Planetary Society's *LightSail* solar sailing spacecraft is scheduled for 2016. An [illustration](#) of the spacecraft, as it is expected to appear in flight, can be seen on the Planetary Society's website.

Example:**LightSail Acceleration**

The intensity of energy from sunlight at a distance of 1 AU from the Sun is 1370 W/m^2 . The *LightSail* spacecraft has sails with total area of 32 m^2 and a total mass of 5.0 kg . Calculate the maximum acceleration *LightSail* spacecraft could achieve from radiation pressure when it is about 1 AU from the Sun.

Strategy

The maximum acceleration can be expected when the sail is opened directly facing the Sun. Use the light intensity to calculate the radiation pressure and from it, the force on the sails. Then use Newton's second law to calculate the acceleration.

Solution

The radiation pressure is

Equation:

$$F = pA = 2uA = \frac{2I}{c}A = \frac{2(1370 \text{ W/m}^2)(32 \text{ m}^2)}{(3.00 \times 10^8 \text{ m/s})} = 2.92 \times 10^{-4} \text{ N}.$$

The resulting acceleration is

Equation:

$$a = \frac{F}{m} = \frac{2.92 \times 10^{-4} \text{ N}}{5.0 \text{ kg}} = 5.8 \times 10^{-5} \text{ m/s}^2.$$

Significance

If this small acceleration continued for a year, the craft would attain a speed of 1829 m/s , or 6600 km/h .

Note:**Exercise:****Problem:**

Check Your Understanding How would the speed and acceleration of a radiation-propelled spacecraft be affected as it moved farther from the Sun on an interplanetary space flight?

Solution:

Its acceleration would decrease because the radiation force is proportional to the intensity of light from the Sun, which decreases with distance. Its speed, however, would not change except for the effects of gravity from the Sun and planets.

Summary

- Electromagnetic waves carry momentum and exert radiation pressure.
- The radiation pressure of an electromagnetic wave is directly proportional to its energy density.
- The pressure is equal to twice the electromagnetic energy intensity if the wave is reflected and equal to the incident energy intensity if the wave is absorbed.

Conceptual Questions

Exercise:

Problem:

Why is the radiation pressure of an electromagnetic wave on a perfectly reflecting surface twice as large as the pressure on a perfectly absorbing surface?

Solution:

The force on a surface acting over time Δt is the momentum that the force would impart to the object. The momentum change of the light is doubled if the light is reflected back compared with when it is absorbed, so the force acting on the object is twice as great.

Exercise:

Problem:

Why did the early Hubble Telescope photos of Comet Ison approaching Earth show it to have merely a fuzzy coma around it, and not the pronounced double tail that developed later (see below)?



(credit: modification of work by NASA,
ESA, J.-Y. Li (Planetary Science Institute),
and the Hubble Comet ISON Imaging
Science Team)

Exercise:

Problem:

- (a) If the electric field and magnetic field in a sinusoidal plane wave were interchanged, in which direction relative to before would the energy propagate?
- (b) What if the electric and the magnetic fields were both changed to their negatives?

Solution:

- a. According to the right hand rule, the direction of energy propagation would reverse.
- b. This would leave the vector \vec{S} , and therefore the propagation direction, the same.

Problems

Exercise:

Problem:

A 150-W lightbulb emits 5% of its energy as electromagnetic radiation. What is the radiation pressure on an absorbing sphere of radius 10 m that surrounds the bulb?

Solution:

$$1.99 \times 10^{-11} \text{ N/m}^2$$

Exercise:**Problem:**

What pressure does light emitted uniformly in all directions from a 100-W incandescent light bulb exert on a mirror at a distance of 3.0 m, if 2.6 W of the power is emitted as visible light?

Exercise:**Problem:**

A microscopic spherical dust particle of radius 2 μm and mass 10 μg is moving in outer space at a constant speed of 30 cm/sec. A wave of light strikes it from the opposite direction of its motion and gets absorbed. Assuming the particle accelerates opposite to the motion uniformly to zero speed in one second, what is the average electric field amplitude in the light?

Solution:

$$F = ma = (p)(\pi r^2), p = \frac{ma}{\pi r^2} = \frac{\varepsilon_0}{2} E_0^2$$

$$E_0 = \sqrt{\frac{2ma}{\varepsilon_0 \pi r^2}} = \sqrt{\frac{2(10^{-8} \text{ kg})(0.30 \text{ m/s}^2)}{(8.854 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2)(\pi)(2 \times 10^{-6} \text{ m})^2}}$$

$$E_0 = 7.34 \times 10^6 \text{ V/m}$$

Exercise:**Problem:**

A Styrofoam spherical ball of radius 2 mm and mass 20 μg is to be suspended by the radiation pressure in a vacuum tube in a lab. How much intensity will be required if the light is completely absorbed the ball?

Exercise:

Problem:

Suppose that \vec{S}_{avg} for sunlight at a point on the surface of Earth is 900 W/m^2 . (a) If sunlight falls perpendicularly on a kite with a reflecting surface of area 0.75 m^2 , what is the average force on the kite due to radiation pressure? (b) How is your answer affected if the kite material is black and absorbs all sunlight?

Solution:

a. $4.50 \times 10^{-6} \text{ N}$; b. it is reduced to half the pressure, $2.25 \times 10^{-6} \text{ N}$

Exercise:**Problem:**

Sunlight reaches the ground with an intensity of about 1.0 kW/m^2 . A sunbather has a body surface area of 0.8 m^2 facing the sun while reclining on a beach chair on a clear day. (a) how much energy from direct sunlight reaches the sunbather's skin per second? (b) What pressure does the sunlight exert if it is absorbed?

Exercise:**Problem:**

Suppose a spherical particle of mass m and radius R in space absorbs light of intensity I for time t . (a) How much work does the radiation pressure do to accelerate the particle from rest in the given time it absorbs the light? (b) How much energy carried by the electromagnetic waves is absorbed by the particle over this time based on the radiant energy incident on the particle?

Solution:

a. $W = \frac{1}{2} \frac{\pi^2 r^4}{mc^2} I^2 t^2$; b. $E = \pi r^2 I t$

Glossary

radiation pressure

force divided by area applied by an electromagnetic wave on a surface

The Electromagnetic Spectrum

By the end of this section, you will be able to:

- Explain how electromagnetic waves are divided into different ranges, depending on wavelength and corresponding frequency
- Describe how electromagnetic waves in different categories are produced
- Describe some of the many practical everyday applications of electromagnetic waves

Electromagnetic waves have a vast range of practical everyday applications that includes such diverse uses as communication by cell phone and radio broadcasting, WiFi, cooking, vision, medical imaging, and treating cancer. In this module, we discuss how electromagnetic waves are classified into categories such as radio, infrared, ultraviolet, and so on. We also summarize some of the main applications for each range.

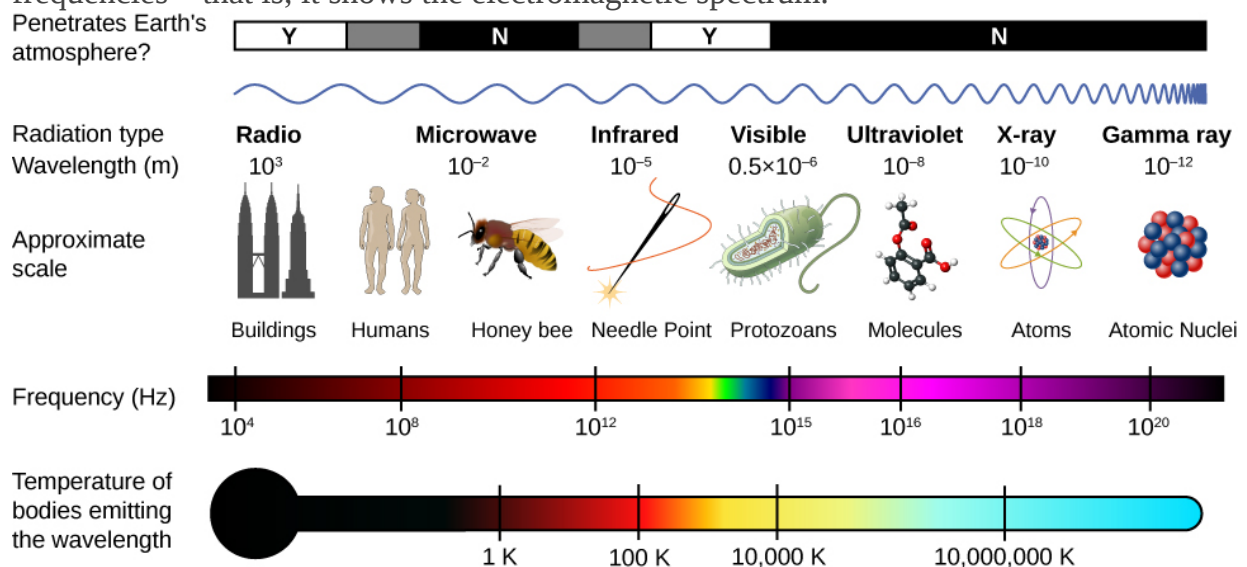
The different categories of electromagnetic waves differ in their wavelength range, or equivalently, in their corresponding frequency ranges. Their properties change smoothly from one frequency range to the next, with different applications in each range. A brief overview of the production and utilization of electromagnetic waves is found in [\[link\]](#).

Type of wave	Production	Applications	Issues
Radio	Accelerating charges	Communications Remote controls MRI	Requires control for band use
Microwaves	Accelerating charges and thermal agitation	Communications Ovens Radar Cell phone use	
Infrared	Thermal agitation and electronic transitions	Thermal imaging Heating	Absorbed by atmosphere Greenhouse effect

Type of wave	Production	Applications	Issues
Visible light	Thermal agitation and electronic transitions	Photosynthesis Human vision	
Ultraviolet	Thermal agitation and electronic transitions	Sterilization Vitamin D production	Ozone depletion Cancer causing
X-rays	Inner electronic transitions and fast collisions	Security Medical diagnosis Cancer therapy	Cancer causing
Gamma rays	Nuclear decay	Nuclear medicine Security Medical diagnosis Cancer therapy	Cancer causing Radiation damage

Electromagnetic Waves

The relationship $c = f\lambda$ between frequency f and wavelength λ applies to all waves and ensures that greater frequency means smaller wavelength. [\[link\]](#) shows how the various types of electromagnetic waves are categorized according to their wavelengths and frequencies—that is, it shows the electromagnetic spectrum.



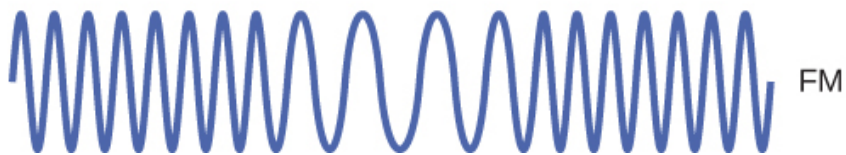
The electromagnetic spectrum, showing the major categories of electromagnetic waves.

Radio Waves

The term **radio waves** refers to electromagnetic radiation with wavelengths greater than about 0.1 m. Radio waves are commonly used for audio communications (i.e., for radios), but the term is used for electromagnetic waves in this range regardless of their application. Radio waves typically result from an alternating current in the wires of a broadcast antenna. They cover a very broad wavelength range and are divided into many subranges, including microwaves, electromagnetic waves used for AM and FM radio, cellular telephones, and TV signals.

There is no lowest frequency of radio waves, but ELF waves, or “extremely low frequency” are among the lowest frequencies commonly encountered, from 3 Hz to 3 kHz. The accelerating charge in the ac currents of electrical power lines produce electromagnetic waves in this range. ELF waves are able to penetrate sea water, which strongly absorbs electromagnetic waves of higher frequency, and therefore are useful for submarine communications.

In order to use an electromagnetic wave to transmit information, the amplitude, frequency, or phase of the wave is *modulated*, or varied in a controlled way that encodes the intended information into the wave. In AM radio transmission, the amplitude of the wave is modulated to mimic the vibrations of the sound being conveyed. Fourier’s theorem implies that the modulated AM wave amounts to a superposition of waves covering some narrow frequency range. Each AM station is assigned a specific carrier frequency that, by international agreement, is allowed to vary by ± 5 kHz. In FM radio transmission, the frequency of the wave is modulated to carry this information, as illustrated in [\[link\]](#), and the frequency of each station is allowed to use 100 kHz on each side of its carrier frequency. The electromagnetic wave produces a current in a receiving antenna, and the radio or television processes the signal to produce the sound and any image. The higher the frequency of the radio wave used to carry the data, the greater the detailed variation of the wave that can be carried by modulating it over each time unit, and the more data that can be transmitted per unit of time. The assigned frequencies for AM broadcasting are 540 to 1600 kHz, and for FM are 88 MHz to 108 MHz.



Electromagnetic waves are used to carry communications signals by varying the wave's amplitude (AM), its frequency (FM), or its phase.

Cell phone conversations, and television voice and video images are commonly transmitted as digital data, by converting the signal into a sequence of binary ones and zeros. This allows clearer data transmission when the signal is weak, and allows using computer algorithms to compress the digital data to transmit more data in each frequency range. Computer data as well is transmitted as a sequence of binary ones and zeros, each one or zero constituting one bit of data.

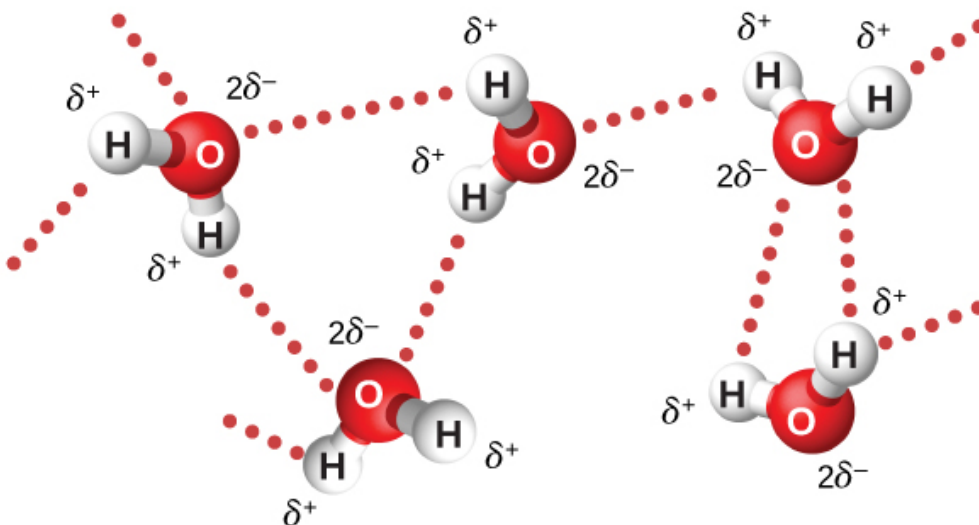
Microwaves

Microwaves are the highest-frequency electromagnetic waves that can be produced by currents in macroscopic circuits and devices. Microwave frequencies range from about 10^9Hz to nearly 10^{12}Hz . Their high frequencies correspond to short wavelengths compared with other radio waves—hence the name “microwave.” Microwaves also occur naturally as the cosmic background radiation left over from the origin of the universe. Along with other ranges of electromagnetic waves, they are part of the radiation that any

object above absolute zero emits and absorbs because of **thermal agitation**, that is, from the thermal motion of its atoms and molecules.

Most satellite-transmitted information is carried on microwaves. **Radar** is a common application of microwaves. By detecting and timing microwave echoes, radar systems can determine the distance to objects as diverse as clouds, aircraft, or even the surface of Venus.

Microwaves of 2.45 GHz are commonly used in microwave ovens. The electrons in a water molecule tend to remain closer to the oxygen nucleus than the hydrogen nuclei ([link](#)). This creates two separated centers of equal and opposite charges, giving the molecule a dipole moment (see [Electric Field](#)). The oscillating electric field of the microwaves inside the oven exerts a torque that tends to align each molecule first in one direction and then in the other, with the motion of each molecule coupled to others around it. This pumps energy into the continual thermal motion of the water to heat the food. The plate under the food contains no water, and remains relatively unheated.



The oscillating electric field in a microwave oven exerts a torque on water molecules because of their dipole moment, and the torque reverses direction 4.90×10^9 times per second. Interactions between the molecules distributes the energy being pumped into them. The δ^+ and δ^- denote the charge distribution on the molecules.

The microwaves in a microwave oven reflect off the walls of the oven, so that the superposition of waves produces standing waves, similar to the standing waves of a

vibrating guitar or violin string (see [Normal Modes of a Standing Sound Wave](#)). A rotating fan acts as a stirrer by reflecting the microwaves in different directions, and food turntables, help spread out the hot spots.

Example:**Why Microwave Ovens Heat Unevenly**

How far apart are the hotspots in a 2.45-GHz microwave oven?

Strategy

Consider the waves along one direction in the oven, being reflected at the opposite wall from where they are generated.

Solution

The antinodes, where maximum intensity occurs, are half the wavelength apart, with separation

Equation:

$$d = \frac{1}{2} \lambda = \frac{1}{2} \frac{c}{f} = \frac{3.00 \times 10^8 \text{ m/s}}{2 (2.45 \times 10^9 \text{ Hz})} = 6.02 \text{ cm.}$$

Significance

The distance between the hot spots in a microwave oven are determined by the wavelength of the microwaves.

A cell phone has a radio receiver and a weak radio transmitter, both of which can quickly tune to hundreds of specifically assigned microwave frequencies. The low intensity of the transmitted signal gives it an intentionally limited range. A ground-based system links the phone to only to the broadcast tower assigned to the specific small area, or cell, and smoothly transitions its connection to the next cell when the signal reception there is the stronger one. This enables a cell phone to be used while changing location.

Microwaves also provide the WiFi that enables owners of cell phones, laptop computers, and similar devices to connect wirelessly to the Internet at home and at coffee shops and airports. A wireless WiFi router is a device that exchanges data over the Internet through the cable or another connection, and uses microwaves to exchange the data wirelessly with devices such as cell phones and computers. The term WiFi itself refers to the standards followed in modulating and analyzing the microwaves so that wireless routers and devices from different manufacturers work compatibly with one another. The computer data in each direction consist of sequences of binary zeros and ones, each corresponding to a binary bit. The microwaves are in the range of 2.4 GHz to 5.0 GHz range.

Other wireless technologies also use microwaves to provide everyday communications between devices. Bluetooth developed alongside WiFi as a standard for radio communication in the 2.4-GHz range between nearby devices, for example, to link to headphones and audio earpieces to devices such as radios, or a driver's cell phone to a hands-free device to allow answering phone calls without fumbling directly with the cell phone.

Microwaves find use also in radio tagging, using RFID (radio frequency identification) technology. Examples are RFID tags attached to store merchandise, transponder for toll booths use attached to the windshield of a car, or even a chip embedded into a pet's skin. The device responds to a microwave signal by emitting a signal of its own with encoded information, allowing stores to quickly ring up items at their cash registers, drivers to charge tolls to their account without stopping, and lost pets to be reunited with their owners. NFC (near field communication) works similarly, except it is much shorter range. Its mechanism of interaction is the induced magnetic field at microwave frequencies between two coils. Cell phones that have NFC capability and the right software can supply information for purchases using the cell phone instead of a physical credit card. The very short range of the data transfer is a desired security feature in this case.

Infrared Radiation

The boundary between the microwave and infrared regions of the electromagnetic spectrum is not well defined (see [\[link\]](#)). **Infrared radiation** is generally produced by thermal motion, and the vibration and rotation of atoms and molecules. Electronic transitions in atoms and molecules can also produce infrared radiation. About half of the solar energy arriving at Earth is in the infrared region, with most of the rest in the visible part of the spectrum. About 23% of the solar energy is absorbed in the atmosphere, about 48% is absorbed at Earth's surface, and about 29% is reflected back into space.[\[footnote\]](#)
<http://earthobservatory.nasa.gov/Features/EnergyBalance/page4.php>

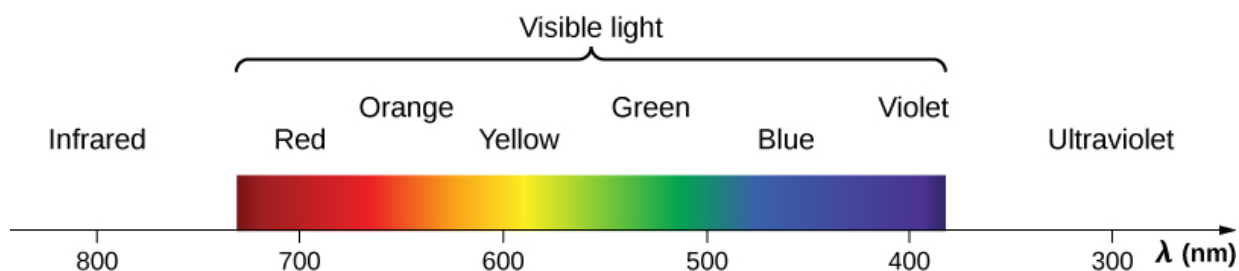
The range of infrared frequencies extends up to the lower limit of visible light, just below red. In fact, infrared means "below red." Water molecules rotate and vibrate particularly well at infrared frequencies. Reconnaissance satellites can detect buildings, vehicles, and even individual humans by their infrared emissions, whose power radiation is proportional to the fourth power of the absolute temperature. More mundanely, we use infrared lamps, including those called *quartz heaters*, to preferentially warm us because we absorb infrared better than our surroundings.

The familiar handheld "remotes" for changing channels and settings on television sets often transmit their signal by modulating an infrared beam. If you try to use a TV remote without the infrared emitter being in direct line of sight with the infrared detector, you may find the television not responding. Some remotes use Bluetooth instead and reduce this annoyance.

Visible Light

Visible light is the narrow segment of the electromagnetic spectrum between about 400 nm and about 750 nm to which the normal human eye responds. Visible light is produced by vibrations and rotations of atoms and molecules, as well as by electronic transitions within atoms and molecules. The receivers or detectors of light largely utilize electronic transitions.

Red light has the lowest frequencies and longest wavelengths, whereas violet has the highest frequencies and shortest wavelengths ([\[link\]](#)). Blackbody radiation from the Sun peaks in the visible part of the spectrum but is more intense in the red than in the violet, making the sun yellowish in appearance.



A small part of the electromagnetic spectrum that includes its visible components. The divisions between infrared, visible, and ultraviolet are not perfectly distinct, nor are those between the seven rainbow colors.

Living things—plants and animals—have evolved to utilize and respond to parts of the electromagnetic spectrum in which they are embedded. We enjoy the beauty of nature through visible light. Plants are more selective. Photosynthesis uses parts of the visible spectrum to make sugars.

Ultraviolet Radiation

Ultraviolet means “above violet.” The electromagnetic frequencies of **ultraviolet radiation (UV)** extend upward from violet, the highest-frequency visible light. The highest-frequency ultraviolet overlaps with the lowest-frequency X-rays. The wavelengths of ultraviolet extend from 400 nm down to about 10 nm at its highest frequencies. Ultraviolet is produced by atomic and molecular motions and electronic transitions.

UV radiation from the Sun is broadly subdivided into three wavelength ranges: UV-A (320–400 nm) is the lowest frequency, then UV-B (290–320 nm) and UV-C (220–290 nm). Most UV-B and all UV-C are absorbed by ozone (O₃) molecules in the upper atmosphere. Consequently, 99% of the solar UV radiation reaching Earth's surface is UV-A.

Sunburn is caused by large exposures to UV-B and UV-C, and repeated exposure can increase the likelihood of skin cancer. The tanning response is a defense mechanism in which the body produces pigments in inert skin layers to reduce exposure of the living cells below.

As examined in a later chapter, the shorter the wavelength of light, the greater the energy change of an atom or molecule that absorbs the light in an electronic transition. This makes short-wavelength ultraviolet light damaging to living cells. It also explains why ultraviolet radiation is better able than visible light to cause some materials to glow, or *fluoresce*.

Besides the adverse effects of ultraviolet radiation, there are also benefits of exposure in nature and uses in technology. Vitamin D production in the skin results from exposure to UV-B radiation, generally from sunlight. Several studies suggest vitamin D deficiency is associated with the development of a range of cancers (prostate, breast, colon), as well as osteoporosis. Low-intensity ultraviolet has applications such as providing the energy to cause certain dyes to fluoresce and emit visible light, for example, in printed money to display hidden watermarks as counterfeit protection.

X-Rays

X-rays have wavelengths from about 10^{-8}m to 10^{-12}m . They have shorter wavelengths, and higher frequencies, than ultraviolet, so that the energy they transfer at an atomic level is greater. As a result, X-rays have adverse effects on living cells similar to those of ultraviolet radiation, but they are more penetrating. Cancer and genetic defects can be induced by X-rays. Because of their effect on rapidly dividing cells, X-rays can also be used to treat and even cure cancer.

The widest use of X-rays is for imaging objects that are opaque to visible light, such as the human body or aircraft parts. In humans, the risk of cell damage is weighed carefully against the benefit of the diagnostic information obtained.

Gamma Rays

Soon after nuclear radioactivity was first detected in 1896, it was found that at least three distinct types of radiation were being emitted, and these were designated as alpha, beta, and gamma rays. The most penetrating nuclear radiation, the **gamma ray** (γ ray), was later found to be an extremely high-frequency electromagnetic wave.

The lower end of the γ -ray frequency range overlaps the upper end of the X-ray range. Gamma rays have characteristics identical to X-rays of the same frequency—they differ only in source. The name “gamma rays” is generally used for electromagnetic radiation emitted by a nucleus, while X-rays are generally produced by bombarding a target with energetic electrons in an X-ray tube. At higher frequencies, γ rays are more penetrating and more damaging to living tissue. They have many of the same uses as X-rays, including cancer therapy. Gamma radiation from radioactive materials is used in nuclear medicine.

Note:

Use this [simulation](#) to explore how light interacts with molecules in our atmosphere.

Explore how light interacts with molecules in our atmosphere.

Identify that absorption of light depends on the molecule and the type of light.

Relate the energy of the light to the resulting motion.

Identify that energy increases from microwave to ultraviolet.

Predict the motion of a molecule based on the type of light it absorbs.

Note:

Exercise:

Problem:

Check Your Understanding How do the electromagnetic waves for the different kinds of electromagnetic radiation differ?

Solution:

They fall into different ranges of wavelength, and therefore also different corresponding ranges of frequency.

Summary

- The relationship among the speed of propagation, wavelength, and frequency for any wave is given by $v = f\lambda$, so that for electromagnetic waves, $c = f\lambda$, where f is the frequency, λ is the wavelength, and c is the speed of light.
- The electromagnetic spectrum is separated into many categories and subcategories, based on the frequency and wavelength, source, and uses of the electromagnetic waves.

Key Equations

Displacement current	$I_d = \varepsilon_0 \frac{d\Phi_E}{dt}$
Gauss's law	$\oint \vec{E} \cdot d\vec{A} = \frac{Q_{in}}{\varepsilon_0}$
Gauss's law for magnetism	$\oint \vec{B} \cdot d\vec{A} = 0$
Faraday's law	$\oint \vec{E} \cdot d\vec{s} = -\frac{d\Phi_m}{dt}$
Ampère-Maxwell law	$\oint \vec{B} \cdot d\vec{s} = \mu_0 I + \varepsilon_0 \mu_0 \frac{d\Phi_E}{dt}$
Wave equation for plane EM wave	$\frac{\partial^2 E_y}{\partial x^2} = \varepsilon_0 \mu_0 \frac{\partial^2 E_y}{\partial t^2}$
Speed of EM waves	$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$
Ratio of E field to B field in electromagnetic wave	$c = \frac{E}{B}$
Energy flux (Poynting) vector	$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}$
Average intensity of an electromagnetic wave	$I = S_{avg} = \frac{c\varepsilon_0 E_0^2}{2} = \frac{cB_0^2}{2\mu_0} = \frac{E_0 B_0}{2\mu_0}$
Radiation pressure	$p = \begin{cases} I/c & \text{Perfect absorber} \\ 2I/c & \text{Perfect reflector} \end{cases}$

Conceptual Questions

Exercise:

Problem:

Compare the speed, wavelength, and frequency of radio waves and X-rays traveling in a vacuum.

Exercise:**Problem:**

Accelerating electric charge emits electromagnetic radiation. How does this apply in each case: (a) radio waves, (b) infrared radiation.

Solution:

a. Radio waves are generally produced by alternating current in a wire or an oscillating electric field between two plates; b. Infrared radiation is commonly produced by heated bodies whose atoms and the charges in them vibrate at about the right frequency.

Exercise:**Problem:**

Compare and contrast the meaning of the prefix “micro” in the names of SI units in the term *microwaves*.

Exercise:**Problem:**

Part of the light passing through the air is scattered in all directions by the molecules comprising the atmosphere. The wavelengths of visible light are larger than molecular sizes, and the scattering is strongest for wavelengths of light closest to sizes of molecules.

(a) Which of the main colors of light is scattered the most? (b) Explain why this would give the sky its familiar background color at midday.

Solution:

a. blue; b. Light of longer wavelengths than blue passes through the air with less scattering, whereas more of the blue light is scattered in different directions in the sky to give it its blue color.

Exercise:

Problem:

When a bowl of soup is removed from a microwave oven, the soup is found to be steaming hot, whereas the bowl is only warm to the touch. Discuss the temperature changes that have occurred in terms of energy transfer.

Exercise:**Problem:**

Certain orientations of a broadcast television antenna give better reception than others for a particular station. Explain.

Solution:

A typical antenna has a stronger response when the wires forming it are orientated parallel to the electric field of the radio wave.

Exercise:

Problem: What property of light corresponds to loudness in sound?

Exercise:

Problem: Is the visible region a major portion of the electromagnetic spectrum?

Solution:

No, it is very narrow and just a small portion of the overall electromagnetic spectrum.

Exercise:**Problem:**

Can the human body detect electromagnetic radiation that is outside the visible region of the spectrum?

Exercise:**Problem:**

Radio waves normally have their E and B fields in specific directions, whereas visible light usually has its E and B fields in random and rapidly changing directions that are perpendicular to each other and to the propagation direction. Can you explain why?

Solution:

Visible light is typically produced by changes of energies of electrons in randomly oriented atoms and molecules. Radio waves are typically emitted by an ac current flowing along a wire, that has fixed orientation and produces electric fields pointed in particular directions.

Exercise:

Problem: Give an example of resonance in the reception of electromagnetic waves.

Exercise:

Problem:

Illustrate that the size of details of an object that can be detected with electromagnetic waves is related to their wavelength, by comparing details observable with two different types (for example, radar and visible light).

Solution:

Radar can observe objects the size of an airplane and uses radio waves of about 0.5 cm in wavelength. Visible light can be used to view single biological cells and has wavelengths of about 10^{-7} m.

Exercise:

In which part of the electromagnetic spectrum are each of these waves:

(a) $f = 10.0$ kHz, (b) $f = \lambda = 750$ nm,

Problem: (c) $f = 1.25 \times 10^8$ Hz, (d) 0.30 nm

Exercise:

Problem:

In what range of electromagnetic radiation are the electromagnetic waves emitted by power lines in a country that uses 50-Hz ac current?

Solution:

ELF radio waves

Exercise:

Problem:

If a microwave oven could be modified to merely tune the waves generated to be in the infrared range instead of using microwaves, how would this affect the uneven heating of the oven?

Exercise:

Problem:

A leaky microwave oven in a home can sometimes cause interference with the homeowner's WiFi system. Why?

Solution:

The frequency of 2.45 GHz of a microwave oven is close to the specific frequencies in the 2.4 GHz band used for WiFi.

Exercise:**Problem:**

When a television news anchor in a studio speaks to a reporter in a distant country, there is sometimes a noticeable lag between when the anchor speaks in the studio and when the remote reporter hears it and replies. Explain what causes this delay.

Problems**Exercise:****Problem:**

How many helium atoms, each with a radius of about 31 pm, must be placed end to end to have a length equal to one wavelength of 470 nm blue light?

Exercise:**Problem:**

If you wish to detect details of the size of atoms (about 0.2 nm) with electromagnetic radiation, it must have a wavelength of about this size. (a) What is its frequency? (b) What type of electromagnetic radiation might this be?

Solution:

a. 1.5×10^{18} Hz; b. X-rays

Exercise:**Problem:**

Find the frequency range of visible light, given that it encompasses wavelengths from 380 to 760 nm.

Exercise:

Problem:

(a) Calculate the wavelength range for AM radio given its frequency range is 540 to 1600 kHz. (b) Do the same for the FM frequency range of 88.0 to 108 MHz.

Solution:

a. The wavelength range is 187 m to 556 m. b. The wavelength range is 2.78 m to 3.41 m.

Exercise:**Problem:**

Radio station WWVB, operated by the National Institute of Standards and Technology (NIST) from Fort Collins, Colorado, at a low frequency of 60 kHz, broadcasts a time synchronization signal whose range covers the entire continental US. The timing of the synchronization signal is controlled by a set of atomic clocks to an accuracy of 1×10^{-12} s, and repeats every 1 minute. The signal is used for devices, such as radio-controlled watches, that automatically synchronize with it at preset local times. WWVB's long wavelength signal tends to propagate close to the ground.

(a) Calculate the wavelength of the radio waves from WWVB.

(b) Estimate the error that the travel time of the signal causes in synchronizing a radio controlled watch in Norfolk, Virginia, which is 1570 mi (2527 km) from Fort Collins, Colorado.

Exercise:**Problem:**

An outdoor WiFi unit for a picnic area has a 100-mW output and a range of about 30 m. What output power would reduce its range to 12 m for use with the same devices as before? Assume there are no obstacles in the way and that microwaves into the ground are simply absorbed.

Solution:

$$P_l = \left(\frac{12 \text{ m}}{30 \text{ m}} \right)^2 (100 \text{ mW}) = 16 \text{ mW}$$

Exercise:

Problem:

7. The prefix “mega” (M) and “kilo” (k), when referring to amounts of computer data, refer to factors of 1024 or 2^{10} rather than 1000 for the prefix *kilo*, and $1024^2 = 2^{20}$ rather than 1,000,000 for the prefix *Mega* (M). If a wireless (WiFi) router transfers 150 Mbps of data, how many bits per second is that in decimal arithmetic?

Exercise:**Problem:**

A computer user finds that his wireless router transmits data at a rate of 75 Mbps (megabits per second). Compare the average time to transmit one bit of data with the time difference between the wifi signal reaching an observer’s cell phone directly and by bouncing back to the observer from a wall 8.00 m past the observer.

Solution:

time for 1 bit = 1.27×10^{-8} s, difference in travel time is 5.34×10^{-8} s

Exercise:**Problem:**

(a) The ideal size (most efficient) for a broadcast antenna with one end on the ground is one-fourth the wavelength ($\lambda/4$) of the electromagnetic radiation being sent out. If a new radio station has such an antenna that is 50.0 m high, what frequency does it broadcast most efficiently? Is this in the AM or FM band? (b) Discuss the analogy of the fundamental resonant mode of an air column closed at one end to the resonance of currents on an antenna that is one-fourth their wavelength.

Exercise:**Problem:**

What are the wavelengths of (a) X-rays of frequency 2.0×10^{17} Hz? (b) Yellow light of frequency 5.1×10^{14} Hz? (c) Gamma rays of frequency 1.0×10^{23} Hz?

Solution:

a. 1.5×10^{-9} m; b. 5.9×10^{-7} m; c. 3.0×10^{-15} m

Exercise:

Problem: For red light of $\lambda = 660$ nm, what are f , ω , and k ?

Exercise:

Problem:

A radio transmitter broadcasts plane electromagnetic waves whose maximum electric field at a particular location is $1.55 \times 10^{-3} \text{ V/m}$. What is the maximum magnitude of the oscillating magnetic field at that location? How does it compare with Earth's magnetic field?

Solution:

$5.17 \times 10^{-12} \text{ T}$, the non-oscillating geomagnetic field of 25–65 μT is much larger

Exercise:**Problem:**

(a) Two microwave frequencies authorized for use in microwave ovens are: 915 and 2450 MHz. Calculate the wavelength of each. (b) Which frequency would produce smaller hot spots in foods due to interference effects?

Exercise:**Problem:**

During normal beating, the heart creates a maximum 4.00-mV potential across 0.300 m of a person's chest, creating a 1.00-Hz electromagnetic wave. (a) What is the maximum electric field strength created? (b) What is the corresponding maximum magnetic field strength in the electromagnetic wave? (c) What is the wavelength of the electromagnetic wave?

Solution:

a. $1.33 \times 10^{-2} \text{ V/m}$; b. $4.44 \times 10^{-11} \text{ T}$; c. $3.00 \times 10^8 \text{ m}$

Exercise:**Problem:**

Distances in space are often quoted in units of light-years, the distance light travels in 1 year. (a) How many meters is a light-year? (b) How many meters is it to Andromeda, the nearest large galaxy, given that it is $2.54 \times 10^6 \text{ ly}$ away? (c) The most distant galaxy yet discovered is $13.4 \times 10^9 \text{ ly}$ away. How far is this in meters?

Exercise:

Problem:

A certain 60.0-Hz ac power line radiates an electromagnetic wave having a maximum electric field strength of 13.0 kV/m. (a) What is the wavelength of this very-low-frequency electromagnetic wave? (b) What type of electromagnetic radiation is this wave (b) What is its maximum magnetic field strength?

Solution:

a. $5.00 \times 10^6 \text{ m}$; b. radio wave; c. $4.33 \times 10^{-5} \text{ T}$

Exercise:**Problem:**

(a) What is the frequency of the 193-nm ultraviolet radiation used in laser eye surgery? (b) Assuming the accuracy with which this electromagnetic radiation can ablate (reshape) the cornea is directly proportional to wavelength, how much more accurate can this UV radiation be than the shortest visible wavelength of light?

Additional Problems**Exercise:****Problem:**

In a region of space, the electric field is pointed along the x -axis, but its magnitude changes as described by

$$E_x = (10 \text{ N/C}) \sin (20x - 500t)$$

$$E_y = E_z = 0$$

where t is in nanoseconds and x is in cm. Find the displacement current through a circle of radius 3 cm in the $x = 0$ plane at $t = 0$.

Solution:

$$I_d = (10 \text{ N/C}) (8.845 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2) \pi (0.03 \text{ m})^2 (5000 \frac{1}{\text{s}}) = 1.25 \times 10^{-9} \text{ A}$$

Exercise:

Problem:

A microwave oven uses electromagnetic waves of frequency $f = 2.45 \times 10^9$ Hz to heat foods. The waves reflect from the inside walls of the oven to produce an interference pattern of standing waves whose antinodes are hot spots that can leave observable pit marks in some foods. The pit marks are measured to be 6.0 cm apart. Use the method employed by Heinrich Hertz to calculate the speed of electromagnetic waves this implies.

Use the [Appendix D](#) for the next two exercises

Exercise:**Problem:**

Galileo proposed measuring the speed of light by uncovering a lantern and having an assistant a known distance away uncover his lantern when he saw the light from Galileo's lantern, and timing the delay. How far away must the assistant be for the delay to equal the human reaction time of about 0.25 s?

Solution:

3.75×10^7 km, which is much greater than Earth's circumference

Exercise:**Problem:**

Show that the wave equation in one dimension

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 f}{\partial t^2}$$

is satisfied by any doubly differentiable function of either the form $f(x - vt)$ or $f(x + vt)$.

Exercise:**Problem:**

On its highest power setting, a microwave oven increases the temperature of 0.400 kg of spaghetti by 45.0°C in 120 s. (a) What was the rate of energy absorption by the spaghetti, given that its specific heat is $3.76 \times 10^3 \text{ J/kg} \cdot ^\circ\text{C}$? Assume the spaghetti is perfectly absorbing. (b) Find the average intensity of the microwaves, given that they are absorbed over a circular area 20.0 cm in diameter. (c) What is the peak electric field strength of the microwave? (d) What is its peak magnetic field strength?

Solution:

a. 564 W ; b. $1.80 \times 10^4 \text{ W/m}^2$; c. $3.68 \times 10^3 \text{ V/m}$; d. $1.23 \times 10^{-5} \text{ T}$

Exercise:

Problem:

A certain microwave oven projects 1.00 kW of microwaves onto a $30\text{-cm-by-}40\text{-cm}$ area. (a) What is its intensity in W/m^2 ? (b) Calculate the maximum electric field strength E_0 in these waves. (c) What is the maximum magnetic field strength B_0 ?

Exercise:

Problem:

Electromagnetic radiation from a 5.00-mW laser is concentrated on a 1.00-mm^2 area. (a) What is the intensity in W/m^2 ? (b) Suppose a 2.00-nC electric charge is in the beam. What is the maximum electric force it experiences? (c) If the electric charge moves at 400 m/s , what maximum magnetic force can it feel?

Solution:

a. $5.00 \times 10^3 \text{ W/m}^2$; b. $3.88 \times 10^{-6} \text{ N}$; c. $5.18 \times 10^{-12} \text{ N}$

Exercise:

Problem:

A 200-turn flat coil of wire 30.0 cm in diameter acts as an antenna for FM radio at a frequency of 100 MHz . The magnetic field of the incoming electromagnetic wave is perpendicular to the coil and has a maximum strength of $1.00 \times 10^{-12} \text{ T}$. (a) What power is incident on the coil? (b) What average emf is induced in the coil over one-fourth of a cycle? (c) If the radio receiver has an inductance of $2.50 \mu\text{H}$, what capacitance must it have to resonate at 100 MHz ?

Exercise:

Problem:

Suppose a source of electromagnetic waves radiates uniformly in all directions in empty space where there are no absorption or interference effects. (a) Show that the intensity is inversely proportional to r^2 , the distance from the source squared. (b) Show that the magnitudes of the electric and magnetic fields are inversely proportional to r .

Solution:

a. $I = \frac{P}{A} = \frac{P}{4\pi r^2} \propto \frac{1}{r^2}$; b. $I \propto E_0^2, B_0^2 \Rightarrow E_0^2, B_0^2 \propto \frac{1}{r^2} \Rightarrow E_0, B_0 \propto \frac{1}{r}$

Exercise:

Problem:

A radio station broadcasts its radio waves with a power of 50,000 W. What would be the intensity of this signal if it is received on a planet orbiting Proxima Centuri, the closest star to our Sun, at 4.243 ly away?

Exercise:**Problem:**

The Poynting vector describes a flow of energy whenever electric and magnetic fields are present. Consider a long cylindrical wire of radius r with a current I in the wire, with resistance R and voltage V . From the expressions for the electric field along the wire and the magnetic field around the wire, obtain the magnitude and direction of the Poynting vector at the surface. Show that it accounts for an energy flow into the wire from the fields around it that accounts for the Ohmic heating of the wire.

Solution:

$$\begin{aligned} \text{Power into the wire} &= \int \vec{S} \cdot d\vec{A} = \left(\frac{1}{\mu_0} EB \right) (2\pi r L) \\ &= \frac{1}{\mu_0} \left(\frac{V}{L} \right) \left(\frac{\mu_0 i}{2\pi r} \right) (2\pi r L) = iV = i^2 R \end{aligned}$$

Exercise:**Problem:**

The Sun's energy strikes Earth at an intensity of 1.37 kW/m^2 . Assume as a model approximation that all of the light is absorbed. (Actually, about 30% of the light intensity is reflected out into space.)

(a) Calculate the total force that the Sun's radiation exerts on Earth.

(b) Compare this to the force of gravity between the Sun and Earth.

Earth's mass is $5.972 \times 10^{24} \text{ kg}$.

Exercise:**Problem:**

If a *Lightsail* spacecraft were sent on a Mars mission, by what ratio of the final force to the initial force would its propulsion be reduced when it reached Mars?

Solution:

0.431

Exercise:

Problem:

Lunar astronauts placed a reflector on the Moon's surface, off which a laser beam is periodically reflected. The distance to the Moon is calculated from the round-trip time. (a) To what accuracy in meters can the distance to the Moon be determined, if this time can be measured to 0.100 ns? (b) What percent accuracy is this, given the average distance to the Moon is 384,400 km?

Exercise:**Problem:**

Radar is used to determine distances to various objects by measuring the round-trip time for an echo from the object. (a) How far away is the planet Venus if the echo time is 1000 s? (b) What is the echo time for a car 75.0 m from a highway police radar unit? (c) How accurately (in nanoseconds) must you be able to measure the echo time to an airplane 12.0 km away to determine its distance within 10.0 m?

Solution:

a. 1.5×10^{11} m; b. 5.0×10^{-7} s; c. 33 ns

Exercise:**Problem:**

Calculate the ratio of the highest to lowest frequencies of electromagnetic waves the eye can see, given the wavelength range of visible light is from 380 to 760 nm. (Note that the ratio of highest to lowest frequencies the ear can hear is 1000.)

Exercise:**Problem:**

How does the wavelength of radio waves for an AM radio station broadcasting at 1030 KHz compare with the wavelength of the lowest audible sound waves (of 20 Hz). The speed of sound in air at 20 °C is about 343 m/s.

Solution:

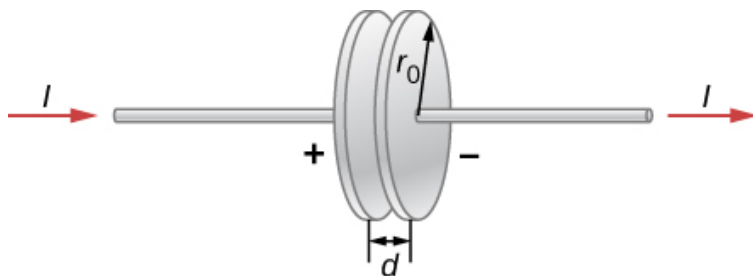
$$\text{sound: } \lambda_{\text{sound}} = \frac{v_s}{f} = \frac{343 \text{ m/s}}{20.0 \text{ Hz}} = 17.2 \text{ m}$$

$$\text{radio: } \lambda_{\text{radio}} = \frac{c}{f} = \frac{3.00 \times 10^8 \text{ m/s}}{1030 \times 10^3 \text{ Hz}} = 291 \text{ m; or } 17.1 \lambda_{\text{sound}}$$

Challenge Problems

Exercise:**Problem:**

A parallel-plate capacitor with plate separation d is connected to a source of emf that places a time-dependent voltage $V(t)$ across its circular plates of radius r_0 and area $A = \pi r_0^2$ (see below).



- (a) Write an expression for the time rate of change of energy inside the capacitor in terms of $V(t)$ and $dV(t)/dt$.
- (b) Assuming that $V(t)$ is increasing with time, identify the directions of the electric field lines inside the capacitor and of the magnetic field lines at the edge of the region between the plates, and then the direction of the Poynting vector \vec{S} at this location.
- (c) Obtain expressions for the time dependence of $E(t)$, for $B(t)$ from the displacement current, and for the magnitude of the Poynting vector at the edge of the region between the plates.
- (d) From \vec{S} , obtain an expression in terms of $V(t)$ and $dV(t)/dt$ for the rate at which electromagnetic field energy enters the region between the plates.
- (e) Compare the results of parts (a) and (d) and explain the relationship between them.

Exercise:**Problem:**

A particle of cosmic dust has a density $\rho = 2.0 \text{ g/cm}^3$. (a) Assuming the dust particles are spherical and light absorbing, and are at the same distance as Earth from the Sun, determine the particle size for which radiation pressure from sunlight is equal to the Sun's force of gravity on the dust particle. (b) Explain how the forces compare if the particle radius is smaller. (c) Explain what this implies about the sizes of dust particle likely to be present in the inner solar system compared with outside the Oort cloud.

Solution:

a. $0.29\ \mu\text{m}$; b. The radiation pressure is greater than the Sun's gravity if the particle size is smaller, because the gravitational force varies as the radius cubed while the radiation pressure varies as the radius squared. c. The radiation force outward implies that particles smaller than this are less likely to be near the Sun than outside the range of the Sun's radiation pressure.

Glossary**gamma ray (γ ray)**

extremely high frequency electromagnetic radiation emitted by the nucleus of an atom, either from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons; the lower end of the γ -ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation

infrared radiation

region of the electromagnetic spectrum with a frequency range that extends from just below the red region of the visible light spectrum up to the microwave region, or from $0.74\ \mu\text{m}$ to $300\ \mu\text{m}$

microwaves

electromagnetic waves with wavelengths in the range from 1 mm to 1 m; they can be produced by currents in macroscopic circuits and devices

radar

common application of microwaves; radar can determine the distance to objects as diverse as clouds and aircraft, as well as determine the speed of a car or the intensity of a rainstorm

radio waves

electromagnetic waves with wavelengths in the range from 1 mm to 100 km; they are produced by currents in wires and circuits and by astronomical phenomena

thermal agitation

thermal motion of atoms and molecules in any object at a temperature above absolute zero, which causes them to emit and absorb radiation

ultraviolet radiation

electromagnetic radiation in the range extending upward in frequency from violet light and overlapping with the lowest X-ray frequencies, with wavelengths from 400 nm down to about 10 nm

visible light

narrow segment of the electromagnetic spectrum to which the normal human eye responds, from about 400 to 750 nm

X-ray

invisible, penetrating form of very high frequency electromagnetic radiation, overlapping both the ultraviolet range and the γ -ray range

Introduction

class="introduction"

Due to total internal reflection, an underwater swimmer's image is reflected back into the water where the camera is located. The circular ripple in the image center is actually on the water surface. Due to the viewing angle, total internal reflection is not occurring at the top edge of this image, and we can see a view of activities on the pool deck.
(credit: modification of work by "jayhem"/Flickr)



Our investigation of light revolves around two questions of fundamental importance: (1) What is the nature of light, and (2) how does light behave under various circumstances? Answers to these questions can be found in Maxwell's equations (in [Electromagnetic Waves](#)), which predict the existence of electromagnetic waves and their behavior. Examples of light include radio and infrared waves, visible light, ultraviolet radiation, and X-rays. Interestingly, not all light phenomena can be explained by Maxwell's theory. Experiments performed early in the twentieth century showed that light has corpuscular, or particle-like, properties. The idea that light can display both wave and particle characteristics is called *wave-particle duality*, which is examined in [Photons and Matter Waves](#).

In this chapter, we study the basic properties of light. In the next few chapters, we investigate the behavior of light when it interacts with optical devices such as mirrors, lenses, and apertures.

The Propagation of Light

By the end of this section, you will be able to:

- Determine the index of refraction, given the speed of light in a medium
- List the ways in which light travels from a source to another location

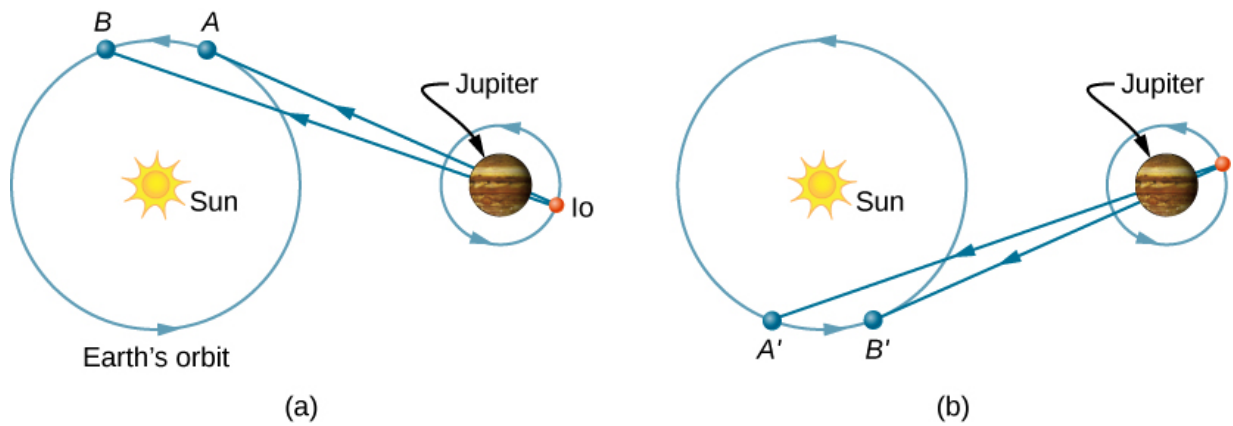
The speed of light in a vacuum c is one of the fundamental constants of physics. As you will see when you reach [Relativity](#), it is a central concept in Einstein's theory of relativity. As the accuracy of the measurements of the speed of light improved, it was found that different observers, even those moving at large velocities with respect to each other, measure the same value for the speed of light. However, the speed of light does vary in a precise manner with the material it traverses. These facts have far-reaching implications, as we will see in later chapters.

The Speed of Light: Early Measurements

The first measurement of the speed of light was made by the Danish astronomer Ole Roemer (1644–1710) in 1675. He studied the orbit of Io, one of the four large moons of Jupiter, and found that it had a period of revolution of 42.5 h around Jupiter. He also discovered that this value fluctuated by a few seconds, depending on the position of Earth in its orbit around the Sun. Roemer realized that this fluctuation was due to the finite speed of light and could be used to determine c .

Roemer found the period of revolution of Io by measuring the time interval between successive eclipses by Jupiter. [\[link\]](#)(a) shows the planetary configurations when such a measurement is made from Earth in the part of its orbit where it is receding from Jupiter. When Earth is at point A, Earth, Jupiter, and Io are aligned. The next time this alignment occurs, Earth is at point B, and the light carrying that information to Earth must travel to that point. Since B is farther from Jupiter than A, light takes more time to reach Earth when Earth is at B. Now imagine it is about 6 months later, and the planets are arranged as in part (b) of the figure. The measurement of Io's period begins with Earth at point A' and Io eclipsed by Jupiter. The next eclipse then occurs when Earth is at point B', to which the light carrying

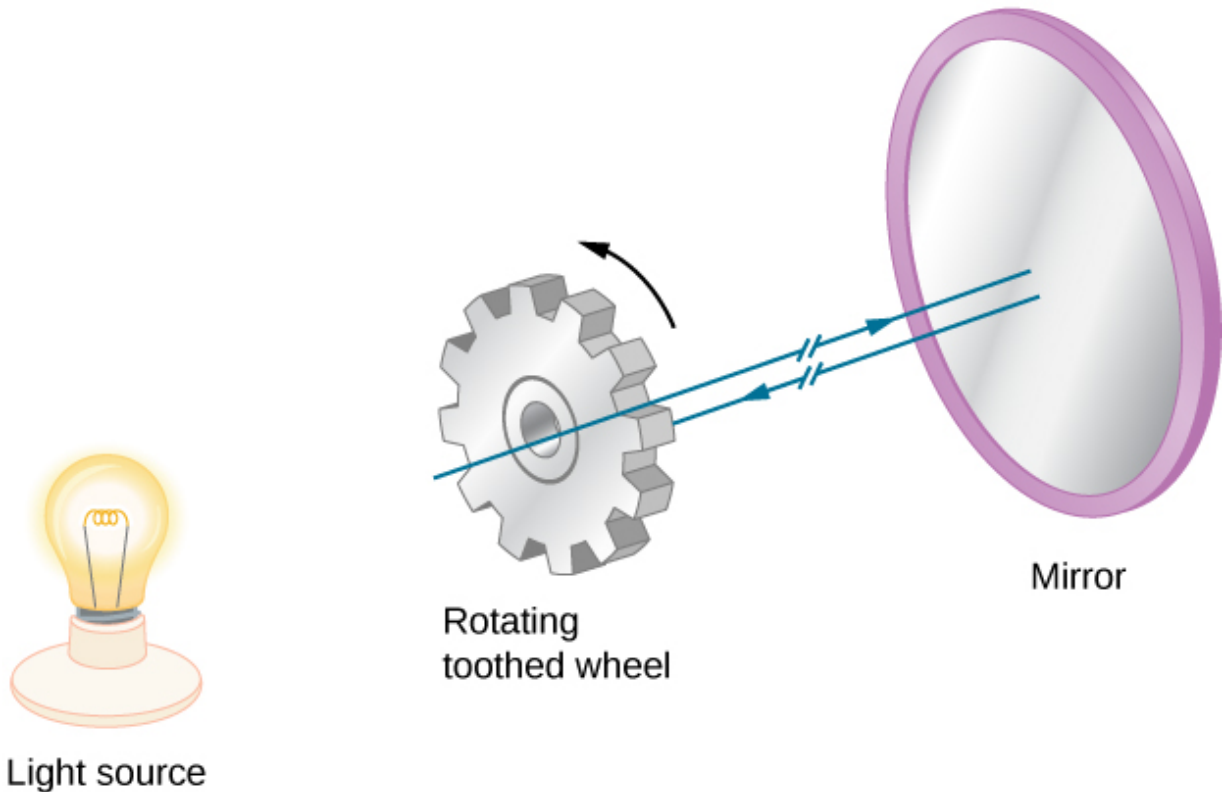
the information of this eclipse must travel. Since B' is closer to Jupiter than A' , light takes less time to reach Earth when it is at B' . This time interval between the successive eclipses of Io seen at A' and B' is therefore less than the time interval between the eclipses seen at A and B . By measuring the difference in these time intervals and with appropriate knowledge of the distance between Jupiter and Earth, Roemer calculated that the speed of light was $2.0 \times 10^8 \text{ m/s}$, which is 33% below the value accepted today.



Roemer's astronomical method for determining the speed of light. Measurements of Io's period done with the configurations of parts (a) and (b) differ, because the light path length and associated travel time increase from A to B (a) but decrease from A' to B' (b).

The first successful terrestrial measurement of the speed of light was made by Armand Fizeau (1819–1896) in 1849. He placed a toothed wheel that could be rotated very rapidly on one hilltop and a mirror on a second hilltop 8 km away ([link](#)). An intense light source was placed behind the wheel, so that when the wheel rotated, it chopped the light beam into a succession of pulses. The speed of the wheel was then adjusted until no light returned to the observer located behind the wheel. This could only happen if the wheel rotated through an angle corresponding to a displacement of $(n + \frac{1}{2})$ teeth, while the pulses traveled down to the mirror and back. Knowing the rotational speed of the wheel, the number of teeth on the wheel, and the

distance to the mirror, Fizeau determined the speed of light to be $3.15 \times 10^8 \text{ m/s}$, which is only 5% too high.



Fizeau's method for measuring the speed of light. The teeth of the wheel block the reflected light upon return when the wheel is rotated at a rate that matches the light travel time to and from the mirror.

The French physicist Jean Bernard Léon Foucault (1819–1868) modified Fizeau's apparatus by replacing the toothed wheel with a rotating mirror. In 1862, he measured the speed of light to be $2.98 \times 10^8 \text{ m/s}$, which is within 0.6% of the presently accepted value. Albert Michelson (1852–1931) also used Foucault's method on several occasions to measure the speed of light. His first experiments were performed in 1878; by 1926, he had refined the technique so well that he found c to be $(2.99796 \pm 4) \times 10^8 \text{ m/s}$.

Today, the speed of light is known to great precision. In fact, the speed of light in a vacuum c is so important that it is accepted as one of the basic physical quantities and has the value

Note:

Equation:

$$c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}$$

where the approximate value of $3.00 \times 10^8 \text{ m/s}$ is used whenever three-digit accuracy is sufficient.

Speed of Light in Matter

The speed of light through matter is less than it is in a vacuum, because light interacts with atoms in a material. The speed of light depends strongly on the type of material, since its interaction varies with different atoms, crystal lattices, and other substructures. We can define a constant of a material that describes the speed of light in it, called the **index of refraction** n :

Note:

Equation:

$$n = \frac{c}{v}$$

where v is the observed speed of light in the material.

Since the speed of light is always less than c in matter and equals c only in a vacuum, the index of refraction is always greater than or equal to one; that is, $n \geq 1$. [\[link\]](#) gives the indices of refraction for some representative substances. The values are listed for a particular wavelength of light, because they vary slightly with wavelength. (This can have important effects, such as colors separated by a prism, as we will see in [Dispersion](#).) Note that for gases, n is close to 1.0. This seems reasonable, since atoms in gases are widely separated, and light travels at c in the vacuum between atoms. It is common to take $n = 1$ for gases unless great precision is needed. Although the speed of light v in a medium varies considerably from its value c in a vacuum, it is still a large speed.

Medium	n
Gases at 0 °C, 1 atm	
Air	1.000293
Carbon dioxide	1.00045
Hydrogen	1.000139
Oxygen	1.000271
Liquids at 20 °C	
Benzene	1.501
Carbon disulfide	1.628
Carbon tetrachloride	1.461

Medium	<i>n</i>
Ethanol	1.361
Glycerine	1.473
Water, fresh	1.333
Solids at 20° C	
Diamond	2.419
Fluorite	1.434
Glass, crown	1.52
Glass, flint	1.66
Ice (at 0° C)	1.309
Polystyrene	1.49
Plexiglas	1.51
Quartz, crystalline	1.544
Quartz, fused	1.458
Sodium chloride	1.544
Zircon	1.923

Index of Refraction in Various Media For light with a wavelength of 589 nm in a vacuum

Example:**Speed of Light in Jewelry**

Calculate the speed of light in zircon, a material used in jewelry to imitate diamond.

Strategy

We can calculate the speed of light in a material v from the index of refraction n of the material, using the equation $n = c/v$.

Solution

Rearranging the equation $n = c/v$ for v gives us

Equation:

$$v = \frac{c}{n}.$$

The index of refraction for zircon is given as 1.923 in [\[link\]](#), and c is given in [\[link\]](#). Entering these values in the equation gives

Equation:

$$v = \frac{3.00 \times 10^8 \text{ m/s}}{1.923} = 1.56 \times 10^8 \text{ m/s}.$$

Significance

This speed is slightly larger than half the speed of light in a vacuum and is still high compared with speeds we normally experience. The only substance listed in [\[link\]](#) that has a greater index of refraction than zircon is diamond. We shall see later that the large index of refraction for zircon makes it sparkle more than glass, but less than diamond.

Note:**Exercise:****Problem:**

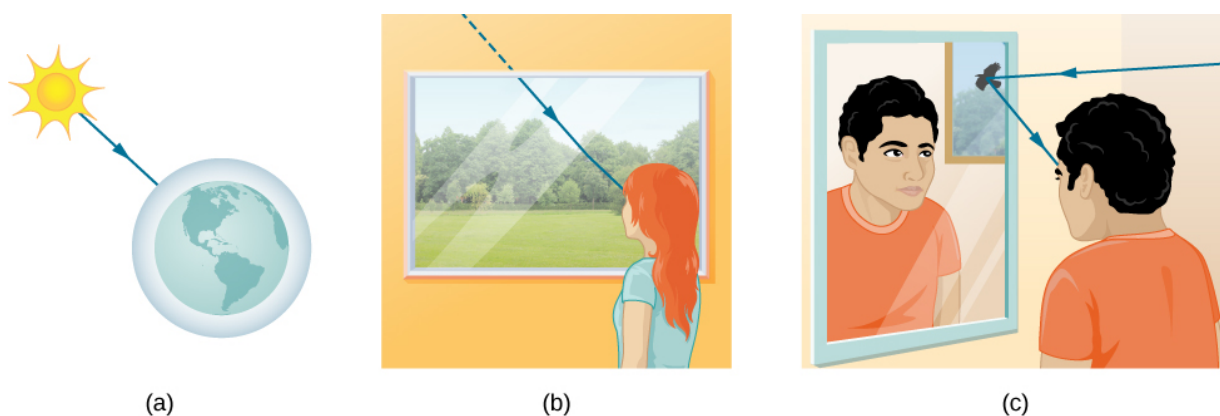
[\[link\]](#) shows that ethanol and fresh water have very similar indices of refraction. By what percentage do the speeds of light in these liquids differ?

Solution:

2.1% (to two significant figures)

The Ray Model of Light

You have already studied some of the wave characteristics of light in the previous chapter on [Electromagnetic Waves](#). In this chapter, we start mainly with the ray characteristics. There are three ways in which light can travel from a source to another location ([\[link\]](#)). It can come directly from the source through empty space, such as from the Sun to Earth. Or light can travel through various media, such as air and glass, to the the observer. Light can also arrive after being reflected, such as by a mirror. In all of these cases, we can model the path of light as a straight line called a **ray**.



Three methods for light to travel from a source to another location. (a) Light reaches the upper atmosphere of Earth, traveling through empty space directly from the source. (b) Light can reach a person by traveling through media like air and glass. (c) Light can also reflect from an object like a mirror. In the situations shown here, light interacts with objects large enough that it travels in straight lines, like a ray.

Experiments show that when light interacts with an object several times larger than its wavelength, it travels in straight lines and acts like a ray. Its wave characteristics are not pronounced in such situations. Since the wavelength of visible light is less than a micron (a thousandth of a millimeter), it acts like a ray in the many common situations in which it encounters objects larger than a micron. For example, when visible light encounters anything large enough that we can observe it with unaided eyes, such as a coin, it acts like a ray, with generally negligible wave characteristics.

In all of these cases, we can model the path of light as straight lines. Light may change direction when it encounters objects (such as a mirror) or in passing from one material to another (such as in passing from air to glass), but it then continues in a straight line or as a ray. The word “ray” comes from mathematics and here means a straight line that originates at some point. It is acceptable to visualize light rays as laser rays. The *ray model of light* describes the path of light as straight lines.

Since light moves in straight lines, changing directions when it interacts with materials, its path is described by geometry and simple trigonometry. This part of optics, where the ray aspect of light dominates, is therefore called **geometric optics**. Two laws govern how light changes direction when it interacts with matter. These are the *law of reflection*, for situations in which light bounces off matter, and the *law of refraction*, for situations in which light passes through matter. We will examine more about each of these laws in upcoming sections of this chapter.

Summary

- The speed of light in a vacuum is
 $c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}$.
- The index of refraction of a material is $n = c/v$, where v is the speed of light in a material and c is the speed of light in a vacuum.
- The ray model of light describes the path of light as straight lines. The part of optics dealing with the ray aspect of light is called geometric optics.

- Light can travel in three ways from a source to another location: (1) directly from the source through empty space; (2) through various media; and (3) after being reflected from a mirror.

Conceptual Questions

Exercise:

Problem:

Under what conditions can light be modeled like a ray? Like a wave?

Solution:

Light can be modeled as a ray when devices are large compared to wavelength, and as a wave when devices are comparable or small compared to wavelength.

Exercise:

Problem:

Why is the index of refraction always greater than or equal to 1?

Exercise:

Problem:

Does the fact that the light flash from lightning reaches you before its sound prove that the speed of light is extremely large or simply that it is greater than the speed of sound? Discuss how you could use this effect to get an estimate of the speed of light.

Solution:

This fact simply proves that the speed of light is greater than that of sound. If one knows the distance to the location of the lightning and the speed of sound, one could, in principle, determine the speed of light from the data. In practice, because the speed of light is so great, the data would have to be known to impractically high precision.

Exercise:**Problem:**

Speculate as to what physical process might be responsible for light traveling more slowly in a medium than in a vacuum.

Problems**Exercise:**

Problem: What is the speed of light in water? In glycerine?

Exercise:

Problem: What is the speed of light in air? In crown glass?

Solution:

$$2.99705 \times 10^8 \text{ m/s}; 1.97 \times 10^8 \text{ m/s}$$

Exercise:**Problem:**

Calculate the index of refraction for a medium in which the speed of light is $2.012 \times 10^8 \text{ m/s}$, and identify the most likely substance based on [\[link\]](#).

Exercise:**Problem:**

In what substance in [\[link\]](#) is the speed of light $2.290 \times 10^8 \text{ m/s}$?

Solution:

ice at 0°C

Exercise:

Problem:

There was a major collision of an asteroid with the Moon in medieval times. It was described by monks at Canterbury Cathedral in England as a red glow on and around the Moon. How long after the asteroid hit the Moon, which is 3.84×10^5 km away, would the light first arrive on Earth?

Exercise:**Problem:**

Components of some computers communicate with each other through optical fibers having an index of refraction $n = 1.55$. What time in nanoseconds is required for a signal to travel 0.200 m through such a fiber?

Solution:

1.03 ns

Exercise:**Problem:**

Compare the time it takes for light to travel 1000 m on the surface of Earth and in outer space.

Exercise:**Problem:**

How far does light travel underwater during a time interval of 1.50×10^{-6} s?

Solution:

337 m

Glossary

geometric optics

part of optics dealing with the ray aspect of light

index of refraction

for a material, the ratio of the speed of light in a vacuum to that in a material

ray

straight line that originates at some point

The Law of Reflection

By the end of this section, you will be able to:

- Explain the reflection of light from polished and rough surfaces
- Describe the principle and applications of corner reflectors

Whenever we look into a mirror, or squint at sunlight glinting from a lake, we are seeing a reflection. When you look at a piece of white paper, you are seeing light scattered from it. Large telescopes use reflection to form an image of stars and other astronomical objects.

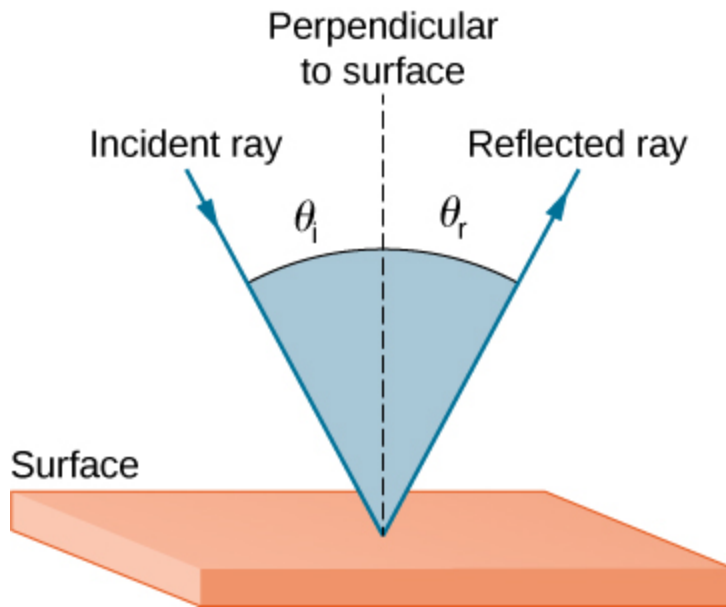
The **law of reflection** states that the angle of reflection equals the angle of incidence, or

Note:

Equation:

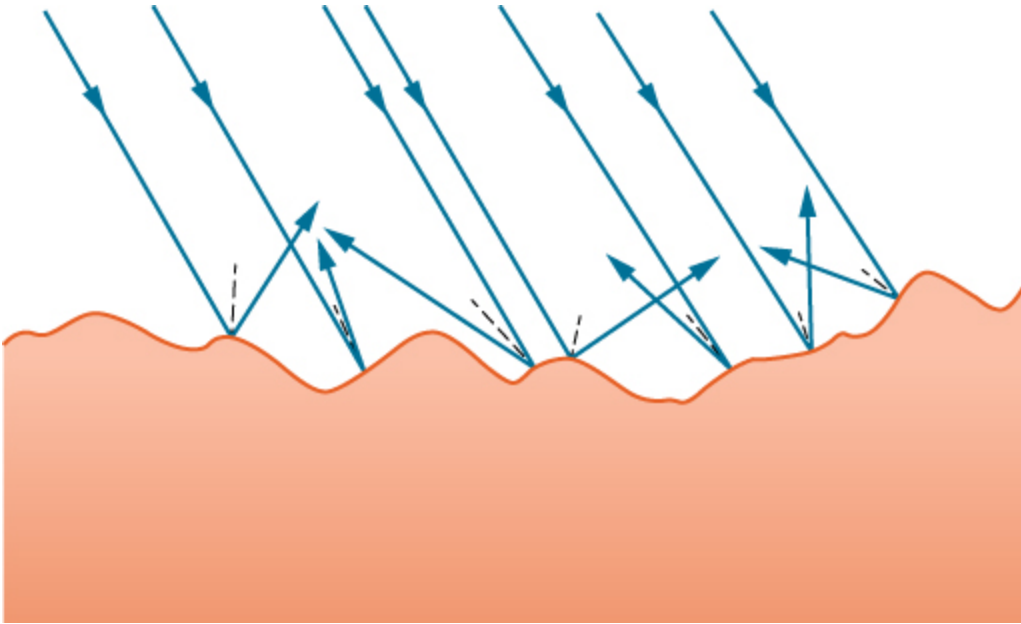
$$\theta_r = \theta_i$$

The law of reflection is illustrated in [\[link\]](#), which also shows how the angle of incidence and angle of reflection are measured relative to the perpendicular to the surface at the point where the light ray strikes.



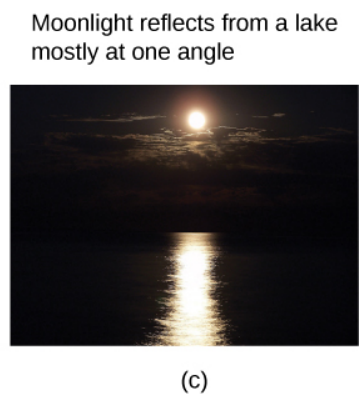
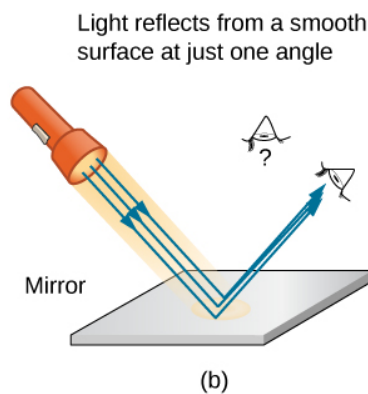
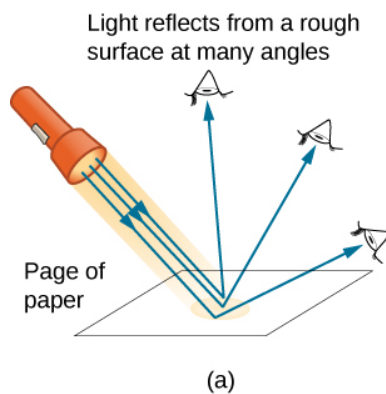
The law of reflection states that the angle of reflection equals the angle of incidence— $\theta_r = \theta_i$. The angles are measured relative to the perpendicular to the surface at the point where the ray strikes the surface.

We expect to see reflections from smooth surfaces, but [\[link\]](#) illustrates how a rough surface reflects light. Since the light strikes different parts of the surface at different angles, it is reflected in many different directions, or diffused. Diffused light is what allows us to see a sheet of paper from any angle, as shown in [\[link\]](#)(a). People, clothing, leaves, and walls all have rough surfaces and can be seen from all sides. A mirror, on the other hand, has a smooth surface (compared with the wavelength of light) and reflects light at specific angles, as illustrated in [\[link\]](#)(b). When the Moon reflects from a lake, as shown in [\[link\]](#)(c), a combination of these effects takes place.



Light is diffused when it reflects from a rough surface.

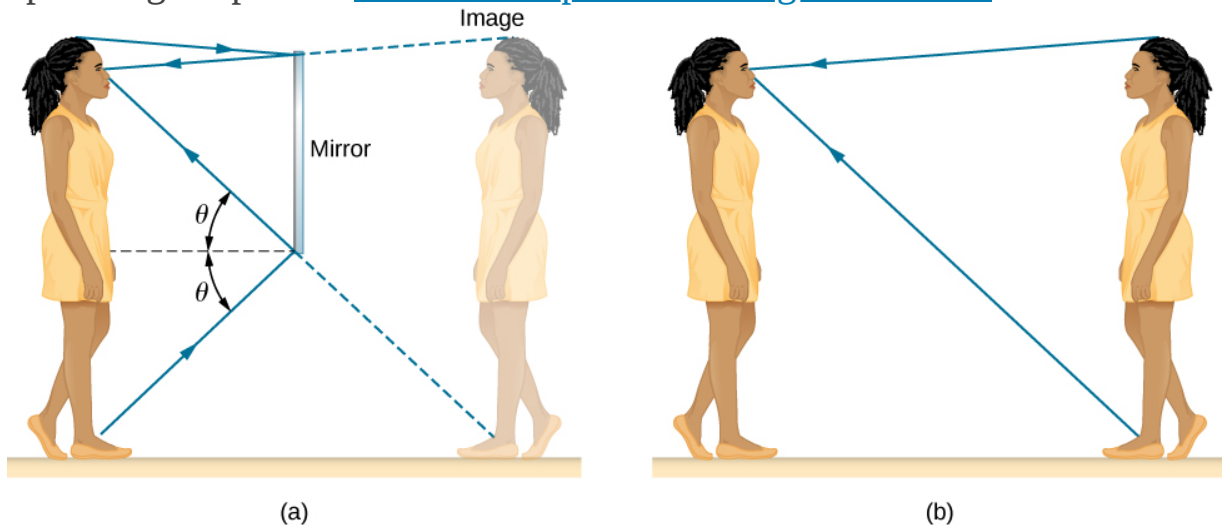
Here, many parallel rays are incident, but they are reflected at many different angles, because the surface is rough.



- (a) When a sheet of paper is illuminated with many parallel incident rays, it can be seen at many different angles, because its surface is rough and diffuses the light. (b) A mirror illuminated by many parallel rays reflects them in only one direction, because its surface is very smooth. Only the observer at a particular angle sees the reflected light. (c) Moonlight is spread out when it is reflected by the lake, because

the surface is shiny but uneven. (credit c: modification of work by Diego Torres Silvestre)

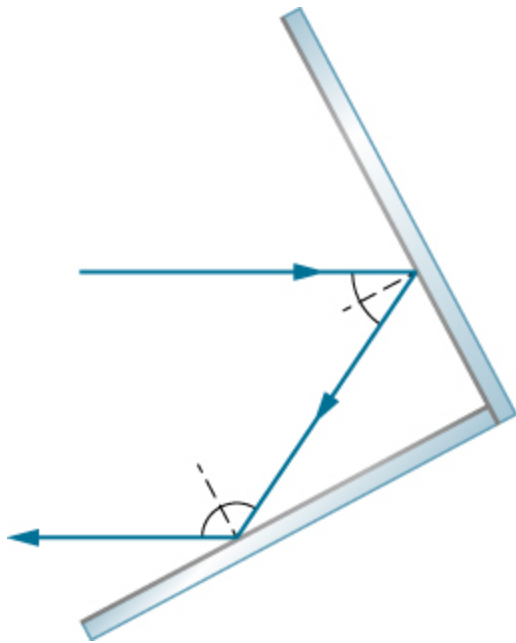
When you see yourself in a mirror, it appears that the image is actually behind the mirror ([link](#)). We see the light coming from a direction determined by the law of reflection. The angles are such that the image is exactly the same distance behind the mirror as you stand in front of the mirror. If the mirror is on the wall of a room, the images in it are all behind the mirror, which can make the room seem bigger. Although these mirror images make objects appear to be where they cannot be (like behind a solid wall), the images are not figments of your imagination. Mirror images can be photographed and videotaped by instruments and look just as they do with our eyes (which are optical instruments themselves). The precise manner in which images are formed by mirrors and lenses is discussed in an upcoming chapter on [Geometric Optics and Image Formation](#).



(a) Your image in a mirror is behind the mirror. The two rays shown are those that strike the mirror at just the correct angles to be reflected into the eyes of the person. The image appears to be behind the mirror at the same distance away as (b) if you were looking at your twin directly, with no mirror.

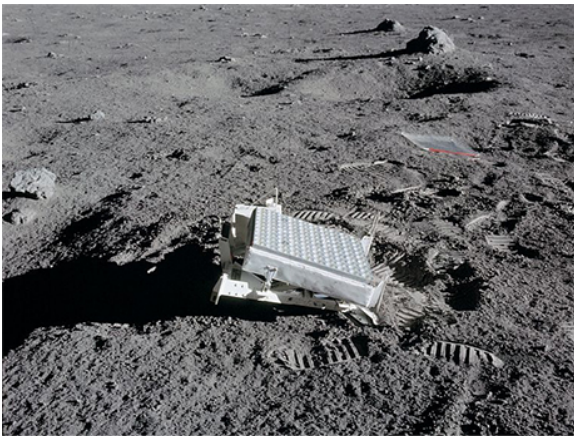
Corner Reflectors (Retroreflectors)

A light ray that strikes an object consisting of two mutually perpendicular reflecting surfaces is reflected back exactly parallel to the direction from which it came ([link](#)). This is true whenever the reflecting surfaces are perpendicular, and it is independent of the angle of incidence. (For proof, see [link](#) at the end of this section.) Such an object is called a **corner reflector**, since the light bounces from its inside corner. Corner reflectors are a subclass of retroreflectors, which all reflect rays back in the directions from which they came. Although the geometry of the proof is much more complex, corner reflectors can also be built with three mutually perpendicular reflecting surfaces and are useful in three-dimensional applications.



A light ray that strikes two mutually perpendicular reflecting surfaces is reflected back exactly parallel to the direction from which it came.

Many inexpensive reflector buttons on bicycles, cars, and warning signs have corner reflectors designed to return light in the direction from which it originated. Rather than simply reflecting light over a wide angle, retroreflection ensures high visibility if the observer and the light source are located together, such as a car's driver and headlights. The Apollo astronauts placed a true corner reflector on the Moon ([link](#)). Laser signals from Earth can be bounced from that corner reflector to measure the gradually increasing distance to the Moon of a few centimeters per year.



(a)



(b)

(a) Astronauts placed a corner reflector on the Moon to measure its gradually increasing orbital distance. (b) The bright spots on these bicycle safety reflectors are reflections of the flash of the camera that took this picture on a dark night. (credit a: modification of work by NASA; credit b: modification of work by “Julo”/Wikimedia Commons)

Working on the same principle as these optical reflectors, corner reflectors are routinely used as radar reflectors ([link](#)) for radio-frequency applications. Under most circumstances, small boats made of fiberglass or wood do not strongly reflect radio waves emitted by radar systems. To

make these boats visible to radar (to avoid collisions, for example), radar reflectors are attached to boats, usually in high places.



A radar reflector hoisted on a sailboat is a type of corner reflector. (credit: Tim Sheerman-Chase)

As a counterexample, if you are interested in building a stealth airplane, radar reflections should be minimized to evade detection. One of the design considerations would then be to avoid building 90° corners into the airframe.

Summary

- When a light ray strikes a smooth surface, the angle of reflection equals the angle of incidence.
- A mirror has a smooth surface and reflects light at specific angles.
- Light is diffused when it reflects from a rough surface.

Conceptual Questions

Exercise:

Problem:

Using the law of reflection, explain how powder takes the shine off of a person's nose. What is the name of the optical effect?

Solution:

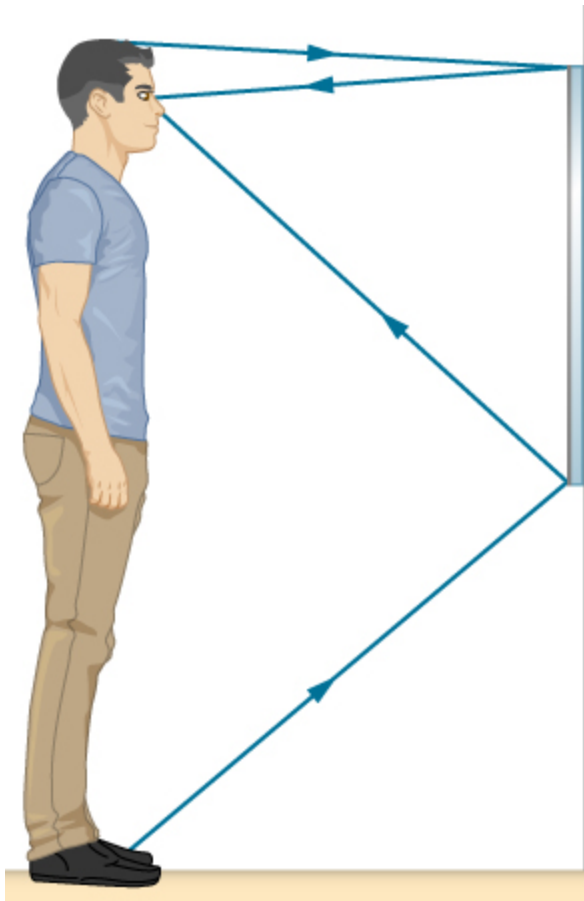
Powder consists of many small particles with randomly oriented surfaces. This leads to diffuse reflection, reducing shine.

Problems

Exercise:

Problem:

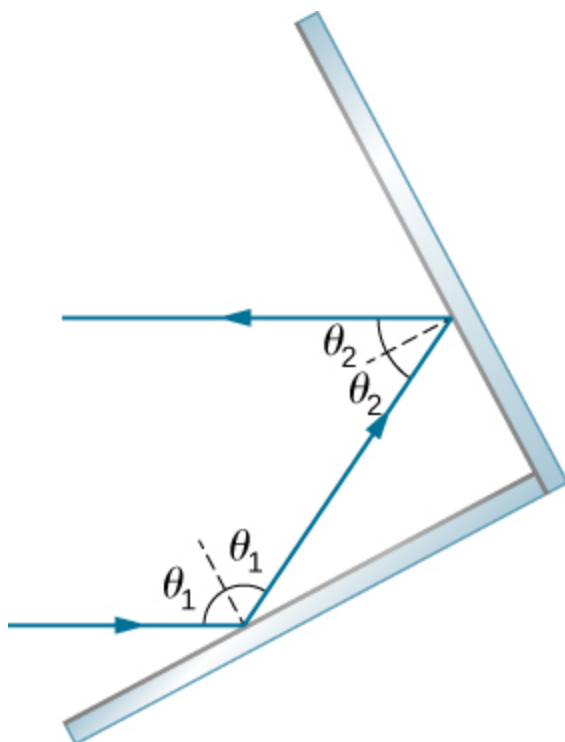
Suppose a man stands in front of a mirror as shown below. His eyes are 1.65 m above the floor and the top of his head is 0.13 m higher. Find the height above the floor of the top and bottom of the smallest mirror in which he can see both the top of his head and his feet. How is this distance related to the man's height?



Exercise:

Problem:

Show that when light reflects from two mirrors that meet each other at a right angle, the outgoing ray is parallel to the incoming ray, as illustrated below.



Solution:

proof

Exercise:

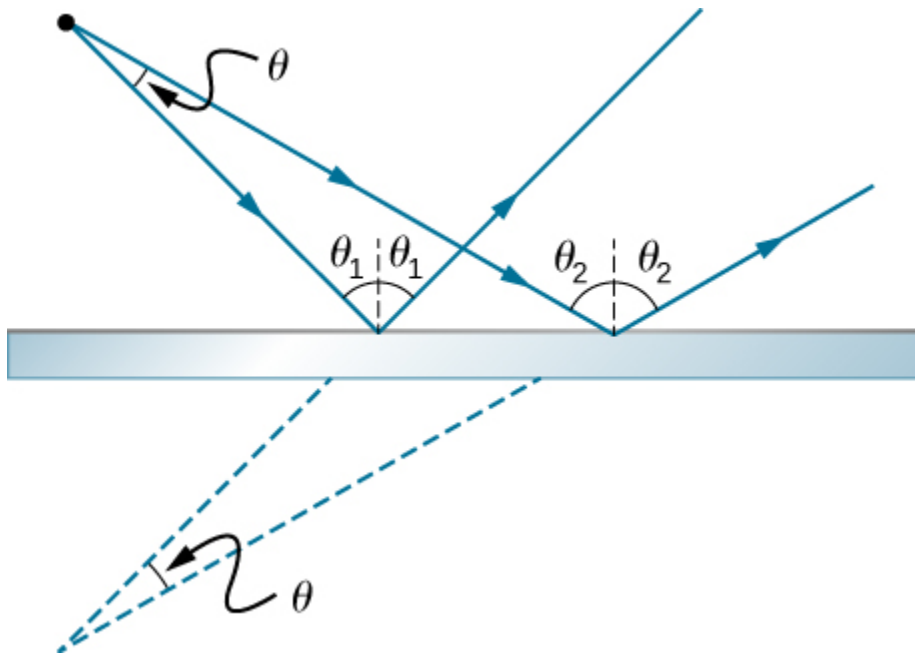
Problem:

On the Moon's surface, lunar astronauts placed a corner reflector, off which a laser beam is periodically reflected. The distance to the Moon is calculated from the round-trip time. What percent correction is needed to account for the delay in time due to the slowing of light in Earth's atmosphere? Assume the distance to the Moon is precisely 3.84×10^8 m and Earth's atmosphere (which varies in density with altitude) is equivalent to a layer 30.0 km thick with a constant index of refraction $n = 1.000293$.

Exercise:

Problem:

A flat mirror is neither converging nor diverging. To prove this, consider two rays originating from the same point and diverging at an angle θ (see below). Show that after striking a plane mirror, the angle between their directions remains θ .



Solution:

proof

Glossary

corner reflector

object consisting of two (or three) mutually perpendicular reflecting surfaces, so that the light that enters is reflected back exactly parallel to the direction from which it came

law of reflection

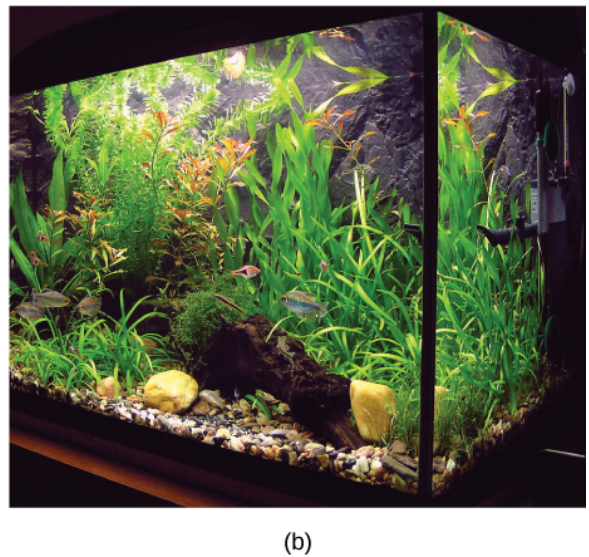
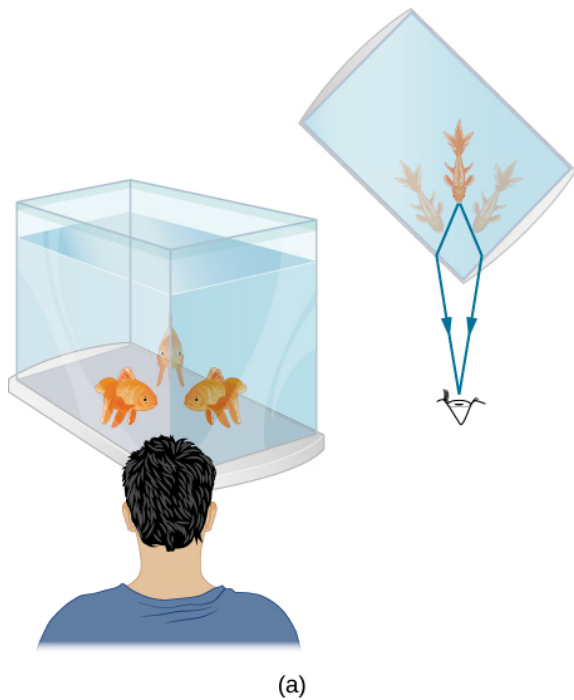
angle of reflection equals the angle of incidence

Refraction

By the end of this section, you will be able to:

- Describe how rays change direction upon entering a medium
- Apply the law of refraction in problem solving

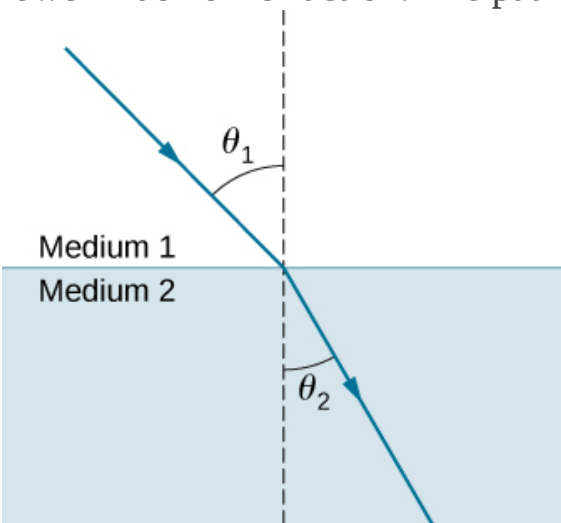
You may often notice some odd things when looking into a fish tank. For example, you may see the same fish appearing to be in two different places ([link](#)). This happens because light coming from the fish to you changes direction when it leaves the tank, and in this case, it can travel two different paths to get to your eyes. The changing of a light ray's direction (loosely called bending) when it passes through substances of different refractive indices is called **refraction** and is related to changes in the speed of light, $v = c/n$. Refraction is responsible for a tremendous range of optical phenomena, from the action of lenses to data transmission through optical fibers.



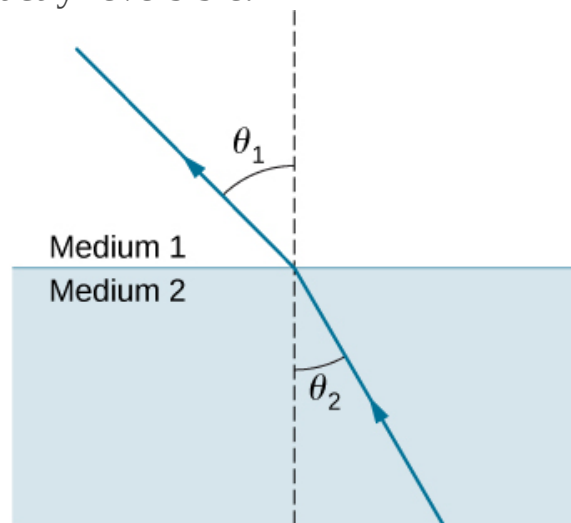
(a) Looking at the fish tank as shown, we can see the same fish in two different locations, because light changes directions when it passes from water to air. In this case, the light can reach the observer by two different paths, so the fish seems to be in two different places. This

bending of light is called refraction and is responsible for many optical phenomena. (b) This image shows refraction of light from a fish near the top of a fish tank.

[\[link\]](#) shows how a ray of light changes direction when it passes from one medium to another. As before, the angles are measured relative to a perpendicular to the surface at the point where the light ray crosses it. (Some of the incident light is reflected from the surface, but for now we concentrate on the light that is transmitted.) The change in direction of the light ray depends on the relative values of the indices of refraction ([The Propagation of Light](#)) of the two media involved. In the situations shown, medium 2 has a greater index of refraction than medium 1. Note that as shown in [\[link\]](#)(a), the direction of the ray moves closer to the perpendicular when it progresses from a medium with a lower index of refraction to one with a higher index of refraction. Conversely, as shown in [\[link\]](#)(b), the direction of the ray moves away from the perpendicular when it progresses from a medium with a higher index of refraction to one with a lower index of refraction. The path is exactly reversible.



(a)



(b)

The change in direction of a light ray depends on how the index of refraction changes when it crosses from one medium to another. In the situations shown here, the index of refraction is greater in medium 2

than in medium 1. (a) A ray of light moves closer to the perpendicular when entering a medium with a higher index of refraction. (b) A ray of light moves away from the perpendicular when entering a medium with a lower index of refraction.

The amount that a light ray changes its direction depends both on the incident angle and the amount that the speed changes. For a ray at a given incident angle, a large change in speed causes a large change in direction and thus a large change in angle. The exact mathematical relationship is the **law of refraction**, or Snell's law, after the Dutch mathematician Willebrord Snell (1591–1626), who discovered it in 1621. While the law has been named after Snell, the Arabian physicist Ibn Sahl found the law of refraction in 984 and used it in his work *On Burning Mirrors and Lenses*. The law of refraction is stated in equation form as

Note:

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

Here n_1 and n_2 are the indices of refraction for media 1 and 2, and θ_1 and θ_2 are the angles between the rays and the perpendicular in media 1 and 2. The incoming ray is called the incident ray, the outgoing ray is called the refracted ray, and the associated angles are the incident angle and the refracted angle, respectively.

Snell's experiments showed that the law of refraction is obeyed and that a characteristic index of refraction n could be assigned to a given medium and its value measured. Snell was not aware that the speed of light varied in different media, a key fact used when we derive the law of refraction theoretically using Huygens's principle in [Huygens's Principle](#).

Example:**Determining the Index of Refraction**

Find the index of refraction for medium 2 in [\[link\]](#)(a), assuming medium 1 is air and given that the incident angle is 30.0° and the angle of refraction is 22.0° .

Strategy

The index of refraction for air is taken to be 1 in most cases (and up to four significant figures, it is 1.000). Thus, $n_1 = 1.00$ here. From the given information, $\theta_1 = 30.0^\circ$ and $\theta_2 = 22.0^\circ$. With this information, the only unknown in Snell's law is n_2 , so we can use Snell's law to find it.

Solution

From Snell's law we have

Equation:

$$\begin{aligned}n_1 \sin \theta_1 &= n_2 \sin \theta_2 \\n_2 &= n_1 \frac{\sin \theta_1}{\sin \theta_2}.\end{aligned}$$

Entering known values,

Equation:

$$n_2 = 1.00 \frac{\sin 30.0^\circ}{\sin 22.0^\circ} = \frac{0.500}{0.375} = 1.33.$$

Significance

This is the index of refraction for water, and Snell could have determined it by measuring the angles and performing this calculation. He would then have found 1.33 to be the appropriate index of refraction for water in all other situations, such as when a ray passes from water to glass. Today, we can verify that the index of refraction is related to the speed of light in a medium by measuring that speed directly.

Note:

Explore [bending of light](#) between two media with different indices of refraction. Use the “Intro” simulation and see how changing from air to water to glass changes the bending angle. Use the protractor tool to

measure the angles and see if you can recreate the configuration in [\[link\]](#). Also by measurement, confirm that the angle of reflection equals the angle of incidence.

Example:**A Larger Change in Direction**

Suppose that in a situation like that in [\[link\]](#), light goes from air to diamond and that the incident angle is 30.0° . Calculate the angle of refraction θ_2 in the diamond.

Strategy

Again, the index of refraction for air is taken to be $n_1 = 1.00$, and we are given $\theta_1 = 30.0^\circ$. We can look up the index of refraction for diamond in [\[link\]](#), finding $n_2 = 2.419$. The only unknown in Snell's law is θ_2 , which we wish to determine.

Solution

Solving Snell's law for $\sin \theta_2$ yields

Equation:

$$\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1.$$

Entering known values,

Equation:

$$\sin \theta_2 = \frac{1.00}{2.419} \sin 30.0^\circ = (0.413)(0.500) = 0.207.$$

The angle is thus

Equation:

$$\theta_2 = \sin^{-1}(0.207) = 11.9^\circ.$$

Significance

For the same 30.0° angle of incidence, the angle of refraction in diamond is significantly smaller than in water (11.9° rather than 22.0° —see [\[link\]](#)). This means there is a larger change in direction in diamond. The cause of a

large change in direction is a large change in the index of refraction (or speed). In general, the larger the change in speed, the greater the effect on the direction of the ray.

Note:

Exercise:

Problem:

Check Your Understanding In [\[link\]](#), the solid with the next highest index of refraction after diamond is zircon. If the diamond in [\[link\]](#) were replaced with a piece of zircon, what would be the new angle of refraction?

Solution:

15.1°

Summary

- The change of a light ray's direction when it passes through variations in matter is called refraction.
- The law of refraction, also called Snell's law, relates the indices of refraction for two media at an interface to the change in angle of a light ray passing through that interface.

Conceptual Questions

Exercise:

Problem:

Diffusion by reflection from a rough surface is described in this chapter. Light can also be diffused by refraction. Describe how this occurs in a specific situation, such as light interacting with crushed ice.

Exercise:**Problem:**

Will light change direction toward or away from the perpendicular when it goes from air to water? Water to glass? Glass to air?

Solution:

“toward” when increasing n (air to water, water to glass); “away” when decreasing n (glass to air)

Exercise:**Problem:**

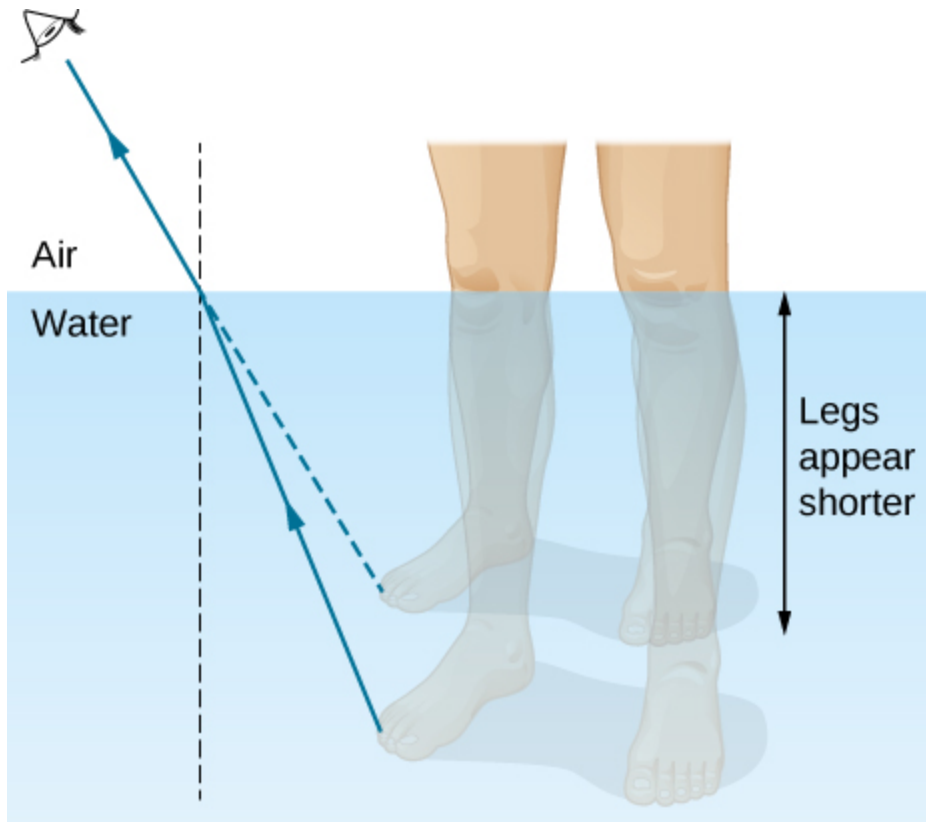
Explain why an object in water always appears to be at a depth shallower than it actually is?

Exercise:**Problem:**

Explain why a person’s legs appear very short when wading in a pool. Justify your explanation with a ray diagram showing the path of rays from the feet to the eye of an observer who is out of the water.

Solution:

A ray from a leg emerges from water after refraction. The observer in air perceives an apparent location for the source, as if a ray traveled in a straight line. See the dashed ray below.



Exercise:

Problem:

Explain why an oar that is partially submerged in water appears bent.

Problems

Unless otherwise specified, for problems 1 through 10, the indices of refraction of glass and water should be taken to be 1.50 and 1.333, respectively.

Exercise:

Problem:

A light beam in air has an angle of incidence of 35° at the surface of a glass plate. What are the angles of reflection and refraction?

Exercise:

Problem:

A light beam in air is incident on the surface of a pond, making an angle of 20° with respect to the surface. What are the angles of reflection and refraction?

Solution:

reflection, 70° ; refraction, 45°

Exercise:**Problem:**

When a light ray crosses from water into glass, it emerges at an angle of 30° with respect to the normal of the interface. What is its angle of incidence?

Exercise:**Problem:**

A pencil flashlight submerged in water sends a light beam toward the surface at an angle of incidence of 30° . What is the angle of refraction in air?

Solution:

42°

Exercise:**Problem:**

Light rays from the Sun make a 30° angle to the vertical when seen from below the surface of a body of water. At what angle above the horizon is the Sun?

Exercise:

Problem:

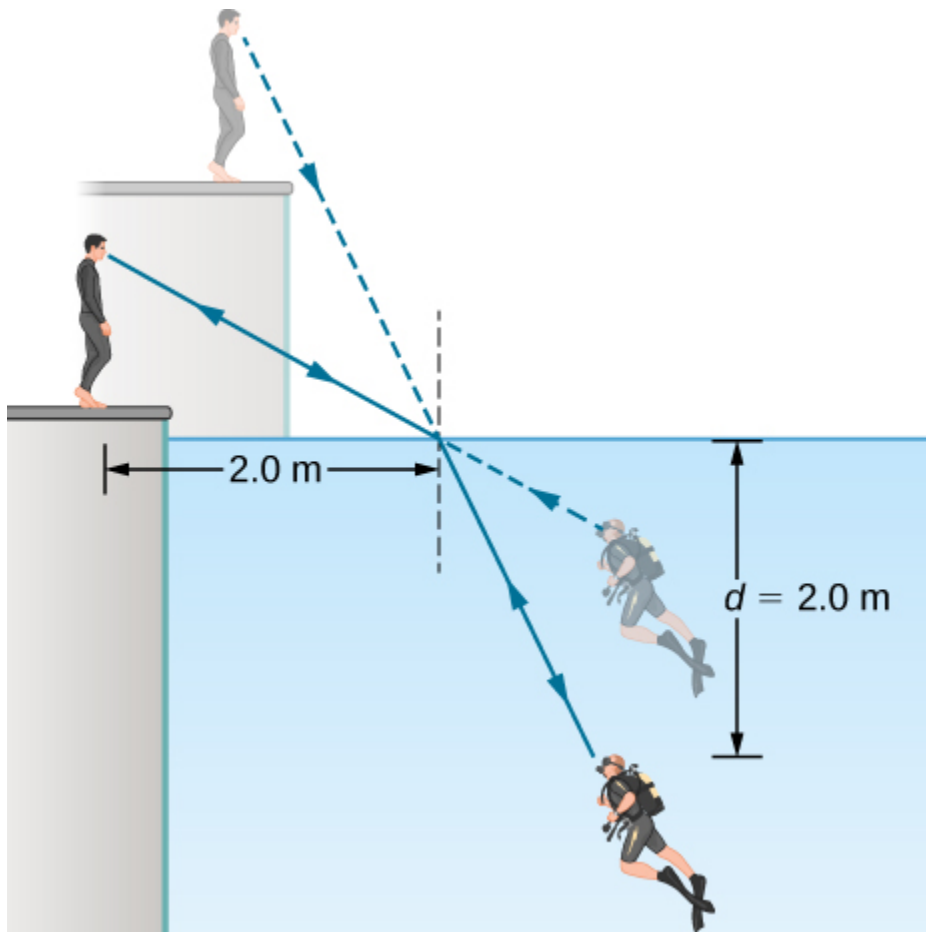
The path of a light beam in air goes from an angle of incidence of 35° to an angle of refraction of 22° when it enters a rectangular block of plastic. What is the index of refraction of the plastic?

Solution:

1.53

Exercise:**Problem:**

A scuba diver training in a pool looks at his instructor as shown below. What angle does the ray from the instructor's face make with the perpendicular to the water at the point where the ray enters? The angle between the ray in the water and the perpendicular to the water is 25.0° .

**Exercise:****Problem:**

(a) Using information in the preceding problem, find the height of the instructor's head above the water, noting that you will first have to calculate the angle of incidence. (b) Find the apparent depth of the diver's head below water as seen by the instructor.

Solution:

a. 2.9 m; b. 1.4 m

Glossary

law of refraction

when a light ray crosses from one medium to another, it changes direction by an amount that depends on the index of refraction of each medium and the sines of the angle of incidence and angle of refraction

refraction

changing of a light ray's direction when it passes through variations in matter

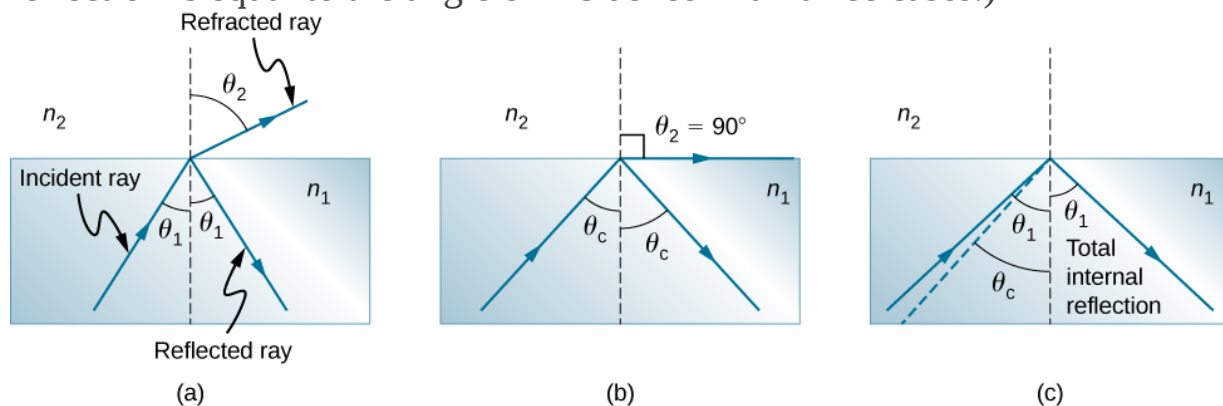
Total Internal Reflection

By the end of this section, you will be able to:

- Explain the phenomenon of total internal reflection
- Describe the workings and uses of optical fibers
- Analyze the reason for the sparkle of diamonds

A good-quality mirror may reflect more than 90% of the light that falls on it, absorbing the rest. But it would be useful to have a mirror that reflects all of the light that falls on it. Interestingly, we can produce total reflection using an aspect of refraction.

Consider what happens when a ray of light strikes the surface between two materials, as shown in [\[link\]](#)(a). Part of the light crosses the boundary and is refracted; the rest is reflected. If, as shown in the figure, the index of refraction for the second medium is less than for the first, the ray bends away from the perpendicular. (Since $n_1 > n_2$, the angle of refraction is greater than the angle of incidence—that is, $\theta_2 > \theta_1$.) Now imagine what happens as the incident angle increases. This causes θ_2 to increase also. The largest the angle of refraction θ_2 can be is 90° , as shown in part (b). The **critical angle** θ_c for a combination of materials is defined to be the incident angle θ_1 that produces an angle of refraction of 90° . That is, θ_c is the incident angle for which $\theta_2 = 90^\circ$. If the incident angle θ_1 is greater than the critical angle, as shown in [\[link\]](#)(c), then all of the light is reflected back into medium 1, a condition called **total internal reflection**. (As the figure shows, the reflected rays obey the law of reflection so that the angle of reflection is equal to the angle of incidence in all three cases.)



(a) A ray of light crosses a boundary where the index of refraction decreases. That is, $n_2 < n_1$. The ray bends away from the perpendicular. (b) The critical angle θ_c is the angle of incidence for which the angle of refraction is 90° . (c) Total internal reflection occurs when the incident angle is greater than the critical angle.

Snell's law states the relationship between angles and indices of refraction. It is given by

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

When the incident angle equals the critical angle ($\theta_1 = \theta_c$), the angle of refraction is 90° ($\theta_2 = 90^\circ$). Noting that $\sin 90^\circ = 1$, Snell's law in this case becomes

Equation:

$$n_1 \sin \theta_1 = n_2.$$

The critical angle θ_c for a given combination of materials is thus

Note:

Equation:

$$\theta_c = \sin^{-1} \left(\frac{n_2}{n_1} \right) \text{ for } n_1 > n_2.$$

Total internal reflection occurs for any incident angle greater than the critical angle θ_c , and it can only occur when the second medium has an index of refraction less than the first. Note that this equation is written for a

light ray that travels in medium 1 and reflects from medium 2, as shown in [\[link\]](#).

Example:**Determining a Critical Angle**

What is the critical angle for light traveling in a polystyrene (a type of plastic) pipe surrounded by air? The index of refraction for polystyrene is 1.49.

Strategy

The index of refraction of air can be taken to be 1.00, as before. Thus, the condition that the second medium (air) has an index of refraction less than the first (plastic) is satisfied, and we can use the equation

Equation:

$$\theta_c = \sin^{-1} \left(\frac{n_2}{n_1} \right)$$

to find the critical angle θ_c , where $n_2 = 1.00$ and $n_1 = 1.49$.

Solution

Substituting the identified values gives

Equation:

$$\theta_c = \sin^{-1} \left(\frac{1.00}{1.49} \right) = \sin^{-1}(0.671) = 42.2^\circ.$$

Significance

This result means that any ray of light inside the plastic that strikes the surface at an angle greater than 42.2° is totally reflected. This makes the inside surface of the clear plastic a perfect mirror for such rays, without any need for the silvering used on common mirrors. Different combinations of materials have different critical angles, but any combination with $n_1 > n_2$ can produce total internal reflection. The same calculation as made here shows that the critical angle for a ray going from water to air is 48.6° , whereas that from diamond to air is 24.4° , and that from flint glass to crown glass is 66.3° .

Note:**Exercise:****Problem:**

Check Your Understanding At the surface between air and water, light rays can go from air to water and from water to air. For which ray is there no possibility of total internal reflection?

Solution:

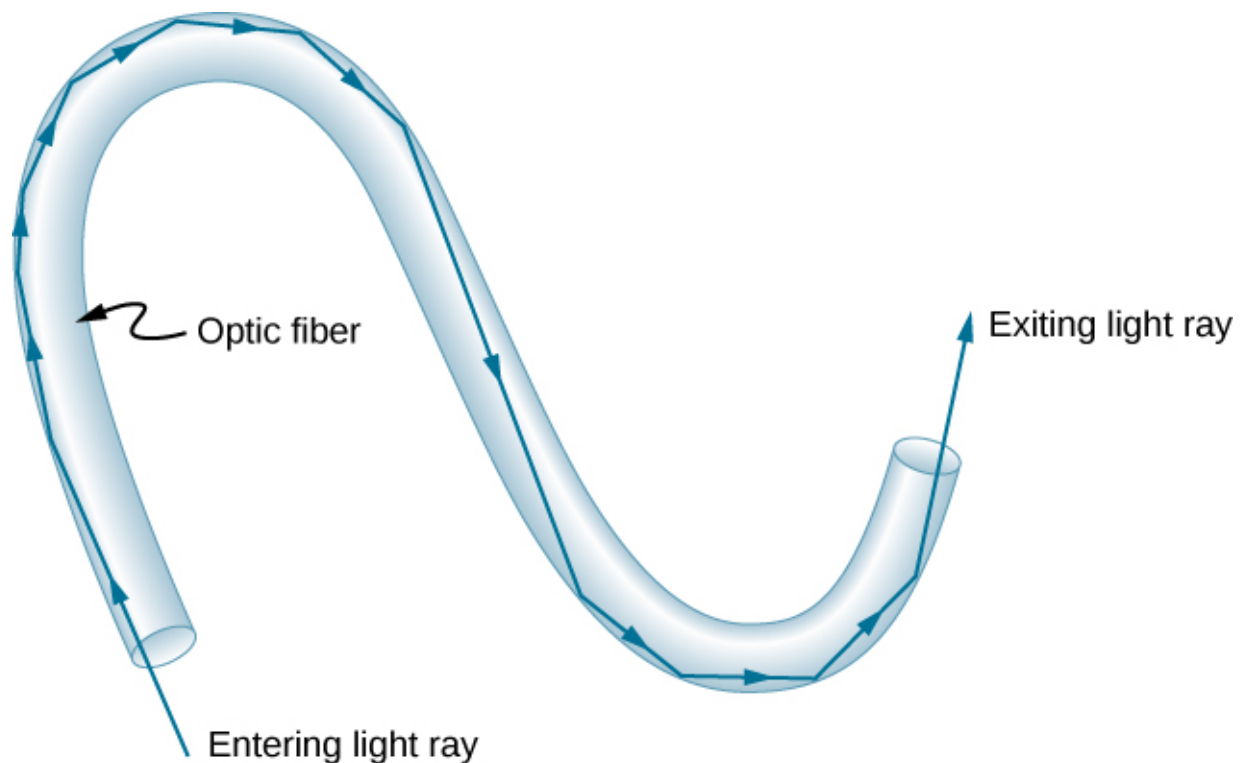
air to water, because the condition that the second medium must have a smaller index of refraction is not satisfied

In the photo that opens this chapter, the image of a swimmer underwater is captured by a camera that is also underwater. The swimmer in the upper half of the photograph, apparently facing upward, is, in fact, a reflected image of the swimmer below. The circular ripple near the photograph's center is actually on the water surface. The undisturbed water surrounding it makes a good reflecting surface when viewed from below, thanks to total internal reflection. However, at the very top edge of this photograph, rays from below strike the surface with incident angles less than the critical angle, allowing the camera to capture a view of activities on the pool deck above water.

Fiber Optics: Endoscopes to Telephones

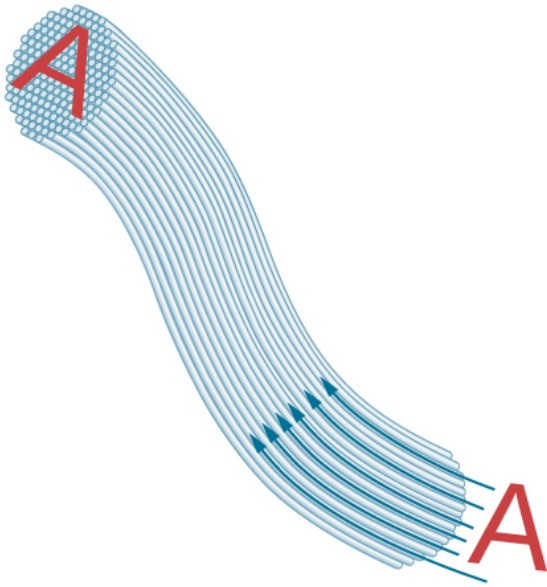
Fiber optics is one application of total internal reflection that is in wide use. In communications, it is used to transmit telephone, internet, and cable TV signals. **Fiber optics** employs the transmission of light down fibers of plastic or glass. Because the fibers are thin, light entering one is likely to strike the inside surface at an angle greater than the critical angle and, thus, be totally reflected ([\[link\]](#)). The index of refraction outside the fiber must be smaller than inside. In fact, most fibers have a varying refractive index to allow more light to be guided along the fiber through total internal

refraction. Rays are reflected around corners as shown, making the fibers into tiny light pipes.

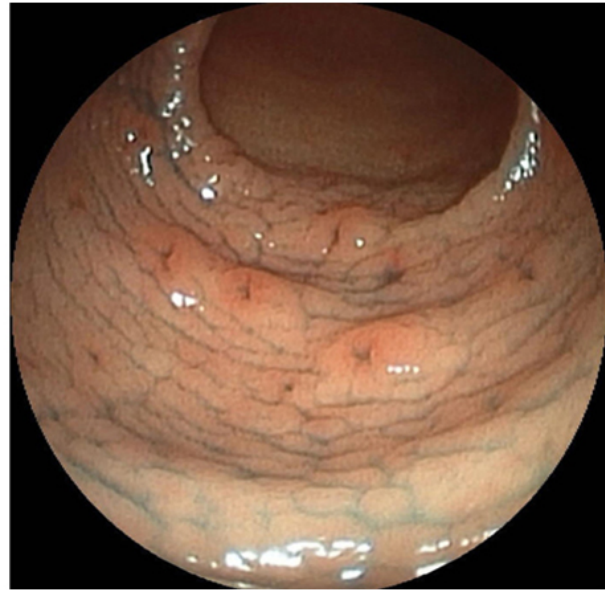


Light entering a thin optic fiber may strike the inside surface at large or grazing angles and is completely reflected if these angles exceed the critical angle. Such rays continue down the fiber, even following it around corners, since the angles of reflection and incidence remain large.

Bundles of fibers can be used to transmit an image without a lens, as illustrated in [\[link\]](#). The output of a device called an endoscope is shown in [\[link\]](#)(b). Endoscopes are used to explore the interior of the body through its natural orifices or minor incisions. Light is transmitted down one fiber bundle to illuminate internal parts, and the reflected light is transmitted back out through another bundle to be observed.



(a)

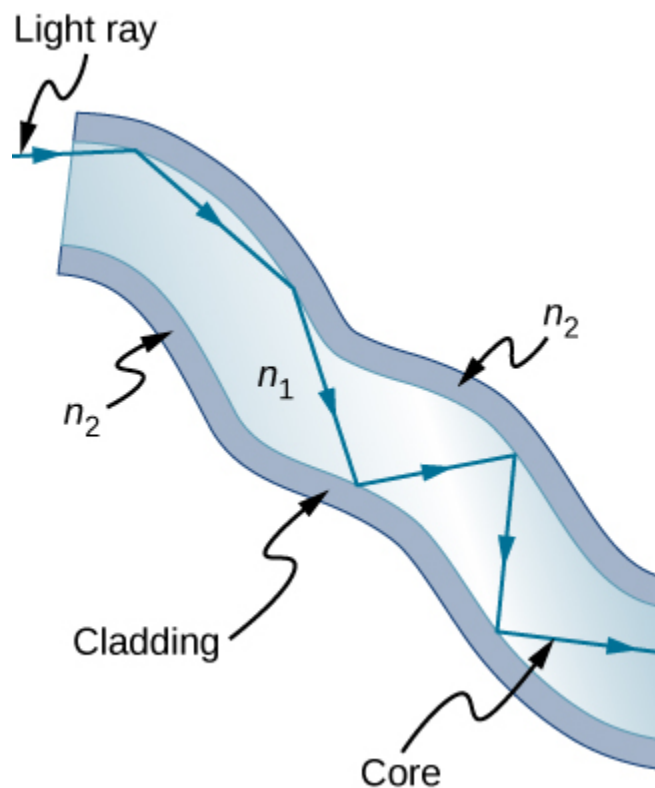


(b)

(a) An image “A” is transmitted by a bundle of optical fibers. (b) An endoscope is used to probe the body, both transmitting light to the interior and returning an image such as the one shown of a human epiglottis (a structure at the base of the tongue). (credit b: modification of work by “Med_Chaos”/Wikimedia Commons)

Fiber optics has revolutionized surgical techniques and observations within the body, with a host of medical diagnostic and therapeutic uses. Surgery can be performed, such as arthroscopic surgery on a knee or shoulder joint, employing cutting tools attached to and observed with the endoscope. Samples can also be obtained, such as by lassoing an intestinal polyp for external examination. The flexibility of the fiber optic bundle allows doctors to navigate it around small and difficult-to-reach regions in the body, such as the intestines, the heart, blood vessels, and joints. Transmission of an intense laser beam to burn away obstructing plaques in major arteries, as well as delivering light to activate chemotherapy drugs, are becoming commonplace. Optical fibers have in fact enabled microsurgery and remote surgery where the incisions are small and the surgeon’s fingers do not need to touch the diseased tissue.

Optical fibers in bundles are surrounded by a cladding material that has a lower index of refraction than the core ([link](#)). The cladding prevents light from being transmitted between fibers in a bundle. Without cladding, light could pass between fibers in contact, since their indices of refraction are identical. Since no light gets into the cladding (there is total internal reflection back into the core), none can be transmitted between clad fibers that are in contact with one another. Instead, the light is propagated along the length of the fiber, minimizing the loss of signal and ensuring that a quality image is formed at the other end. The cladding and an additional protective layer make optical fibers durable as well as flexible.



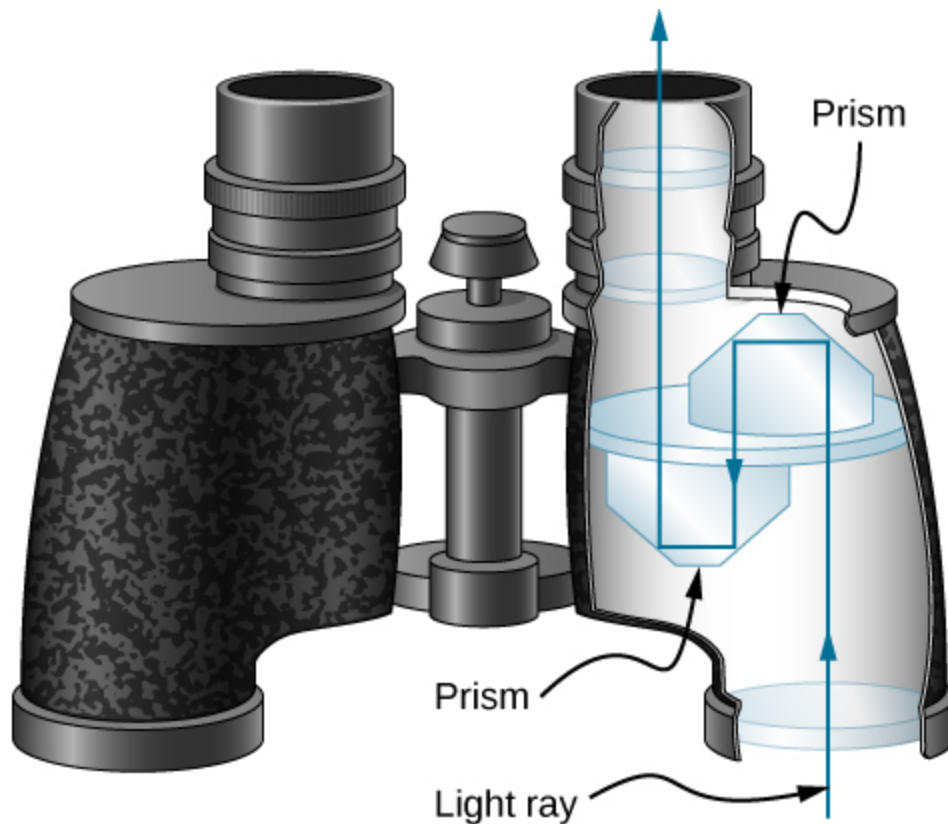
Fibers in bundles are clad by a material that has a lower index of refraction than the core to ensure total internal reflection, even when fibers are in contact with one another.

Special tiny lenses that can be attached to the ends of bundles of fibers have been designed and fabricated. Light emerging from a fiber bundle can be focused through such a lens, imaging a tiny spot. In some cases, the spot can be scanned, allowing quality imaging of a region inside the body. Special minute optical filters inserted at the end of the fiber bundle have the capacity to image the interior of organs located tens of microns below the surface without cutting the surface—an area known as noninvasive diagnostics. This is particularly useful for determining the extent of cancers in the stomach and bowel.

In another type of application, optical fibers are commonly used to carry signals for telephone conversations and internet communications. Extensive optical fiber cables have been placed on the ocean floor and underground to enable optical communications. Optical fiber communication systems offer several advantages over electrical (copper)-based systems, particularly for long distances. The fibers can be made so transparent that light can travel many kilometers before it becomes dim enough to require amplification—much superior to copper conductors. This property of optical fibers is called low loss. Lasers emit light with characteristics that allow far more conversations in one fiber than are possible with electric signals on a single conductor. This property of optical fibers is called high bandwidth. Optical signals in one fiber do not produce undesirable effects in other adjacent fibers. This property of optical fibers is called reduced crosstalk. We shall explore the unique characteristics of laser radiation in a later chapter.

Corner Reflectors and Diamonds

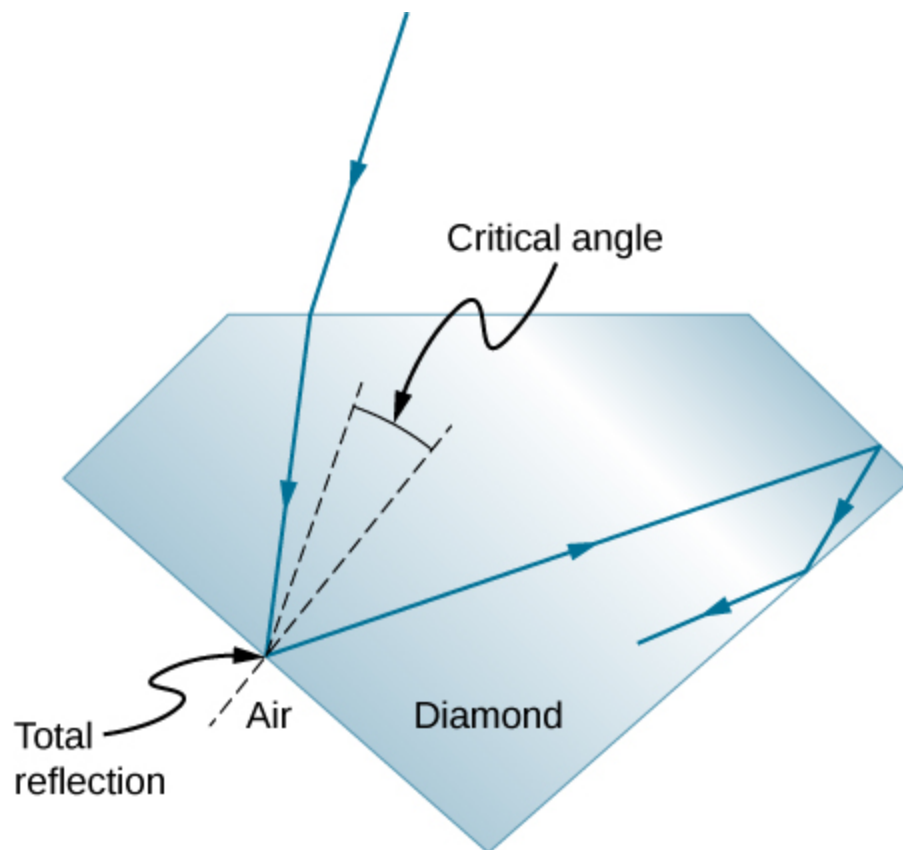
Corner reflectors ([The Law of Reflection](#)) are perfectly efficient when the conditions for total internal reflection are satisfied. With common materials, it is easy to obtain a critical angle that is less than 45° . One use of these perfect mirrors is in binoculars, as shown in [\[link\]](#). Another use is in periscopes found in submarines.



These binoculars employ corner reflectors (prisms) with total internal reflection to get light to the observer's eyes.

Total internal reflection, coupled with a large index of refraction, explains why diamonds sparkle more than other materials. The critical angle for a diamond-to-air surface is only 24.4° , so when light enters a diamond, it has trouble getting back out ([\[link\]](#)). Although light freely enters the diamond, it can exit only if it makes an angle less than 24.4° . Facets on diamonds are specifically intended to make this unlikely. Good diamonds are very clear, so that the light makes many internal reflections and is concentrated before exiting—hence the bright sparkle. (Zircon is a natural gemstone that has an exceptionally large index of refraction, but it is not as large as diamond, so it is not as highly prized. Cubic zirconia is manufactured and has an even higher index of refraction (≈ 2.17), but it is still less than that of diamond.) The colors you see emerging from a clear diamond are not due to the

diamond's color, which is usually nearly colorless. The colors result from dispersion, which we discuss in [Dispersion](#). Colored diamonds get their color from structural defects of the crystal lattice and the inclusion of minute quantities of graphite and other materials. The Argyle Mine in Western Australia produces around 90% of the world's pink, red, champagne, and cognac diamonds, whereas around 50% of the world's clear diamonds come from central and southern Africa.



Light cannot easily escape a diamond, because its critical angle with air is so small. Most reflections are total, and the facets are placed so that light can exit only in particular ways—thus concentrating the light and making the diamond sparkle brightly.

Note:

Explore [refraction and reflection of light](#) between two media with different indices of refraction. Try to make the refracted ray disappear with total internal reflection. Use the protractor tool to measure the critical angle and compare with the prediction from [\[link\]](#).

Summary

- The incident angle that produces an angle of refraction of 90° is called the critical angle.
- Total internal reflection is a phenomenon that occurs at the boundary between two media, such that if the incident angle in the first medium is greater than the critical angle, then all the light is reflected back into that medium.
- Fiber optics involves the transmission of light down fibers of plastic or glass, applying the principle of total internal reflection.
- Cladding prevents light from being transmitted between fibers in a bundle.
- Diamonds sparkle due to total internal reflection coupled with a large index of refraction.

Conceptual Questions

Exercise:**Problem:**

A ring with a colorless gemstone is dropped into water. The gemstone becomes invisible when submerged. Can it be a diamond? Explain.

Solution:

The gemstone becomes invisible when its index of refraction is the same, or at least similar to, the water surrounding it. Because diamond has a particularly high index of refraction, it can still sparkle as a result of total internal reflection, not invisible.

Exercise:**Problem:**

The most common type of mirage is an illusion that light from faraway objects is reflected by a pool of water that is not really there. Mirages are generally observed in deserts, when there is a hot layer of air near the ground. Given that the refractive index of air is lower for air at higher temperatures, explain how mirages can be formed.

Exercise:**Problem:**

How can you use total internal reflection to estimate the index of refraction of a medium?

Solution:

One can measure the critical angle by looking for the onset of total internal reflection as the angle of incidence is varied. [\[link\]](#) can then be applied to compute the index of refraction.

Problems**Exercise:****Problem:**

Verify that the critical angle for light going from water to air is 48.6° , as discussed at the end of [\[link\]](#), regarding the critical angle for light traveling in a polystyrene (a type of plastic) pipe surrounded by air.

Exercise:**Problem:**

(a) At the end of [\[link\]](#), it was stated that the critical angle for light going from diamond to air is 24.4° . Verify this. (b) What is the critical angle for light going from zircon to air?

Solution:

a. 24.42° ; b. 31.33°

Exercise:**Problem:**

An optical fiber uses flint glass clad with crown glass. What is the critical angle?

Exercise:**Problem:**

At what minimum angle will you get total internal reflection of light traveling in water and reflected from ice?

Solution:

79.11°

Exercise:**Problem:**

Suppose you are using total internal reflection to make an efficient corner reflector. If there is air outside and the incident angle is 45.0° , what must be the minimum index of refraction of the material from which the reflector is made?

Exercise:**Problem:**

You can determine the index of refraction of a substance by determining its critical angle. (a) What is the index of refraction of a substance that has a critical angle of 68.4° when submerged in water? What is the substance, based on [\[link\]](#)? (b) What would the critical angle be for this substance in air?

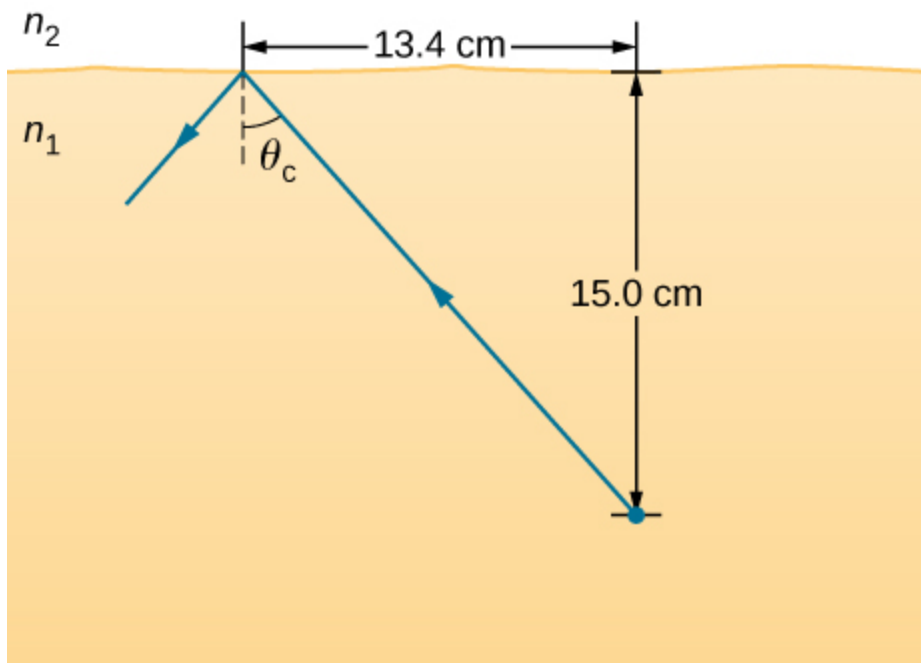
Solution:

a. 1.43, fluorite; b. 44.2°

Exercise:

Problem:

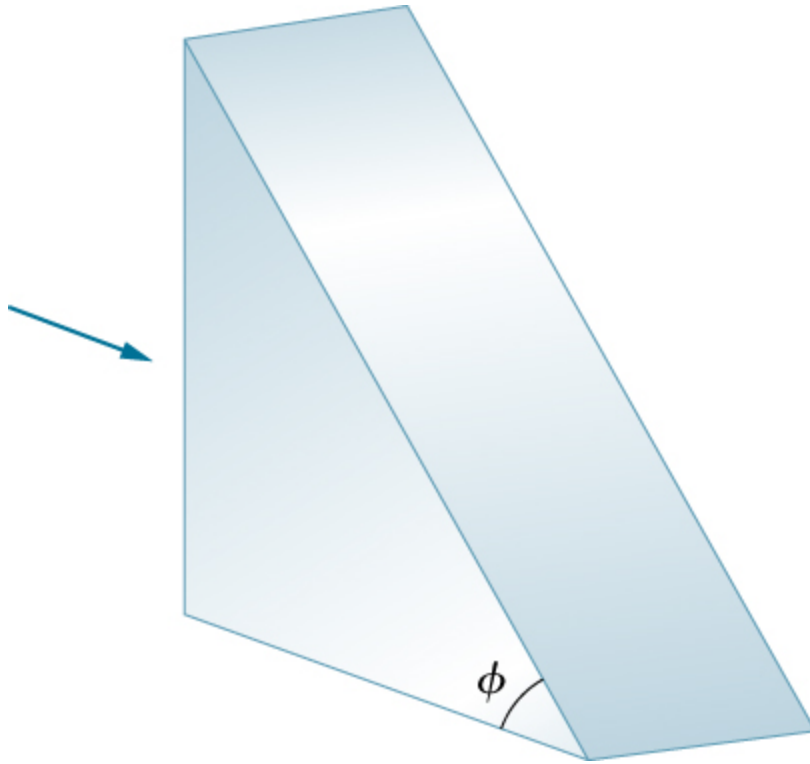
A ray of light, emitted beneath the surface of an unknown liquid with air above it, undergoes total internal reflection as shown below. What is the index of refraction for the liquid and its likely identification?



Exercise:

Problem:

Light rays fall normally on the vertical surface of the glass prism ($n = 1.50$) shown below. (a) What is the largest value for ϕ such that the ray is totally reflected at the slanted face? (b) Repeat the calculation of part (a) if the prism is immersed in water.



Solution:

a. 48.2° ; b. 27.3°

Glossary

critical angle

incident angle that produces an angle of refraction of 90°

fiber optics

field of study of the transmission of light down fibers of plastic or glass, applying the principle of total internal reflection

total internal reflection

phenomenon at the boundary between two media such that all the light is reflected and no refraction occurs

Dispersion

By the end of this section, you will be able to:

- Explain the cause of dispersion in a prism
- Describe the effects of dispersion in producing rainbows
- Summarize the advantages and disadvantages of dispersion

Everyone enjoys the spectacle of a rainbow glimmering against a dark stormy sky. How does sunlight falling on clear drops of rain get broken into the rainbow of colors we see? The same process causes white light to be broken into colors by a clear glass prism or a diamond ([link](#)).



(a)

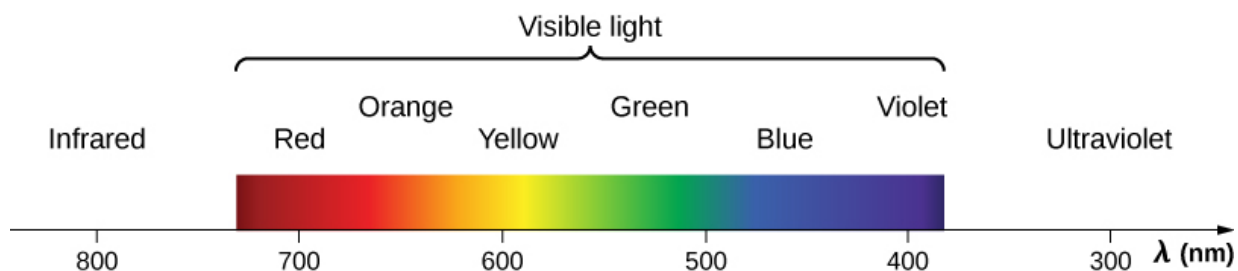


(b)

The colors of the rainbow (a) and those produced by a prism (b) are identical.
(credit a: modification of work by “Alfredo55”/Wikimedia Commons; credit b: modification of work by NASA)

We see about six colors in a rainbow—red, orange, yellow, green, blue, and violet; sometimes indigo is listed, too. These colors are associated with different wavelengths of light, as shown in [link](#). When our eye receives pure-wavelength light, we tend to see only one of the six colors, depending on wavelength. The thousands of other hues we can sense in other situations are our eye’s response to various mixtures of wavelengths. White light, in particular, is a fairly uniform mixture of all visible wavelengths. Sunlight, considered to be white, actually appears to be a bit yellow, because of its mixture of wavelengths, but it does contain all visible wavelengths. The sequence of colors in rainbows is the same sequence as the colors shown in the figure. This implies that white light is spread out in a rainbow according to wavelength.

Dispersion is defined as the spreading of white light into its full spectrum of wavelengths. More technically, dispersion occurs whenever the propagation of light depends on wavelength.



Even though rainbows are associated with six colors, the rainbow is a continuous distribution of colors according to wavelengths.

Any type of wave can exhibit dispersion. For example, sound waves, all types of electromagnetic waves, and water waves can be dispersed according to wavelength. Dispersion may require special circumstances and can result in spectacular displays such as in the production of a rainbow. This is also true for sound, since all frequencies ordinarily travel at the same speed. If you listen to sound through a long tube, such as a vacuum cleaner hose, you can easily hear it dispersed by interaction with the tube. Dispersion, in fact, can reveal a great deal about what the wave has encountered that disperses its wavelengths. The dispersion of electromagnetic radiation from outer space, for example, has revealed much about what exists between the stars—the so-called interstellar medium.

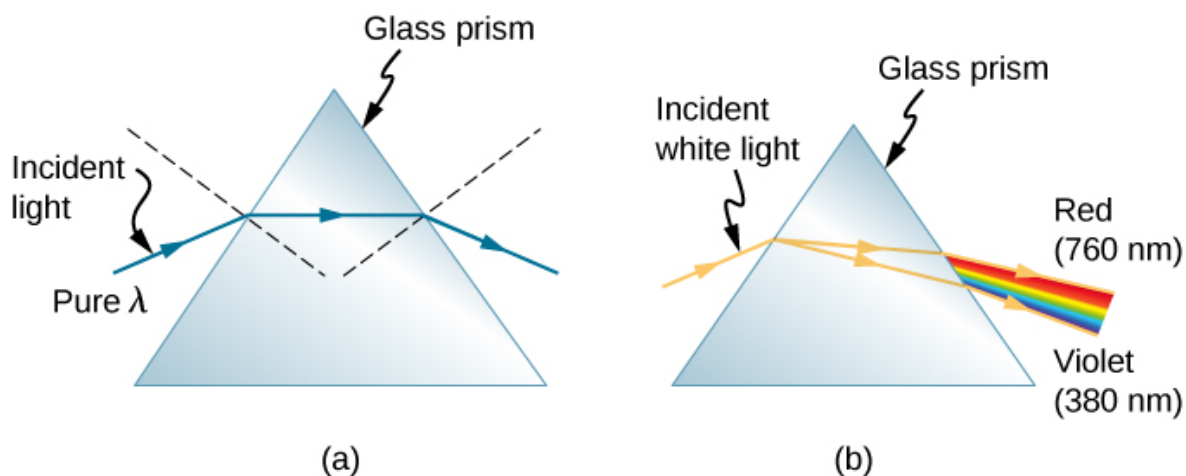
Note:

Nick Moore's [video](#) discusses dispersion of a pulse as he taps a long spring. Follow his explanation as Moore replays the high-speed footage showing high frequency waves outrunning the lower frequency waves.

Refraction is responsible for dispersion in rainbows and many other situations. The angle of refraction depends on the index of refraction, as we know from Snell's law. We know that the index of refraction n depends on the medium. But for a given medium, n also depends on wavelength ([\[link\]](#)). Note that for a given medium, n increases as wavelength decreases and is greatest for violet light. Thus, violet light is bent more than red light, as shown for a prism in [\[link\]\(b\)](#). White light is dispersed into the same sequence of wavelengths as seen in [\[link\]](#) and [\[link\]](#).

Medium	Red (660 nm)	Orange (610 nm)	Yellow (580 nm)	Green (550 nm)	Blue (470 nm)	Violet (410 nm)
Water	1.331	1.332	1.333	1.335	1.338	1.342
Diamond	2.410	2.415	2.417	2.426	2.444	2.458
Glass, crown	1.512	1.514	1.518	1.519	1.524	1.530
Glass, flint	1.662	1.665	1.667	1.674	1.684	1.698
Polystyrene	1.488	1.490	1.492	1.493	1.499	1.506
Quartz, fused	1.455	1.456	1.458	1.459	1.462	1.468

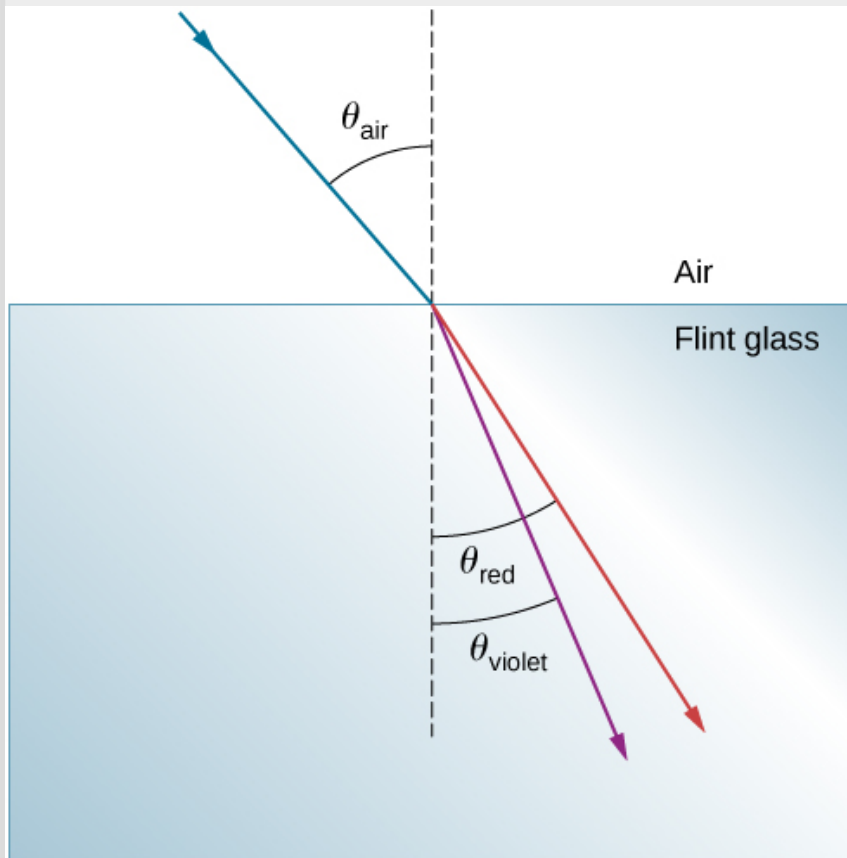
Index of Refraction n in Selected Media at Various Wavelengths



(a) A pure wavelength of light falls onto a prism and is refracted at both surfaces. (b) White light is dispersed by the prism (shown exaggerated). Since the index of refraction varies with wavelength, the angles of refraction vary with wavelength. A sequence of red to violet is produced, because the index of refraction increases steadily with decreasing wavelength.

Example:**Dispersion of White Light by Flint Glass**

A beam of white light goes from air into flint glass at an incidence angle of 43.2° . What is the angle between the red (660 nm) and violet (410 nm) parts of the refracted light?

**Strategy**

Values for the indices of refraction for flint glass at various wavelengths are listed in [\[link\]](#). Use these values to calculate the angle of refraction for each color and then take the difference to find the dispersion angle.

Solution

Applying the law of refraction for the red part of the beam

Equation:

$$n_{\text{air}} \sin \theta_{\text{air}} = n_{\text{red}} \sin \theta_{\text{red}},$$

we can solve for the angle of refraction as

Equation:

$$\theta_{\text{red}} = \sin^{-1} \left(\frac{n_{\text{air}} \sin \theta_{\text{air}}}{n_{\text{red}}} \right) = \sin^{-1} \left[\frac{(1.000) \sin 43.2^\circ}{(1.512)} \right] = 27.0^\circ.$$

Similarly, the angle of incidence for the violet part of the beam is

Equation:

$$\theta_{\text{violet}} = \sin^{-1} \left(\frac{n_{\text{air}} \sin \theta_{\text{air}}}{n_{\text{violet}}} \right) = \sin^{-1} \left[\frac{(1.000) \sin 43.2^\circ}{(1.530)} \right] = 26.4^\circ.$$

The difference between these two angles is

Equation:

$$\theta_{\text{red}} - \theta_{\text{violet}} = 27.0^\circ - 26.4^\circ = 0.6^\circ.$$

Significance

Although 0.6° may seem like a negligibly small angle, if this beam is allowed to propagate a long enough distance, the dispersion of colors becomes quite noticeable.

Note:

Exercise:

Problem:

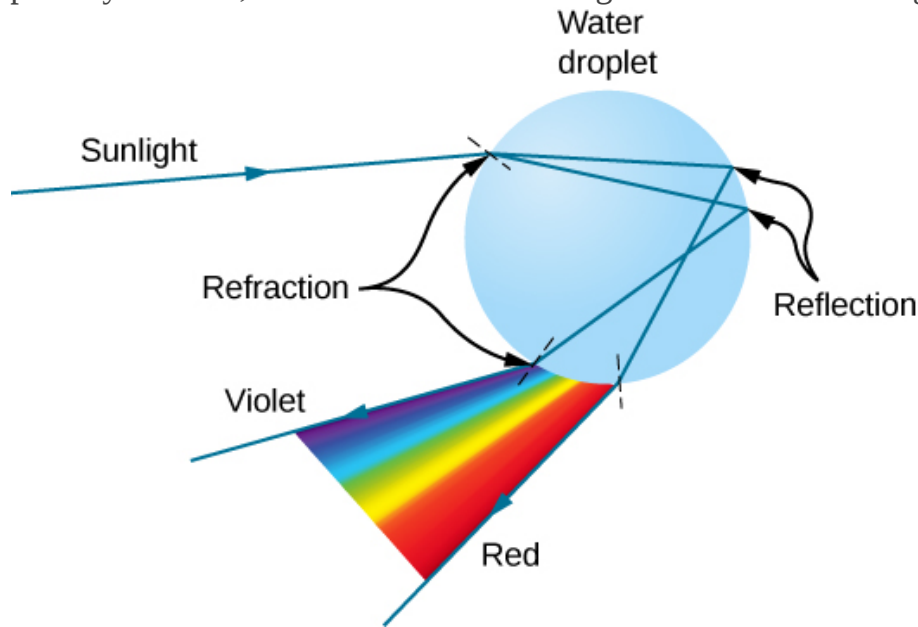
Check Your Understanding In the preceding example, how much distance inside the block of flint glass would the red and the violet rays have to progress before they are separated by 1.0 mm?

Solution:

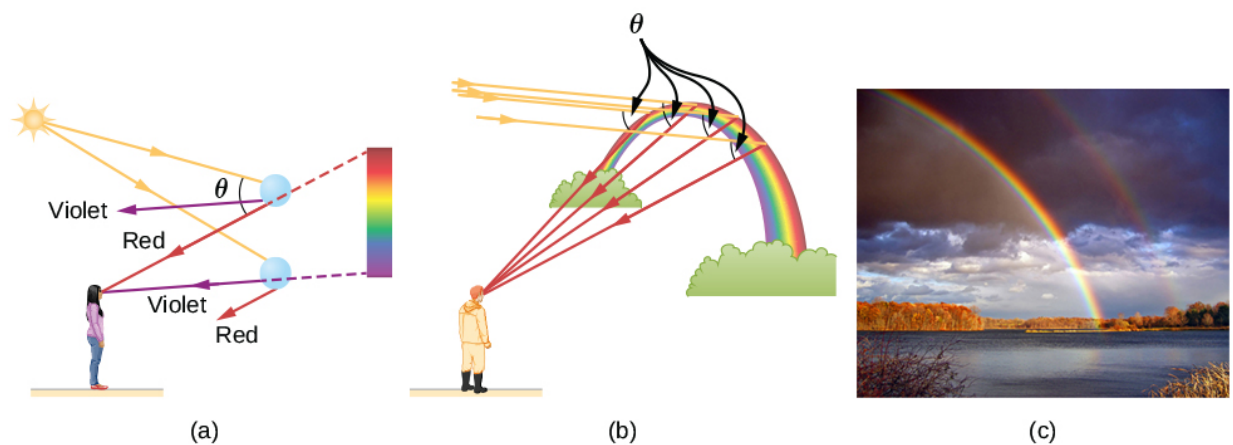
9.3 cm

Rainbows are produced by a combination of refraction and reflection. You may have noticed that you see a rainbow only when you look away from the Sun. Light enters a drop of water and is reflected from the back of the drop ([link](#)). The light is refracted both as it enters and as it leaves the drop. Since the index of refraction of water varies with wavelength, the light is dispersed, and a rainbow is observed ([link](#)(a)). (No dispersion occurs at the back surface, because the law of reflection does not depend on wavelength.) The actual rainbow of colors seen by an observer depends on the myriad rays being refracted and reflected toward the observer's eyes from numerous drops of water. The effect is most spectacular when the background is dark, as in stormy weather, but can also be observed in waterfalls and lawn sprinklers. The arc of a rainbow comes from the need to be looking at a specific angle relative to the direction of the Sun, as

illustrated in part (b). If two reflections of light occur within the water drop, another “secondary” rainbow is produced. This rare event produces an arc that lies above the primary rainbow arc, as in part (c), and produces colors in the reverse order of the primary rainbow, with red at the lowest angle and violet at the largest angle.



A ray of light falling on this water drop enters and is reflected from the back of the drop. This light is refracted and dispersed both as it enters and as it leaves the drop.



(a) Different colors emerge in different directions, and so you must look at different locations to see the various colors of a rainbow. (b) The arc of a rainbow results from the fact that a line between the observer and any point on the arc must make the correct angle with the parallel rays of sunlight for the observer to receive

the refracted rays. (c) Double rainbow. (credit c: modification of work by “Nicholas”/Wikimedia Commons)

Dispersion may produce beautiful rainbows, but it can cause problems in optical systems. White light used to transmit messages in a fiber is dispersed, spreading out in time and eventually overlapping with other messages. Since a laser produces a nearly pure wavelength, its light experiences little dispersion, an advantage over white light for transmission of information. In contrast, dispersion of electromagnetic waves coming to us from outer space can be used to determine the amount of matter they pass through.

Summary

- The spreading of white light into its full spectrum of wavelengths is called dispersion.
- Rainbows are produced by a combination of refraction and reflection, and involve the dispersion of sunlight into a continuous distribution of colors.
- Dispersion produces beautiful rainbows but also causes problems in certain optical systems.

Conceptual Questions

Exercise:

Problem:

Is it possible that total internal reflection plays a role in rainbows? Explain in terms of indices of refraction and angles, perhaps referring to that shown below. Some of us have seen the formation of a double rainbow; is it physically possible to observe a triple rainbow?



(credit: "Chad"/Flickr)

Exercise:

Problem:

A high-quality diamond may be quite clear and colorless, transmitting all visible wavelengths with little absorption. Explain how it can sparkle with flashes of brilliant color when illuminated by white light.

Solution:

In addition to total internal reflection, rays that refract into and out of diamond crystals are subject to dispersion due to varying values of n across the spectrum, resulting in a sparkling display of colors.

Problems

Exercise:

Problem:

(a) What is the ratio of the speed of red light to violet light in diamond, based on [\[link\]](#)? (b) What is this ratio in polystyrene? (c) Which is more dispersive?

Exercise:

Problem:

A beam of white light goes from air into water at an incident angle of 75.0° . At what angles are the red (660 nm) and violet (410 nm) parts of the light refracted?

Solution:

46.5° for red, 46.0° for violet

Exercise:**Problem:**

By how much do the critical angles for red (660 nm) and violet (410 nm) light differ in a diamond surrounded by air?

Exercise:**Problem:**

(a) A narrow beam of light containing yellow (580 nm) and green (550 nm) wavelengths goes from polystyrene to air, striking the surface at a 30.0° incident angle. What is the angle between the colors when they emerge? (b) How far would they have to travel to be separated by 1.00 mm?

Solution:

a. 0.04° ; b. 1.3 m

Exercise:**Problem:**

A parallel beam of light containing orange (610 nm) and violet (410 nm) wavelengths goes from fused quartz to water, striking the surface between them at a 60.0° incident angle. What is the angle between the two colors in water?

Exercise:**Problem:**

A ray of 610-nm light goes from air into fused quartz at an incident angle of 55.0° . At what incident angle must 470 nm light enter flint glass to have the same angle of refraction?

Solution:

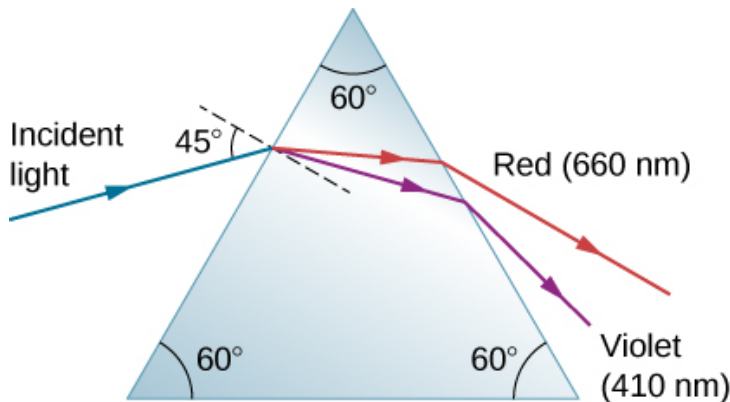
72.8°

Exercise:**Problem:**

A narrow beam of light containing red (660 nm) and blue (470 nm) wavelengths travels from air through a 1.00-cm-thick flat piece of crown glass and back to air again. The beam strikes at a 30.0° incident angle. (a) At what angles do the two colors emerge? (b) By what distance are the red and blue separated when they emerge?

Exercise:**Problem:**

A narrow beam of white light enters a prism made of crown glass at a 45.0° incident angle, as shown below. At what angles, θ_R and θ_V , do the red (660 nm) and violet (410 nm) components of the light emerge from the prism?

**Solution:**

53.5° for red, 55.2° for violet

Glossary

dispersion

spreading of light into its spectrum of wavelengths

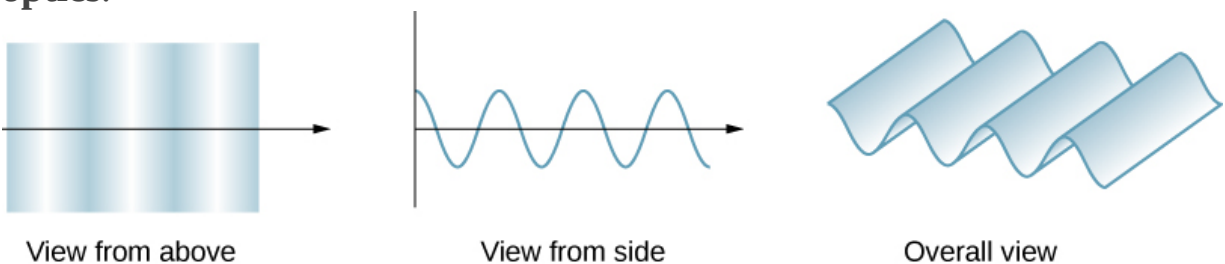
Huygens's Principle

By the end of this section, you will be able to:

- Describe Huygens's principle
- Use Huygens's principle to explain the law of reflection
- Use Huygens's principle to explain the law of refraction
- Use Huygens's principle to explain diffraction

So far in this chapter, we have been discussing optical phenomena using the ray model of light. However, some phenomena require analysis and explanations based on the wave characteristics of light. This is particularly true when the wavelength is not negligible compared to the dimensions of an optical device, such as a slit in the case of *diffraction*. Huygens's principle is an indispensable tool for this analysis.

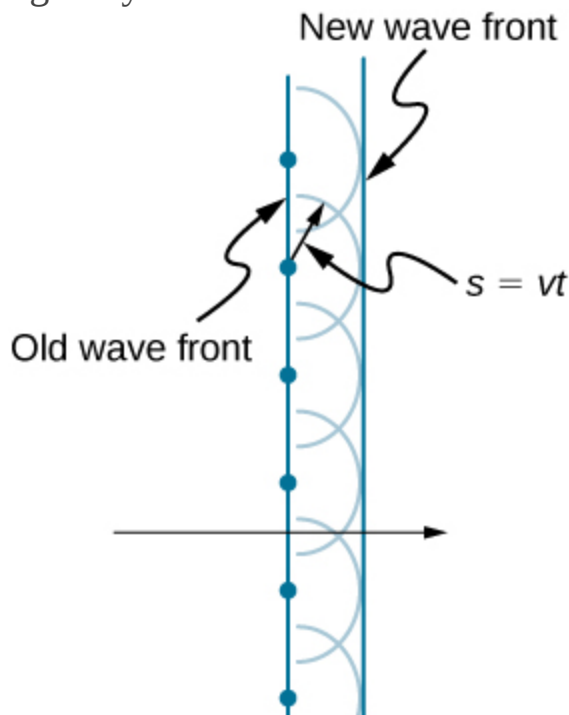
[\[link\]](#) shows how a transverse wave looks as viewed from above and from the side. A light wave can be imagined to propagate like this, although we do not actually see it wiggling through space. From above, we view the wave fronts (or wave crests) as if we were looking down on ocean waves. The side view would be a graph of the electric or magnetic field. The view from above is perhaps more useful in developing concepts about **wave optics**.



A transverse wave, such as an electromagnetic light wave, as viewed from above and from the side. The direction of propagation is perpendicular to the wave fronts (or wave crests) and is represented by a ray.

The Dutch scientist Christiaan Huygens (1629–1695) developed a useful technique for determining in detail how and where waves propagate. Starting from some known position, **Huygens's principle** states that every point on a wave front is a source of wavelets that spread out in the forward direction at the same speed as the wave itself. The new wave front is tangent to all of the wavelets.

[\[link\]](#) shows how Huygens's principle is applied. A wave front is the long edge that moves, for example, with the crest or the trough. Each point on the wave front emits a semicircular wave that moves at the propagation speed v . We can draw these wavelets at a time t later, so that they have moved a distance $s = vt$. The new wave front is a plane tangent to the wavelets and is where we would expect the wave to be a time t later. Huygens's principle works for all types of waves, including water waves, sound waves, and light waves. It is useful not only in describing how light waves propagate but also in explaining the laws of reflection and refraction. In addition, we will see that Huygens's principle tells us how and where light rays interfere.

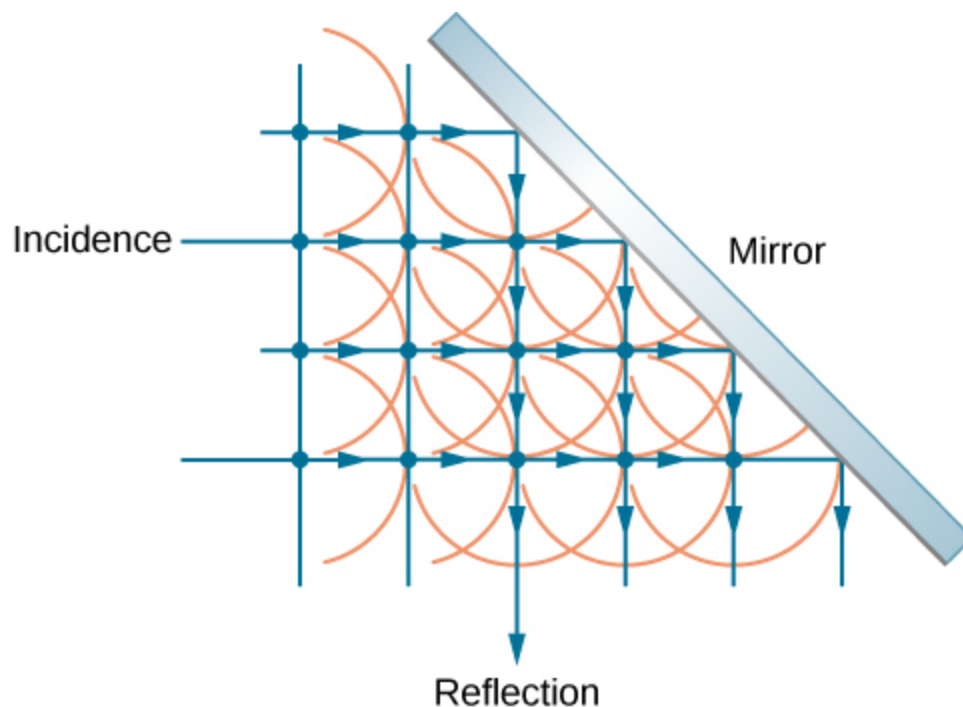


Huygens's principle applied to a straight wave front. Each point on the wave front emits

a semicircular wavelet that moves a distance $s = vt$. The new wave front is a line tangent to the wavelets.

Reflection

[\[link\]](#) shows how a mirror reflects an incoming wave at an angle equal to the incident angle, verifying the law of reflection. As the wave front strikes the mirror, wavelets are first emitted from the left part of the mirror and then from the right. The wavelets closer to the left have had time to travel farther, producing a wave front traveling in the direction shown.

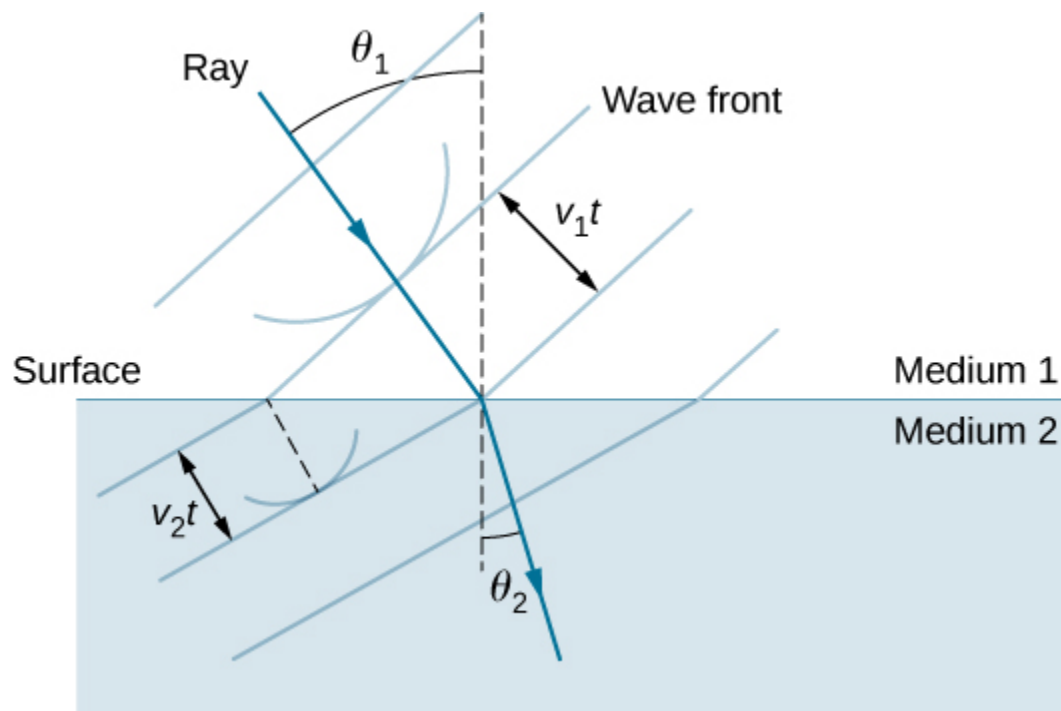


Huygens's principle applied to a plane wave front striking a mirror. The wavelets shown were emitted as each point on the wave front struck the mirror. The tangent to these wavelets shows that the new wave

front has been reflected at an angle equal to the incident angle. The direction of propagation is perpendicular to the wave front, as shown by the downward-pointing arrows.

Refraction

The law of refraction can be explained by applying Huygens's principle to a wave front passing from one medium to another ([link](#)). Each wavelet in the figure was emitted when the wave front crossed the interface between the media. Since the speed of light is smaller in the second medium, the waves do not travel as far in a given time, and the new wave front changes direction as shown. This explains why a ray changes direction to become closer to the perpendicular when light slows down. Snell's law can be derived from the geometry in [link](#) ([link](#)).



Huygens's principle applied to a plane wave front traveling

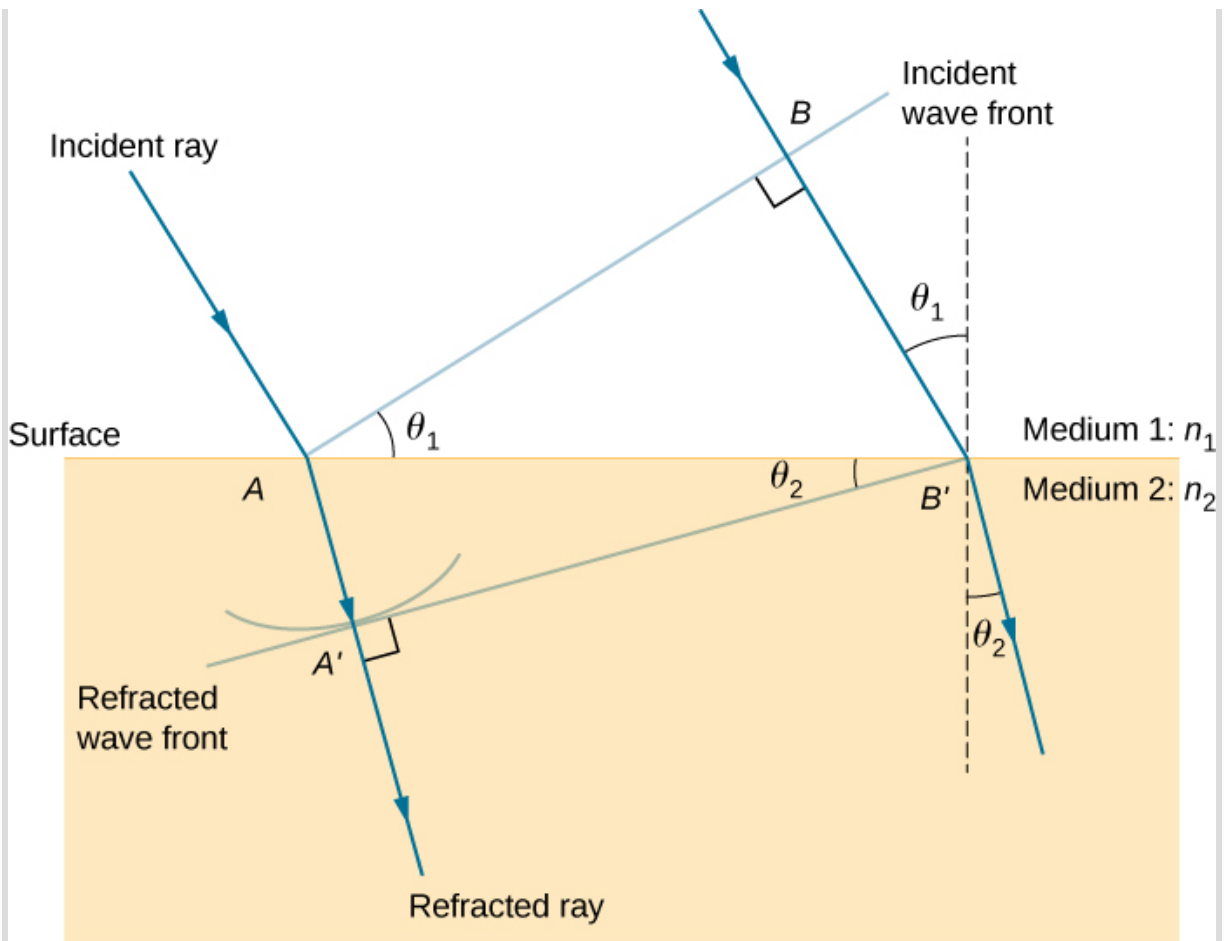
from one medium to another, where its speed is less. The ray bends toward the perpendicular, since the wavelets have a lower speed in the second medium.

Example:**Deriving the Law of Refraction**

By examining the geometry of the wave fronts, derive the law of refraction.

Strategy

Consider [\[link\]](#), which expands upon [\[link\]](#). It shows the incident wave front just reaching the surface at point A , while point B is still well within medium 1. In the time Δt it takes for a wavelet from B to reach B' on the surface at speed $v_1 = c/n_1$, a wavelet from A travels into medium 2 a distance of $AA' = v_2\Delta t$, where $v_2 = c/n_2$. Note that in this example, v_2 is slower than v_1 because $n_1 < n_2$.



Geometry of the law of refraction from medium 1 to medium 2.

Solution

The segment on the surface AB' is shared by both the triangle ABB' inside medium 1 and the triangle $AA'B'$ inside medium 2. Note that from the geometry, the angle $\angle BAB'$ is equal to the angle of incidence, θ_1 . Similarly, $\angle AB'A'$ is θ_2 .

The length of AB' is given in two ways as

Equation:

$$AB' = \frac{BB'}{\sin \theta_1} = \frac{AA'}{\sin \theta_2}.$$

Inverting the equation and substituting $AA' = c\Delta t/n_2$ from above and similarly $BB' = c\Delta t/n_1$, we obtain

Equation:

$$\frac{\sin \theta_1}{c\Delta t/n_1} = \frac{\sin \theta_2}{c\Delta t/n_2}.$$

Cancellation of $c\Delta t$ allows us to simplify this equation into the familiar form

Equation:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

Significance

Although the law of refraction was established experimentally by Snell and stated in [Refraction](#), its derivation here requires Huygens's principle and the understanding that the speed of light is different in different media.

Note:**Exercise:****Problem:**

Check Your Understanding In [\[link\]](#), we had $n_1 < n_2$. If n_2 were decreased such that $n_1 > n_2$ and the speed of light in medium 2 is faster than in medium 1, what would happen to the length of AA' ? What would happen to the wave front $A'B'$ and the direction of the refracted ray?

Solution:

AA' becomes longer, $A'B'$ tilts further away from the surface, and the refracted ray tilts away from the normal.

Note:

This [applet](#) by Walter Fendt shows an animation of reflection and refraction using Huygens's wavelets while you control the parameters. Be sure to click on "Next step" to display the wavelets. You can see the reflected and refracted wave fronts forming.

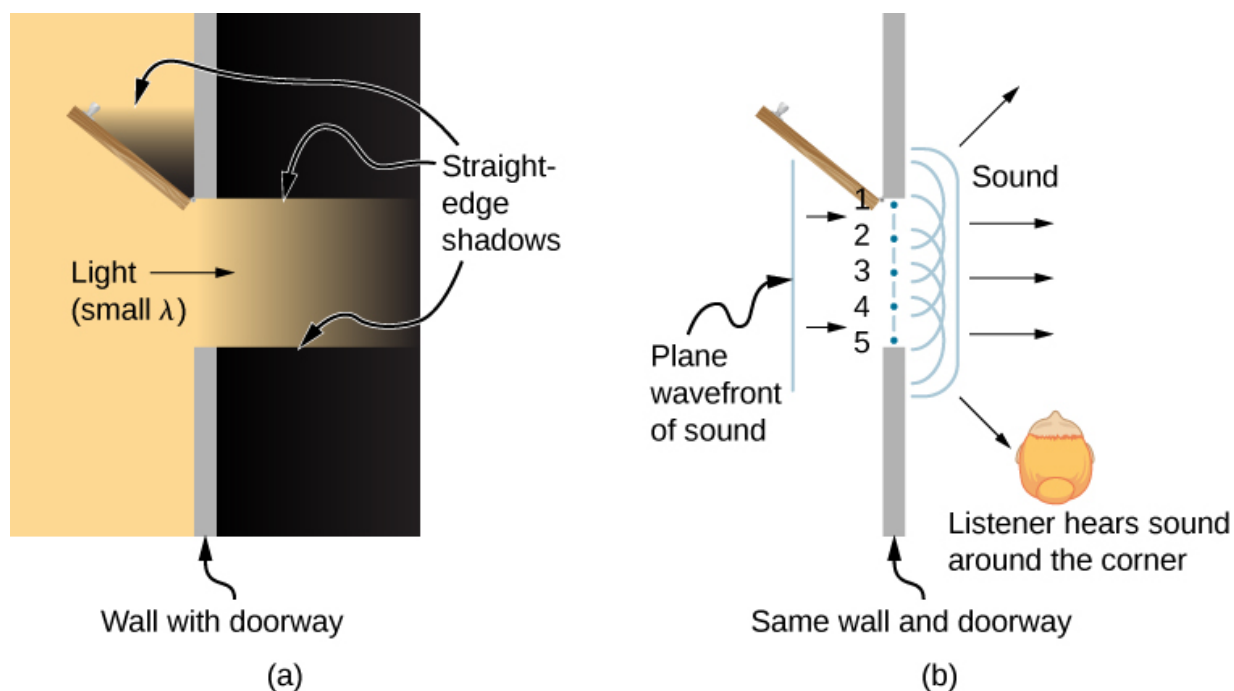
Diffraction

What happens when a wave passes through an opening, such as light shining through an open door into a dark room? For light, we observe a sharp shadow of the doorway on the floor of the room, and no visible light bends around corners into other parts of the room. When sound passes through a door, we hear it everywhere in the room and thus observe that sound spreads out when passing through such an opening ([link](#)). What is the difference between the behavior of sound waves and light waves in this case? The answer is that light has very short wavelengths and acts like a ray. Sound has wavelengths on the order of the size of the door and bends around corners (for frequency of 1000 Hz,

Equation:

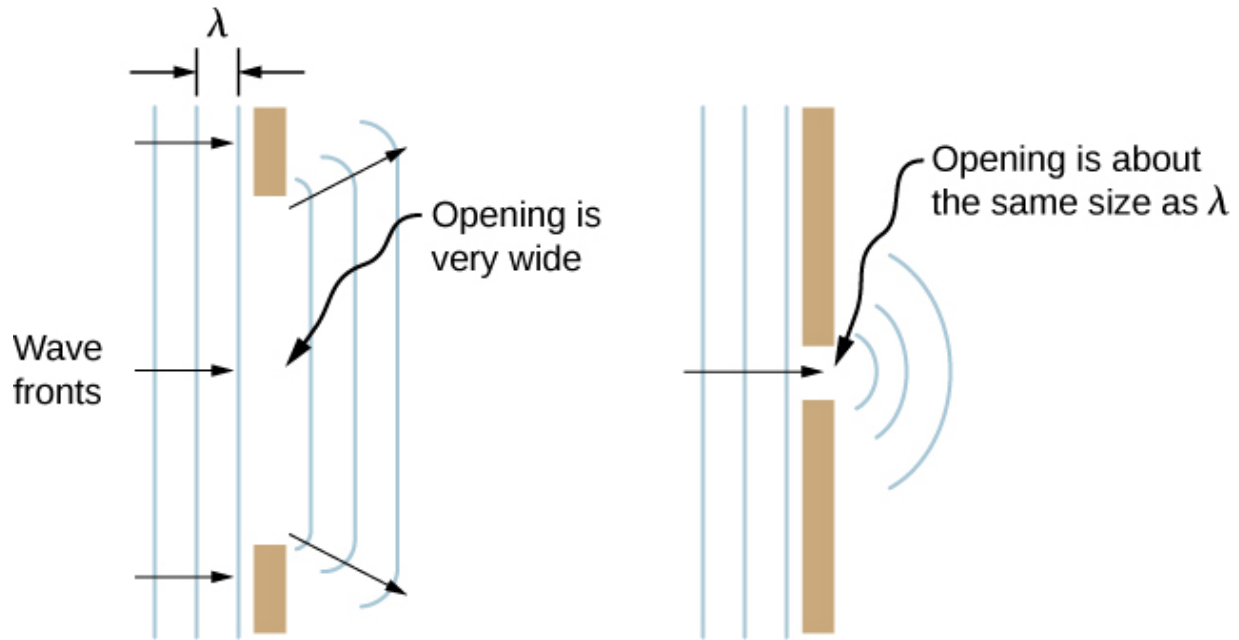
$$\lambda = \frac{c}{f} = \frac{330 \text{ m/s}}{1000 \text{ s}^{-1}} = 0.33 \text{ m},$$

about three times smaller than the width of the doorway).



(a) Light passing through a doorway makes a sharp outline on the floor. Since light's wavelength is very small compared with the size of the door, it acts like a ray. (b) Sound waves bend into all parts of the room, a wave effect, because their wavelength is similar to the size of the door.

If we pass light through smaller openings such as slits, we can use Huygens's principle to see that light bends as sound does ([\[link\]](#)). The bending of a wave around the edges of an opening or an obstacle is called diffraction. Diffraction is a wave characteristic and occurs for all types of waves. If diffraction is observed for some phenomenon, it is evidence that the phenomenon is a wave. Thus, the horizontal diffraction of the laser beam after it passes through the slits in [\[link\]](#) is evidence that light is a wave. You will learn about diffraction in much more detail in the chapter on [Diffraction](#).



Huygens's principle applied to a plane wave front striking an opening. The edges of the wave front bend after passing through the opening, a process called diffraction. The amount of bending is more extreme for a small opening, consistent with the fact that wave characteristics are most noticeable for interactions with objects about the same size as the wavelength.

Summary

- According to Huygens's principle, every point on a wave front is a source of wavelets that spread out in the forward direction at the same speed as the wave itself. The new wave front is tangent to all of the wavelets.
- A mirror reflects an incoming wave at an angle equal to the incident angle, verifying the law of reflection.
- The law of refraction can be explained by applying Huygens's principle to a wave front passing from one medium to another.
- The bending of a wave around the edges of an opening or an obstacle is called diffraction.

Conceptual Questions

Exercise:

Problem:

How do wave effects depend on the size of the object with which the wave interacts? For example, why does sound bend around the corner of a building while light does not?

Exercise:

Problem: Does Huygens's principle apply to all types of waves?

Solution:

yes

Exercise:

Problem:

If diffraction is observed for some phenomenon, it is evidence that the phenomenon is a wave. Does the reverse hold true? That is, if diffraction is not observed, does that mean the phenomenon is not a wave?

Glossary

Huygens's principle

every point on a wave front is a source of wavelets that spread out in the forward direction at the same speed as the wave itself; the new wave front is a plane tangent to all of the wavelets

wave optics

part of optics dealing with the wave aspect of light

Polarization

By the end of this section, you will be able to:

- Explain the change in intensity as polarized light passes through a polarizing filter
- Calculate the effect of polarization by reflection and Brewster's angle
- Describe the effect of polarization by scattering
- Explain the use of polarizing materials in devices such as LCDs

Polarizing sunglasses are familiar to most of us. They have a special ability to cut the glare of light reflected from water or glass ([link](#)). They have this ability because of a wave characteristic of light called polarization. What is polarization? How is it produced? What are some of its uses? The answers to these questions are related to the wave character of light.



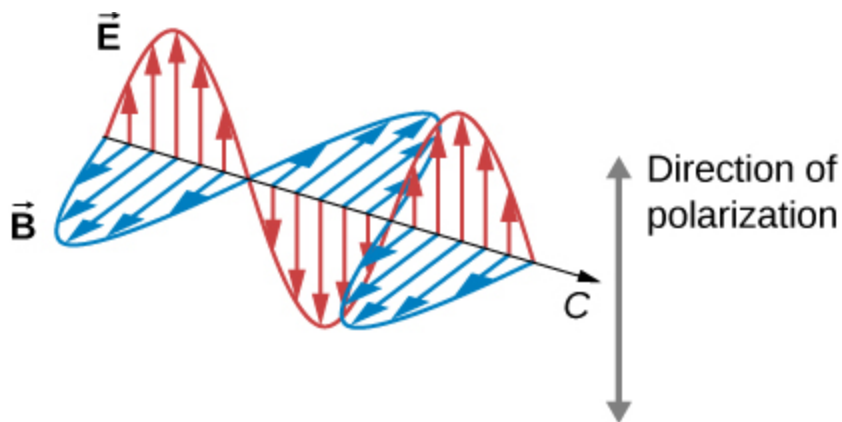
(a)

(b)

These two photographs of a river show the effect of a polarizing filter in reducing glare in light reflected from the surface of water. Part (b) of this figure was taken with a polarizing filter and part (a) was not. As a result, the reflection of clouds and sky observed in part (a) is not observed in part (b). Polarizing sunglasses are particularly useful on snow and water. (credit a and credit b: modifications of work by “Amithshs”/Wikimedia Commons)

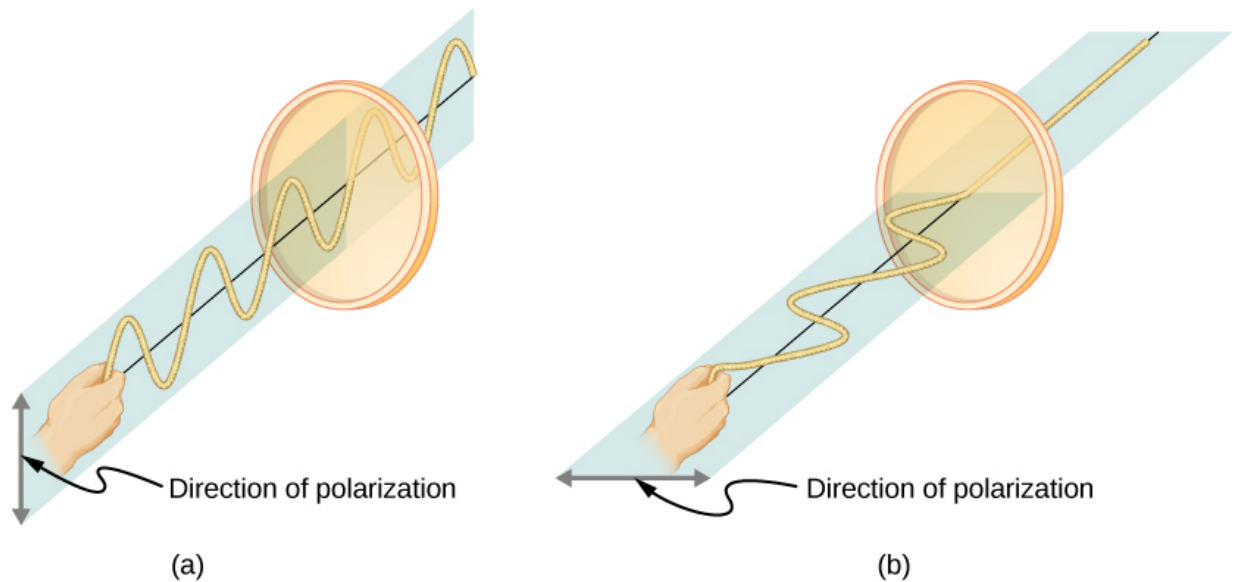
Malus's Law

Light is one type of electromagnetic (EM) wave. As noted in the previous chapter on [Electromagnetic Waves](#), EM waves are *transverse waves* consisting of varying electric and magnetic fields that oscillate perpendicular to the direction of propagation ([\[link\]](#)). However, in general, there are no specific directions for the oscillations of the electric and magnetic fields; they vibrate in any randomly oriented plane perpendicular to the direction of propagation. **Polarization** is the attribute that a wave's oscillations do have a definite direction relative to the direction of propagation of the wave. (This is not the same type of polarization as that discussed for the separation of charges.) Waves having such a direction are said to be **polarized**. For an EM wave, we define the **direction of polarization** to be the direction parallel to the electric field. Thus, we can think of the electric field arrows as showing the direction of polarization, as in [\[link\]](#).



An EM wave, such as light, is a transverse wave. The electric (\vec{E}) and magnetic (\vec{B}) fields are perpendicular to the direction of propagation. The direction of polarization of the wave is the direction of the electric field.

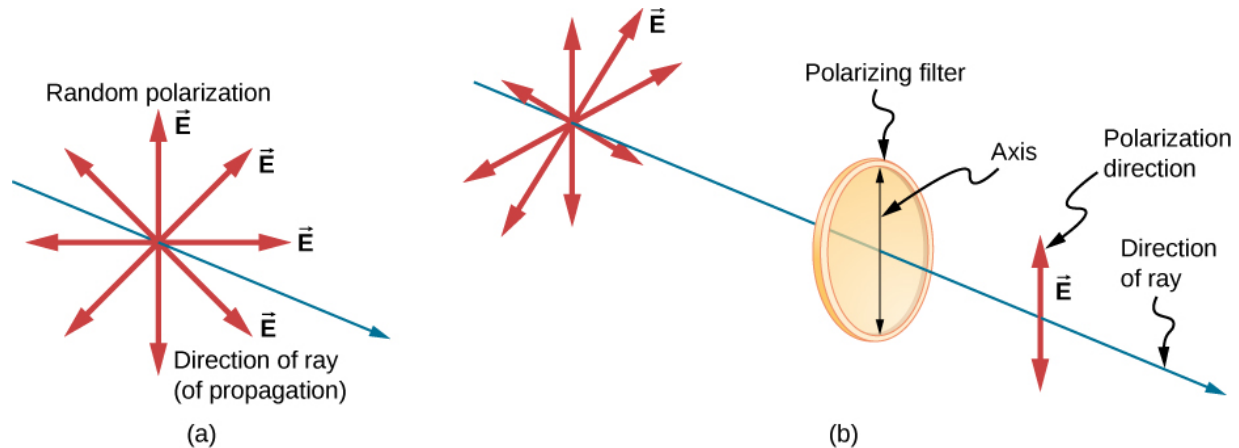
To examine this further, consider the transverse waves in the ropes shown in [\[link\]](#). The oscillations in one rope are in a vertical plane and are said to be **vertically polarized**. Those in the other rope are in a horizontal plane and are **horizontally polarized**. If a vertical slit is placed on the first rope, the waves pass through. However, a vertical slit blocks the horizontally polarized waves. For EM waves, the direction of the electric field is analogous to the disturbances on the ropes.



The transverse oscillations in one rope (a) are in a vertical plane, and those in the other rope (b) are in a horizontal plane. The first is said to be vertically polarized, and the other is said to be horizontally polarized. Vertical slits pass vertically polarized waves and block horizontally polarized waves.

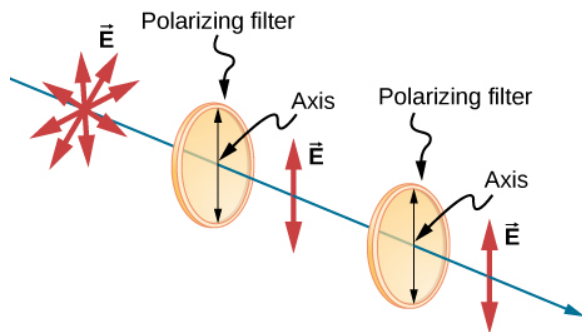
The Sun and many other light sources produce waves that have the electric fields in random directions ([\[link\]](#)(a)). Such light is said to be **unpolarized**, because it is composed of many waves with all possible directions of polarization. Polaroid materials—which were invented by the founder of the Polaroid Corporation, Edwin Land—act as a polarizing slit for light, allowing only polarization in one direction to pass through. Polarizing

filters are composed of long molecules aligned in one direction. If we think of the molecules as many slits, analogous to those for the oscillating ropes, we can understand why only light with a specific polarization can get through. The axis of a polarizing filter is the direction along which the filter passes the electric field of an EM wave.

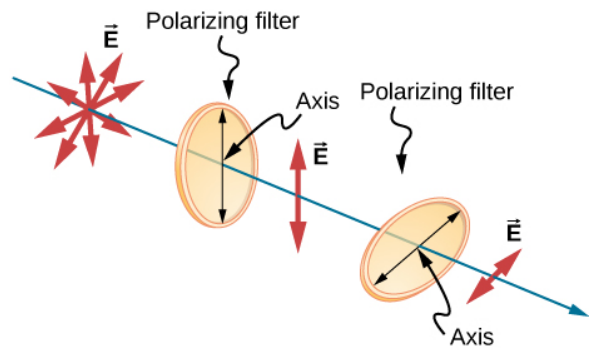


The slender arrow represents a ray of unpolarized light. The bold arrows represent the direction of polarization of the individual waves composing the ray. (a) If the light is unpolarized, the arrows point in all directions. (b) A polarizing filter has a polarization axis that acts as a slit passing through electric fields parallel to its direction. The direction of polarization of an EM wave is defined to be the direction of its electric field.

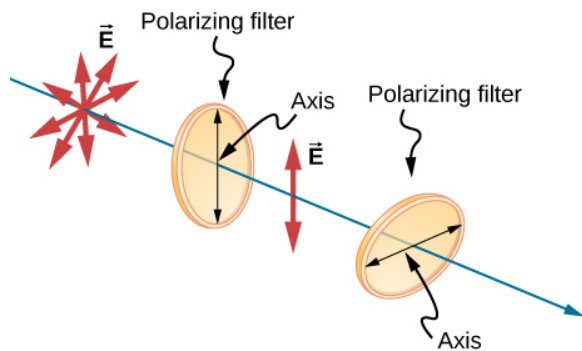
[\[link\]](#) shows the effect of two polarizing filters on originally unpolarized light. The first filter polarizes the light along its axis. When the axes of the first and second filters are aligned (parallel), then all of the polarized light passed by the first filter is also passed by the second filter. If the second polarizing filter is rotated, only the component of the light parallel to the second filter's axis is passed. When the axes are perpendicular, no light is passed by the second filter.



(a)



(b)



(c)



(d)

The effect of rotating two polarizing filters, where the first polarizes the light. (a) All of the polarized light is passed by the second polarizing filter, because its axis is parallel to the first. (b) As the second filter is rotated, only part of the light is passed. (c) When the second filter is perpendicular to the first, no light is passed. (d) In this photograph, a polarizing filter is placed above two others. Its axis is perpendicular to the filter on the right (dark area) and parallel to the filter on the left (lighter area). (credit d: modification of work by P.P. Urone)

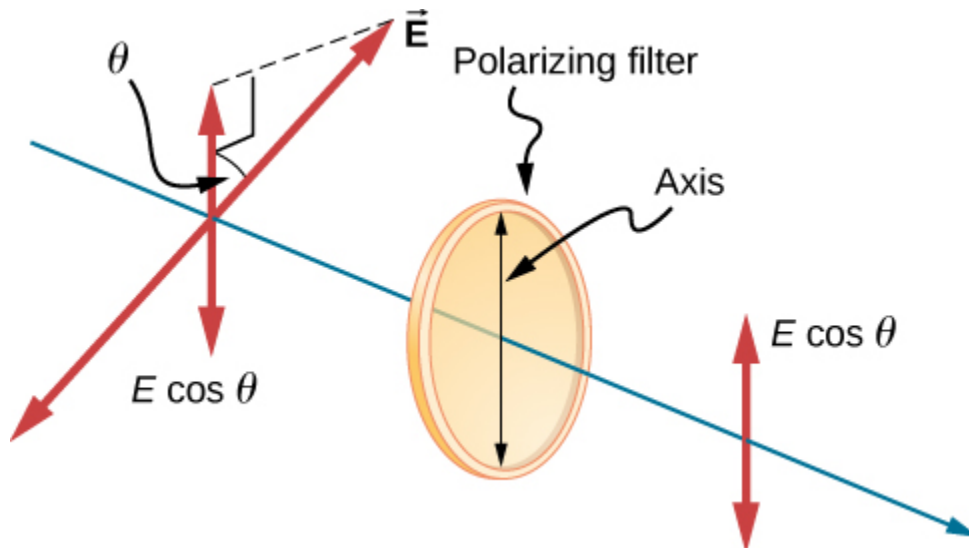
Only the component of the EM wave parallel to the axis of a filter is passed. Let us call the angle between the direction of polarization and the axis of a filter θ . If the electric field has an amplitude E , then the transmitted part of the wave has an amplitude $E \cos \theta$ ([\[link\]](#)). Since the intensity of a wave is proportional to its amplitude squared, the intensity I of the transmitted wave is related to the incident wave by

Note:

Equation:

$$I = I_0 \cos^2 \theta$$

where I_0 is the intensity of the polarized wave before passing through the filter. This equation is known as **Malus's law**.



A polarizing filter transmits only the component of the wave parallel to its axis, reducing the intensity of any light not polarized parallel to its axis.

Note:

This [Open Source Physics animation](#) helps you visualize the electric field vectors as light encounters a polarizing filter. You can rotate the filter—

note that the angle displayed is in radians. You can also rotate the animation for 3D visualization.

Example:**Calculating Intensity Reduction by a Polarizing Filter**

What angle is needed between the direction of polarized light and the axis of a polarizing filter to reduce its intensity by 90.0%?

Strategy

When the intensity is reduced by 90.0%, it is 10.0% or 0.100 times its original value. That is, $I = 0.100 I_0$. Using this information, the equation $I = I_0 \cos^2 \theta$ can be used to solve for the needed angle.

Solution

Solving the equation $I = I_0 \cos^2 \theta$ for $\cos \theta$ and substituting with the relationship between I and I_0 gives

Equation:

$$\cos \theta = \sqrt{\frac{I}{I_0}} = \sqrt{\frac{0.100 I_0}{I_0}} = 0.3162.$$

Solving for θ yields

Equation:

$$\theta = \cos^{-1} 0.3162 = 71.6^\circ.$$

Significance

A fairly large angle between the direction of polarization and the filter axis is needed to reduce the intensity to 10.0% of its original value. This seems reasonable based on experimenting with polarizing films. It is interesting that at an angle of 45° , the intensity is reduced to 50% of its original value. Note that 71.6° is 18.4° from reducing the intensity to zero, and that at an angle of 18.4° , the intensity is reduced to 90.0% of its original value, giving evidence of symmetry.

Note:

Exercise:

Problem:

Check Your Understanding Although we did not specify the direction in [\[link\]](#), let's say the polarizing filter was rotated clockwise by 71.6° to reduce the light intensity by 90.0%. What would be the intensity reduction if the polarizing filter were rotated counterclockwise by 71.6° ?

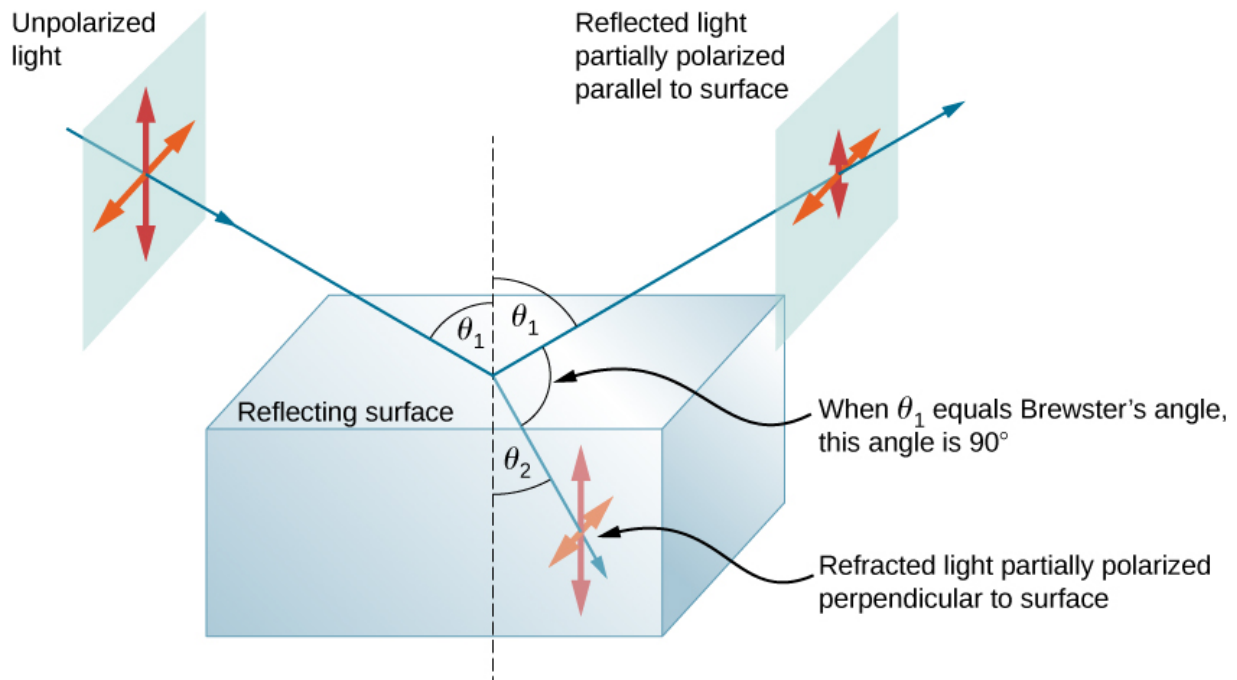
Solution:

also 90.0%

Polarization by Reflection

By now, you can probably guess that polarizing sunglasses cut the glare in reflected light, because that light is polarized. You can check this for yourself by holding polarizing sunglasses in front of you and rotating them while looking at light reflected from water or glass. As you rotate the sunglasses, you will notice the light gets bright and dim, but not completely black. This implies the reflected light is partially polarized and cannot be completely blocked by a polarizing filter.

[\[link\]](#) illustrates what happens when unpolarized light is reflected from a surface. Vertically polarized light is preferentially refracted at the surface, so the reflected light is left more horizontally polarized. The reasons for this phenomenon are beyond the scope of this text, but a convenient mnemonic for remembering this is to imagine the polarization direction to be like an arrow. Vertical polarization is like an arrow perpendicular to the surface and is more likely to stick and not be reflected. Horizontal polarization is like an arrow bouncing on its side and is more likely to be reflected. Sunglasses with vertical axes thus block more reflected light than unpolarized light from other sources.



Polarization by reflection. Unpolarized light has equal amounts of vertical and horizontal polarization. After interaction with a surface, the vertical components are preferentially absorbed or refracted, leaving the reflected light more horizontally polarized. This is akin to arrows striking on their sides and bouncing off, whereas arrows striking on their tips go into the surface.

Since the part of the light that is not reflected is refracted, the amount of polarization depends on the indices of refraction of the media involved. It can be shown that reflected light is completely polarized at an angle of reflection θ_b given by

Note:
Equation:

$$\tan \theta_b = \frac{n_2}{n_1}$$

where n_1 is the medium in which the incident and reflected light travel and n_2 is the index of refraction of the medium that forms the interface that reflects the light. This equation is known as **Brewster's law** and θ_b is known as **Brewster's angle**, named after the nineteenth-century Scottish physicist who discovered them.

Note:

This [Open Source Physics animation](#) shows incident, reflected, and refracted light as rays and EM waves. Try rotating the animation for 3D visualization and also change the angle of incidence. Near Brewster's angle, the reflected light becomes highly polarized.

Example:

Calculating Polarization by Reflection

(a) At what angle will light traveling in air be completely polarized horizontally when reflected from water? (b) From glass?

Strategy

All we need to solve these problems are the indices of refraction. Air has $n_1 = 1.00$, water has $n_2 = 1.333$, and crown glass has $n'_2 = 1.520$. The equation $\tan \theta_b = \frac{n_2}{n_1}$ can be directly applied to find θ_b in each case.

Solution

- a. Putting the known quantities into the equation

Equation:

$$\tan \theta_b = \frac{n_2}{n_1}$$

gives

Equation:

$$\tan \theta_b = \frac{n_2}{n_1} = \frac{1.333}{1.00} = 1.333.$$

Solving for the angle θ_b yields

Equation:

$$\theta_b = \tan^{-1} 1.333 = 53.1^\circ.$$

b. Similarly, for crown glass and air,

Equation:

$$\tan \theta'_b = \frac{n'_2}{n_1} = \frac{1.520}{1.00} = 1.52.$$

Thus,

Equation:

$$\theta'_b = \tan^{-1} 1.52 = 56.7^\circ.$$

Significance

Light reflected at these angles could be completely blocked by a good polarizing filter held with its axis vertical. Brewster's angle for water and air are similar to those for glass and air, so that sunglasses are equally effective for light reflected from either water or glass under similar circumstances. Light that is not reflected is refracted into these media. Therefore, at an incident angle equal to Brewster's angle, the refracted light is slightly polarized vertically. It is not completely polarized vertically, because only a small fraction of the incident light is reflected, so a significant amount of horizontally polarized light is refracted.

Note:

Exercise:

Problem:

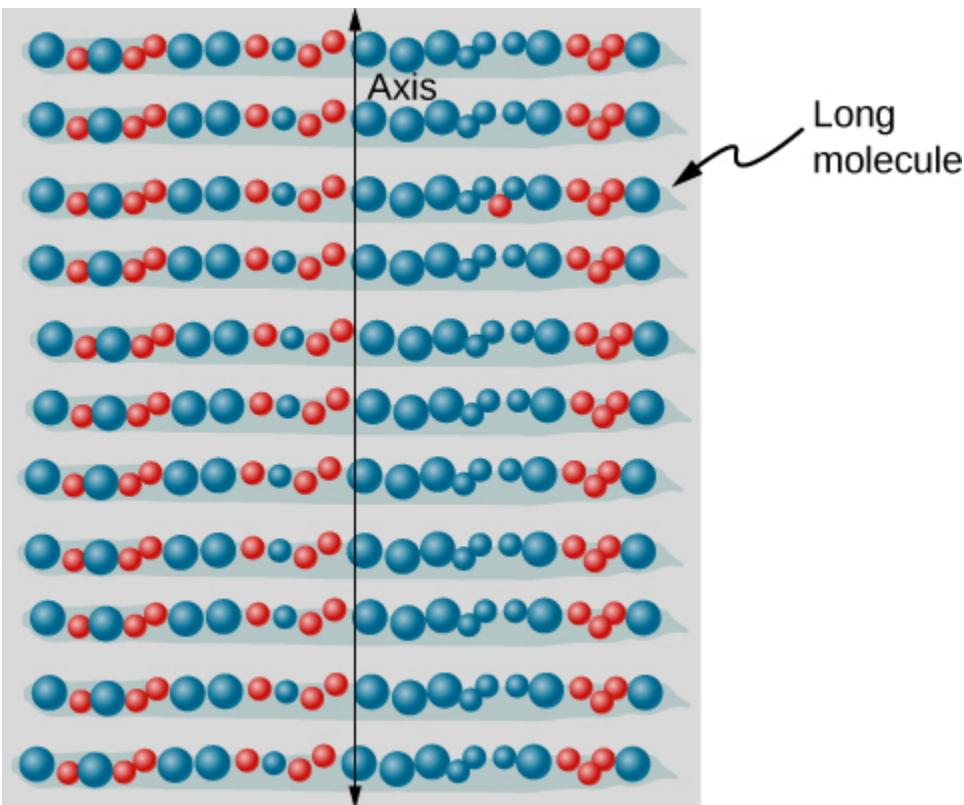
Check Your Understanding What happens at Brewster's angle if the original incident light is already 100% vertically polarized?

Solution:

There will be only refraction but no reflection.

Atomic Explanation of Polarizing Filters

Polarizing filters have a polarization axis that acts as a slit. This slit passes EM waves (often visible light) that have an electric field parallel to the axis. This is accomplished with long molecules aligned perpendicular to the axis, as shown in [\[link\]](#).



Long molecules are aligned perpendicular to the axis of a polarizing filter. In an EM wave, the component of the electric field perpendicular to these molecules passes through the filter, whereas the component parallel to the molecules is absorbed.

[\[link\]](#) illustrates how the component of the electric field parallel to the long molecules is absorbed. An EM wave is composed of oscillating electric and magnetic fields. The electric field is strong compared with the magnetic field and is more effective in exerting force on charges in the molecules. The most affected charged particles are the electrons, since electron masses are small. If an electron is forced to oscillate, it can absorb energy from the EM wave. This reduces the field in the wave and, hence, reduces its intensity. In long molecules, electrons can more easily oscillate parallel to the molecule than in the perpendicular direction. The electrons are bound to the molecule and are more restricted in their movement perpendicular to the molecule. Thus, the electrons can absorb EM waves that have a component of their electric field parallel to the molecule. The electrons are much less responsive to electric fields perpendicular to the molecule and allow these fields to pass. Thus, the axis of the polarizing filter is perpendicular to the length of the molecule.

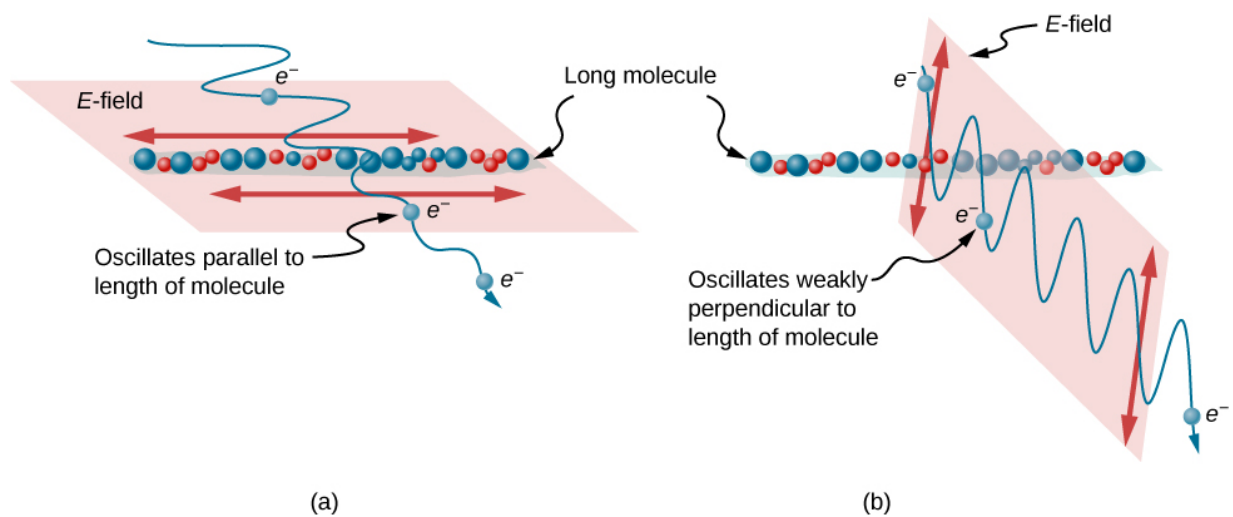
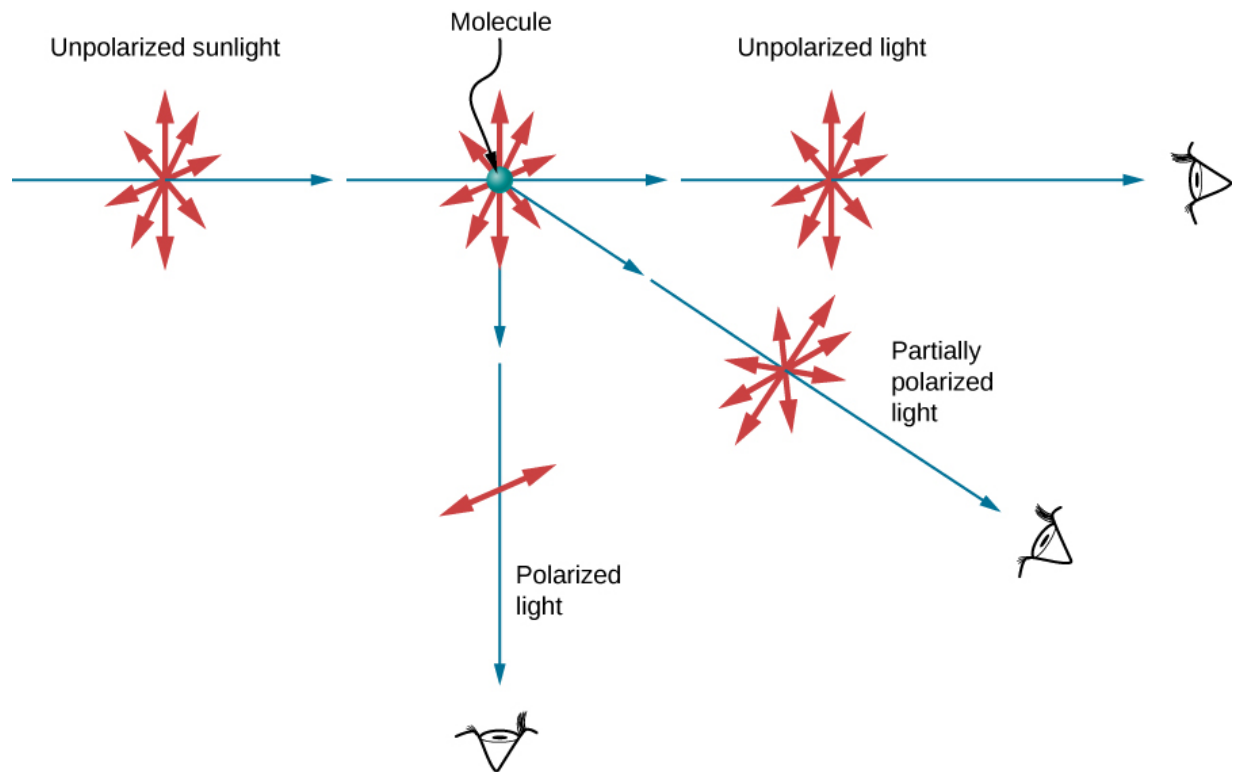


Diagram of an electron in a long molecule oscillating parallel to the molecule. The oscillation of the electron absorbs energy and reduces the intensity of the component of the EM wave that is parallel to the molecule.

Polarization by Scattering

If you hold your polarizing sunglasses in front of you and rotate them while looking at blue sky, you will see the sky get bright and dim. This is a clear indication that light scattered by air is partially polarized. [\[link\]](#) helps illustrate how this happens. Since light is a transverse EM wave, it vibrates the electrons of air molecules perpendicular to the direction that it is traveling. The electrons then radiate like small antennae. Since they are oscillating perpendicular to the direction of the light ray, they produce EM radiation that is polarized perpendicular to the direction of the ray. When viewing the light along a line perpendicular to the original ray, as in the figure, there can be no polarization in the scattered light parallel to the original ray, because that would require the original ray to be a longitudinal wave. Along other directions, a component of the other polarization can be projected along the line of sight, and the scattered light is only partially polarized. Furthermore, multiple scattering can bring light to your eyes from other directions and can contain different polarizations.



Polarization by scattering. Unpolarized light scattering from air molecules shakes their electrons perpendicular to the direction of the original ray. The scattered light therefore has a polarization perpendicular to the original direction and none parallel to the original direction.

Photographs of the sky can be darkened by polarizing filters, a trick used by many photographers to make clouds brighter by contrast. Scattering from other particles, such as smoke or dust, can also polarize light. Detecting polarization in scattered EM waves can be a useful analytical tool in determining the scattering source.

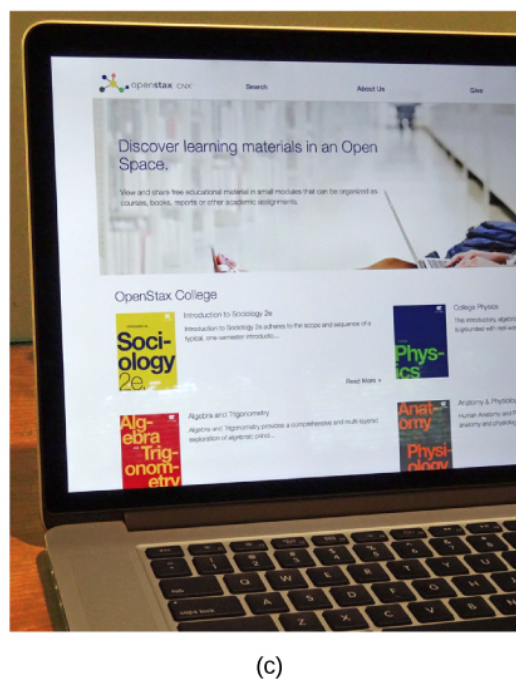
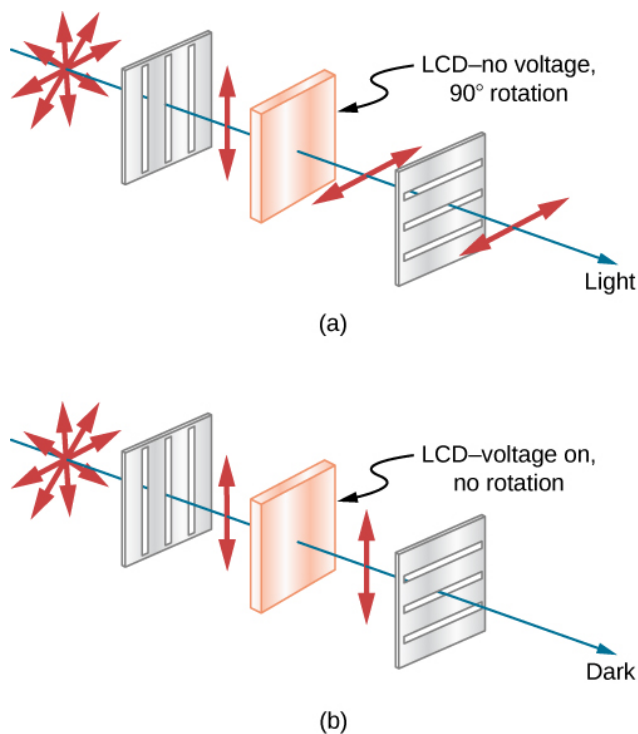
A range of optical effects are used in sunglasses. Besides being polarizing, sunglasses may have colored pigments embedded in them, whereas others use either a nonreflective or reflective coating. A recent development is photochromic lenses, which darken in the sunlight and become clear indoors. Photochromic lenses are embedded with organic microcrystalline

molecules that change their properties when exposed to UV in sunlight, but become clear in artificial lighting with no UV.

Liquid Crystals and Other Polarization Effects in Materials

Although you are undoubtedly aware of liquid crystal displays (LCDs) found in watches, calculators, computer screens, cellphones, flat screen televisions, and many other places, you may not be aware that they are based on polarization. Liquid crystals are so named because their molecules can be aligned even though they are in a liquid. Liquid crystals have the property that they can rotate the polarization of light passing through them by 90° . Furthermore, this property can be turned off by the application of a voltage, as illustrated in [\[link\]](#). It is possible to manipulate this characteristic quickly and in small, well-defined regions to create the contrast patterns we see in so many LCD devices.

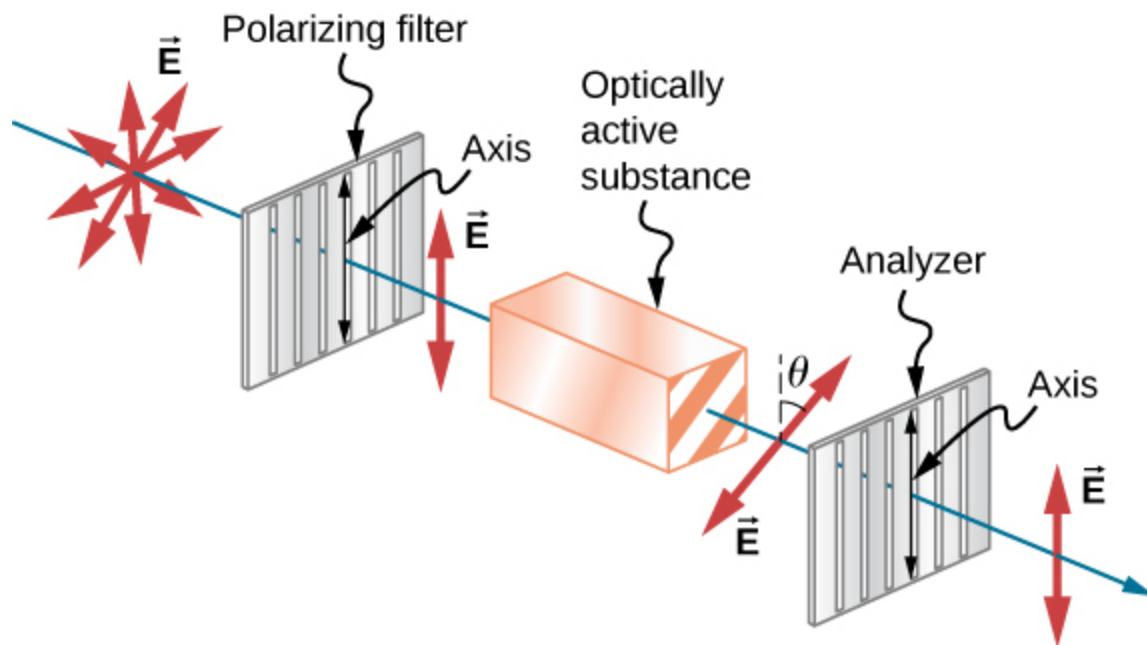
In flat screen LCD televisions, a large light is generated at the back of the TV. The light travels to the front screen through millions of tiny units called pixels (picture elements). One of these is shown in [\[link\]](#)(a) and (b). Each unit has three cells, with red, blue, or green filters, each controlled independently. When the voltage across a liquid crystal is switched off, the liquid crystal passes the light through the particular filter. We can vary the picture contrast by varying the strength of the voltage applied to the liquid crystal.



(a) Polarized light is rotated 90° by a liquid crystal and then passed by a polarizing filter that has its axis perpendicular to the direction of the original polarization. (b) When a voltage is applied to the liquid crystal, the polarized light is not rotated and is blocked by the filter, making the region dark in comparison with its surroundings. (c) LCDs can be made color specific, small, and fast enough to use in laptop computers and TVs. (credit c: modification of work by Jane Whitney)

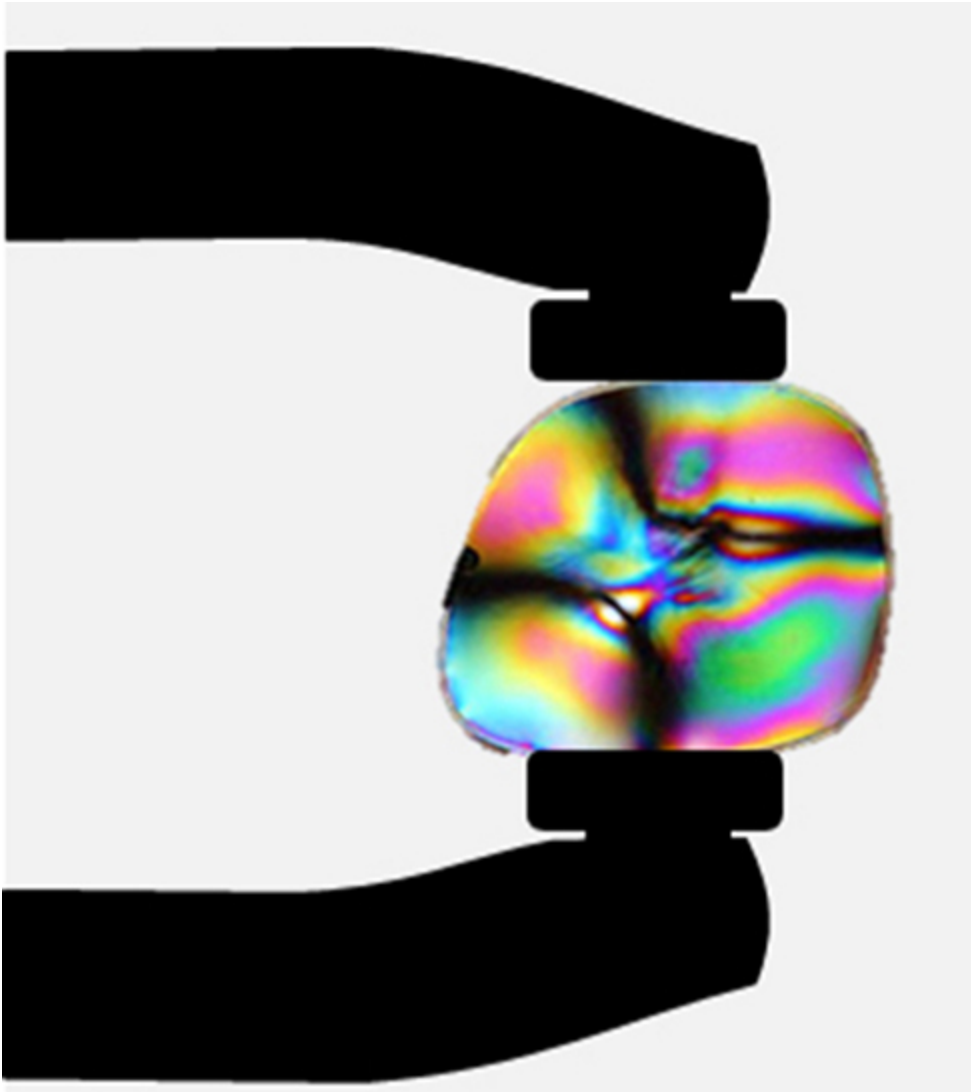
Many crystals and solutions rotate the plane of polarization of light passing through them. Such substances are said to be **optically active**. Examples include sugar water, insulin, and collagen ([link](#)). In addition to depending on the type of substance, the amount and direction of rotation depend on several other factors. Among these is the concentration of the substance, the distance the light travels through it, and the wavelength of light. Optical activity is due to the asymmetrical shape of molecules in the substance, such as being helical. Measurements of the rotation of polarized light passing through substances can thus be used to measure concentrations, a standard technique for sugars. It can also give information on the shapes of

molecules, such as proteins, and factors that affect their shapes, such as temperature and pH.



Optical activity is the ability of some substances to rotate the plane of polarization of light passing through them. The rotation is detected with a polarizing filter or analyzer.

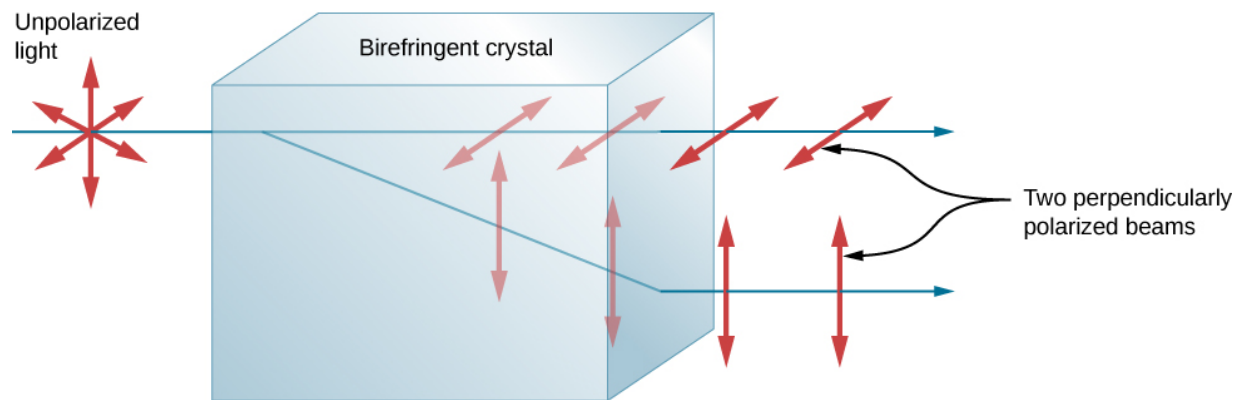
Glass and plastic become optically active when stressed: the greater the stress, the greater the effect. Optical stress analysis on complicated shapes can be performed by making plastic models of them and observing them through crossed filters, as seen in [\[link\]](#). It is apparent that the effect depends on wavelength as well as stress. The wavelength dependence is sometimes also used for artistic purposes.



Optical stress analysis of a plastic lens placed between crossed polarizers. (credit: “Infopro”/Wikimedia Commons)

Another interesting phenomenon associated with polarized light is the ability of some crystals to split an unpolarized beam of light into two polarized beams. This occurs because the crystal has one value for the index of refraction of polarized light but a different value for the index of refraction of light polarized in the perpendicular direction, so that each component has its own angle of refraction. Such crystals are said to be

birefringent, and, when aligned properly, two perpendicularly polarized beams will emerge from the crystal ([link](#)). Birefringent crystals can be used to produce polarized beams from unpolarized light. Some birefringent materials preferentially absorb one of the polarizations. These materials are called dichroic and can produce polarization by this preferential absorption. This is fundamentally how polarizing filters and other polarizers work.



Birefringent materials, such as the common mineral calcite, split unpolarized beams of light into two with two different values of index of refraction.

Summary

- Polarization is the attribute that wave oscillations have a definite direction relative to the direction of propagation of the wave. The direction of polarization is defined to be the direction parallel to the electric field of the EM wave.
- Unpolarized light is composed of many rays having random polarization directions.
- Unpolarized light can be polarized by passing it through a polarizing filter or other polarizing material. The process of polarizing light decreases its intensity by a factor of 2.

- The intensity, I , of polarized light after passing through a polarizing filter is $I = I_0 \cos^2 \theta$, where I_0 is the incident intensity and θ is the angle between the direction of polarization and the axis of the filter.
- Polarization is also produced by reflection.
- Brewster's law states that reflected light is completely polarized at the angle of reflection θ_b , known as Brewster's angle.
- Polarization can also be produced by scattering.
- Several types of optically active substances rotate the direction of polarization of light passing through them.

Key Equations

Speed of light	$c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}$
Index of refraction	$n = \frac{c}{v}$
Law of reflection	$\theta_r = \theta_i$
Law of refraction (Snell's law)	$n_1 \sin \theta_1 = n_2 \sin \theta_2$
Critical angle	$\theta_c = \sin^{-1} \left(\frac{n_2}{n_1} \right)$ for $n_1 > n_2$
Malus's law	$I = I_0 \cos^2 \theta$
Brewster's law	$\tan \theta_b = \frac{n_2}{n_1}$

Conceptual Questions

Exercise:

Problem: Can a sound wave in air be polarized? Explain.

Solution:

No. Sound waves are not transverse waves.

Exercise:

Problem:

No light passes through two perfect polarizing filters with perpendicular axes. However, if a third polarizing filter is placed between the original two, some light can pass. Why is this? Under what circumstances does most of the light pass?

Exercise:

Problem:

Explain what happens to the energy carried by light that it is dimmed by passing it through two crossed polarizing filters.

Solution:

Energy is absorbed into the filters.

Exercise:

Problem:

When particles scattering light are much smaller than its wavelength, the amount of scattering is proportional to $\frac{1}{\lambda}$. Does this mean there is more scattering for small λ than large λ ? How does this relate to the fact that the sky is blue?

Exercise:

Problem:

Using the information given in the preceding question, explain why sunsets are red.

Solution:

Sunsets are viewed with light traveling straight from the Sun toward us. When blue light is scattered out of this path, the remaining red light dominates the overall appearance of the setting Sun.

Exercise:**Problem:**

When light is reflected at Brewster's angle from a smooth surface, it is 100% polarized parallel to the surface. Part of the light will be refracted into the surface. Describe how you would do an experiment to determine the polarization of the refracted light. What direction would you expect the polarization to have and would you expect it to be 100%?

Exercise:**Problem:**

If you lie on a beach looking at the water with your head tipped slightly sideways, your polarized sunglasses do not work very well. Why not?

Solution:

The axis of polarization for the sunglasses has been rotated 90° .

Problems**Exercise:**

Problem:

What angle is needed between the direction of polarized light and the axis of a polarizing filter to cut its intensity in half?

Exercise:**Problem:**

The angle between the axes of two polarizing filters is 45.0° . By how much does the second filter reduce the intensity of the light coming through the first?

Solution:

0.500

Exercise:**Problem:**

Two polarizing sheets P_1 and P_2 are placed together with their transmission axes oriented at an angle θ to each other. What is θ when only 25% of the maximum transmitted light intensity passes through them?

Exercise:**Problem:**

Suppose that in the preceding problem the light incident on P_1 is unpolarized. At the determined value of θ , what fraction of the incident light passes through the combination?

Solution:

0.125 or $1/8$

Exercise:

Problem:

If you have completely polarized light of intensity 150 W/m^2 , what will its intensity be after passing through a polarizing filter with its axis at an 89.0° angle to the light's polarization direction?

Exercise:**Problem:**

What angle would the axis of a polarizing filter need to make with the direction of polarized light of intensity 1.00 kW/m^2 to reduce the intensity to 10.0 W/m^2 ?

Solution:

84.3°

Exercise:**Problem:**

At the end of [\[link\]](#), it was stated that the intensity of polarized light is reduced to 90.0% of its original value by passing through a polarizing filter with its axis at an angle of 18.4° to the direction of polarization. Verify this statement.

Exercise:**Problem:**

Show that if you have three polarizing filters, with the second at an angle of 45.0° to the first and the third at an angle of 90.0° to the first, the intensity of light passed by the first will be reduced to 25.0% of its value. (This is in contrast to having only the first and third, which reduces the intensity to zero, so that placing the second between them increases the intensity of the transmitted light.)

Solution:

$$0.250 I_0$$

Exercise:**Problem:**

Three polarizing sheets are placed together such that the transmission axis of the second sheet is oriented at 25.0° to the axis of the first, whereas the transmission axis of the third sheet is oriented at 40.0° (in the same sense) to the axis of the first. What fraction of the intensity of an incident unpolarized beam is transmitted by the combination?

Exercise:**Problem:**

In order to rotate the polarization axis of a beam of linearly polarized light by 90.0° , a student places sheets P_1 and P_2 with their transmission axes at 45.0° and 90.0° , respectively, to the beam's axis of polarization. (a) What fraction of the incident light passes through P_1 and (b) through the combination? (c) Repeat your calculations for part (b) for transmission-axis angles of 30.0° and 90.0° , respectively.

Solution:

a. 0.500; b. 0.250; c. 0.187

Exercise:**Problem:**

It is found that when light traveling in water falls on a plastic block, Brewster's angle is 50.0° . What is the refractive index of the plastic?

Exercise:**Problem:**

At what angle will light reflected from diamond be completely polarized?

Solution:

67.54°

Exercise:

Problem:

What is Brewster's angle for light traveling in water that is reflected from crown glass?

Exercise:

Problem:

A scuba diver sees light reflected from the water's surface. At what angle relative to the water's surface will this light be completely polarized?

Solution:

53.1°

Additional Problems

Exercise:

Problem:

From his measurements, Roemer estimated that it took 22 min for light to travel a distance equal to the diameter of Earth's orbit around the Sun. (a) Use this estimate along with the known diameter of Earth's orbit to obtain a rough value of the speed of light. (b) Light actually takes 16.5 min to travel this distance. Use this time to calculate the speed of light.

Exercise:

Problem:

Cornu performed Fizeau's measurement of the speed of light using a wheel of diameter 4.00 cm that contained 180 teeth. The distance from the wheel to the mirror was 22.9 km. Assuming he measured the speed of light accurately, what was the angular velocity of the wheel?

Solution:

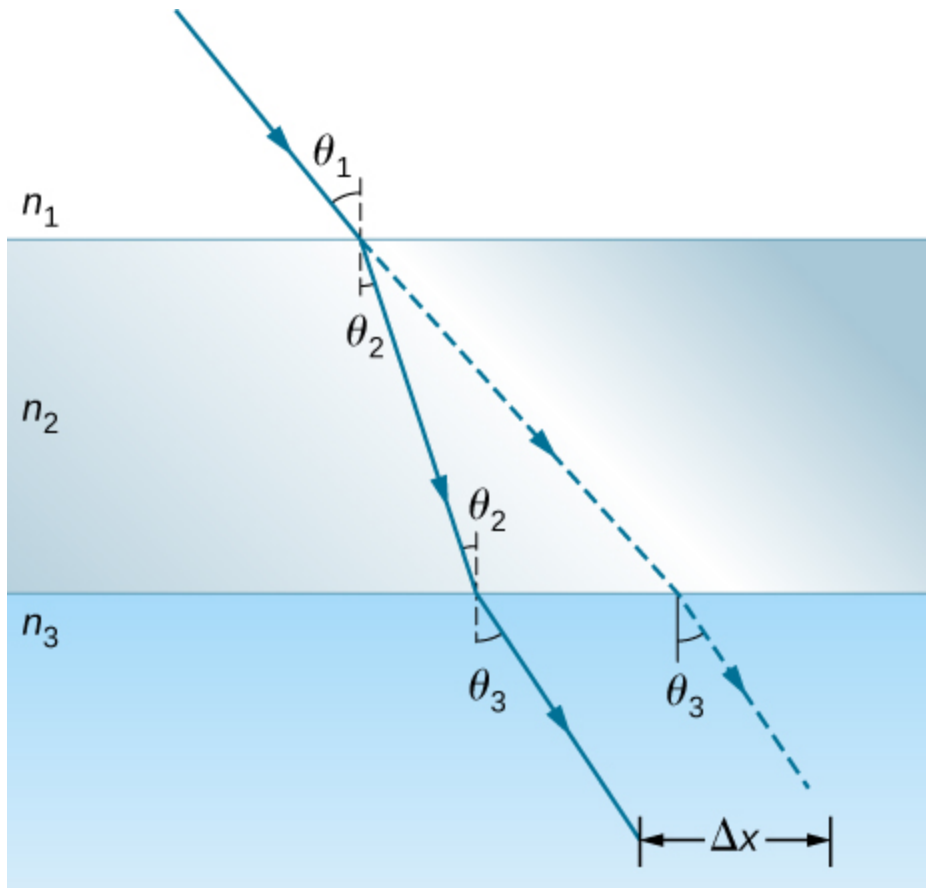
114 radian/s

Exercise:**Problem:**

Suppose you have an unknown clear substance immersed in water, and you wish to identify it by finding its index of refraction. You arrange to have a beam of light enter it at an angle of 45.0° , and you observe the angle of refraction to be 40.3° . What is the index of refraction of the substance and its likely identity?

Exercise:**Problem:**

Shown below is a ray of light going from air through crown glass into water, such as going into a fish tank. Calculate the amount the ray is displaced by the glass (Δx), given that the incident angle is 40.0° and the glass is 1.00 cm thick.



Solution:

3.72 mm

Exercise:

Problem:

Considering the previous problem, show that θ_3 is the same as it would be if the second medium were not present.

Exercise:

Problem:

At what angle is light inside crown glass completely polarized when reflected from water, as in a fish tank?

Solution:

41.2°

Exercise:

Problem:

Light reflected at 55.6° from a window is completely polarized. What is the window's index of refraction and the likely substance of which it is made?

Exercise:

Problem:

(a) Light reflected at 62.5° from a gemstone in a ring is completely polarized. Can the gem be a diamond? (b) At what angle would the light be completely polarized if the gem was in water?

Solution:

a. 1.92. The gem is not a diamond (it is zircon). b. 55.2°

Exercise:

Problem:

If θ_b is Brewster's angle for light reflected from the top of an interface between two substances, and θ'_b is Brewster's angle for light reflected from below, prove that $\theta_b + \theta'_b = 90.0^\circ$.

Exercise:

Problem:

Unreasonable results Suppose light travels from water to another substance, with an angle of incidence of 10.0° and an angle of refraction of 14.9°. (a) What is the index of refraction of the other substance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. 0.898; b. We cannot have $n < 1.00$, since this would imply a speed greater than c . c. The refracted angle is too big relative to the angle of incidence.

Exercise:

Problem:

Unreasonable results Light traveling from water to a gemstone strikes the surface at an angle of 80.0° and has an angle of refraction of 15.2° . (a) What is the speed of light in the gemstone? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem:

If a polarizing filter reduces the intensity of polarized light to 50.0% of its original value, by how much are the electric and magnetic fields reduced?

Solution:

$$0.707 B_1$$

Exercise:

Problem:

Suppose you put on two pairs of polarizing sunglasses with their axes at an angle of 15.0° . How much longer will it take the light to deposit a given amount of energy in your eye compared with a single pair of sunglasses? Assume the lenses are clear except for their polarizing characteristics.

Exercise:

Problem:

(a) On a day when the intensity of sunlight is 1.00 kW/m^2 , a circular lens 0.200 m in diameter focuses light onto water in a black beaker. Two polarizing sheets of plastic are placed in front of the lens with their axes at an angle of 20.0° . Assuming the sunlight is unpolarized and the polarizers are 100% efficient, what is the initial rate of heating of the water in $^\circ\text{C/s}$, assuming it is 80.0% absorbed? The aluminum beaker has a mass of 30.0 grams and contains 250 grams of water. (b) Do the polarizing filters get hot? Explain.

Solution:

a. $1.69 \times 10^{-2} \text{ }^\circ\text{C/s}$; b. yes

Challenge Problems**Exercise:****Problem:**

Light shows staged with lasers use moving mirrors to swing beams and create colorful effects. Show that a light ray reflected from a mirror changes direction by 2θ when the mirror is rotated by an angle θ .

Exercise:**Problem:**

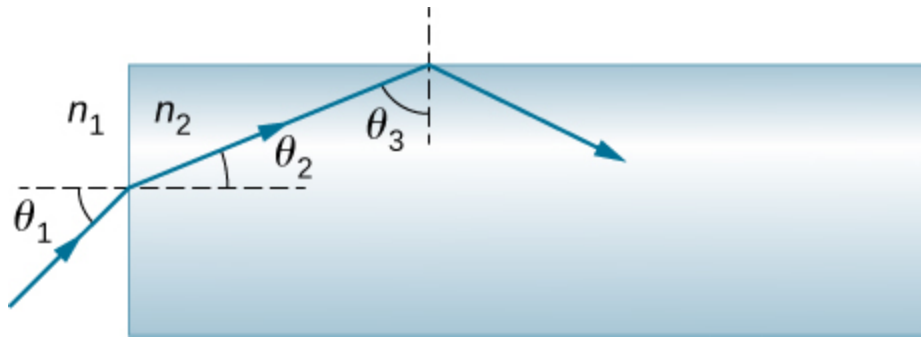
Consider sunlight entering Earth's atmosphere at sunrise and sunset—that is, at a 90.0° incident angle. Taking the boundary between nearly empty space and the atmosphere to be sudden, calculate the angle of refraction for sunlight. This lengthens the time the Sun appears to be above the horizon, both at sunrise and sunset. Now construct a problem in which you determine the angle of refraction for different models of the atmosphere, such as various layers of varying density. Your instructor may wish to guide you on the level of complexity to consider and on how the index of refraction varies with air density.

Solution:

First part: 88.6° . The remainder depends on the complexity of the solution the reader constructs.

Exercise:**Problem:**

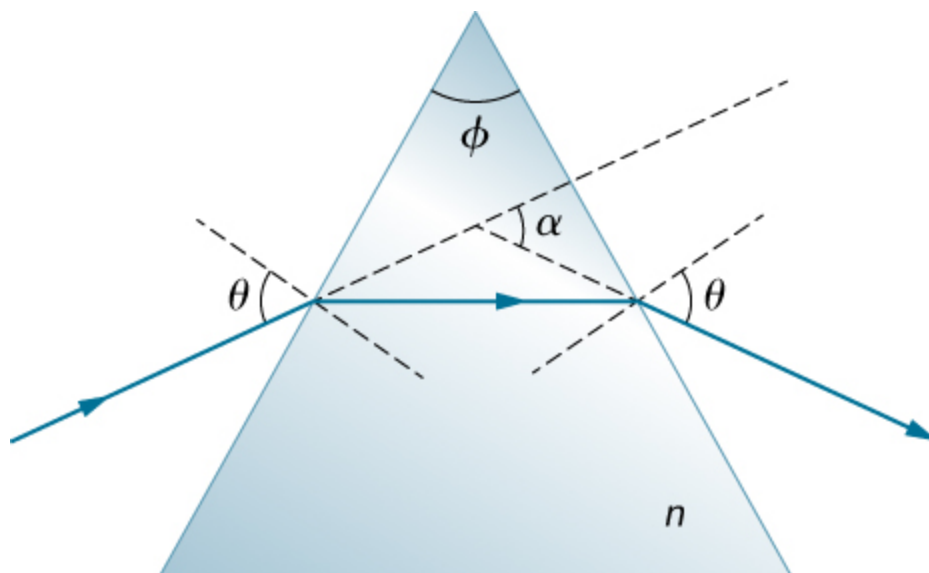
A light ray entering an optical fiber surrounded by air is first refracted and then reflected as shown below. Show that if the fiber is made from crown glass, any incident ray will be totally internally reflected.

**Exercise:****Problem:**

A light ray falls on the left face of a prism (see below) at the angle of incidence θ for which the emerging beam has an angle of refraction θ at the right face. Show that the index of refraction n of the glass prism is given by

$$n = \frac{\sin \frac{1}{2}(\alpha + \phi)}{\sin \frac{1}{2}\phi}$$

where ϕ is the vertex angle of the prism and α is the angle through which the beam has been deviated. If $\alpha = 37.0^\circ$ and the base angles of the prism are each 50.0° , what is n ?



Solution:

proof; 1.33

Exercise:

Problem:

If the apex angle ϕ in the previous problem is 20.0° and $n = 1.50$, what is the value of α ?

Exercise:

Problem:

The light incident on polarizing sheet P_1 is linearly polarized at an angle of 30.0° with respect to the transmission axis of P_1 . Sheet P_2 is placed so that its axis is parallel to the polarization axis of the incident light, that is, also at 30.0° with respect to P_1 . (a) What fraction of the incident light passes through P_1 ? (b) What fraction of the incident light is passed by the combination? (c) By rotating P_2 , a maximum in transmitted intensity is obtained. What is the ratio of this maximum intensity to the intensity of transmitted light when P_2 is at 30.0° with respect to P_1 ?

Solution:

a. 0.750; b. 0.563; c. 1.33

Exercise:

Problem:

Prove that if I is the intensity of light transmitted by two polarizing filters with axes at an angle θ and I' is the intensity when the axes are at an angle $90.0^\circ - \theta$, then $I + I' = I_0$, the original intensity. (*Hint: Use the trigonometric identities $\cos 90.0^\circ - \theta = \sin \theta$ and $\cos^2 \theta + \sin^2 \theta = 1$.*)

Glossary

birefringent

refers to crystals that split an unpolarized beam of light into two beams

Brewster's angle

angle of incidence at which the reflected light is completely polarized

Brewster's law

$\tan \theta_b = \frac{n_2}{n_1}$, where n_1 is the medium in which the incident and reflected light travel and n_2 is the index of refraction of the medium that forms the interface that reflects the light

direction of polarization

direction parallel to the electric field for EM waves

horizontally polarized

oscillations are in a horizontal plane

Malus's law

where I_0 is the intensity of the polarized wave before passing through the filter

optically active

substances that rotate the plane of polarization of light passing through them

polarization

attribute that wave oscillations have a definite direction relative to the direction of propagation of the wave

polarized

refers to waves having the electric and magnetic field oscillations in a definite direction

unpolarized

refers to waves that are randomly polarized

vertically polarized

oscillations are in a vertical plane

Introduction

class="introduction"

Cloud Gate
is a public
sculpture by
Anish

Kapoor
located in
Millennium
Park in

Chicago. Its
stainless
steel plates
reflect and
distort
images
around it,
including
the Chicago
skyline.

Dedicated in
2006, it has
become a
popular
tourist
attraction,
illustrating
how art can
use the
principles of
physical
optics to
startle and
entertain.

(credit:
modificatio

n of work
by Dhilung
Kirat)



This chapter introduces the major ideas of geometric optics, which describe the formation of images due to reflection and refraction. It is called “geometric” optics because the images can be characterized using geometric constructions, such as ray diagrams. We have seen that visible light is an electromagnetic wave; however, its wave nature becomes evident only when light interacts with objects with dimensions comparable to the wavelength (about 500 nm for visible light). Therefore, the laws of geometric optics only apply to light interacting with objects much larger than the wavelength of the light.

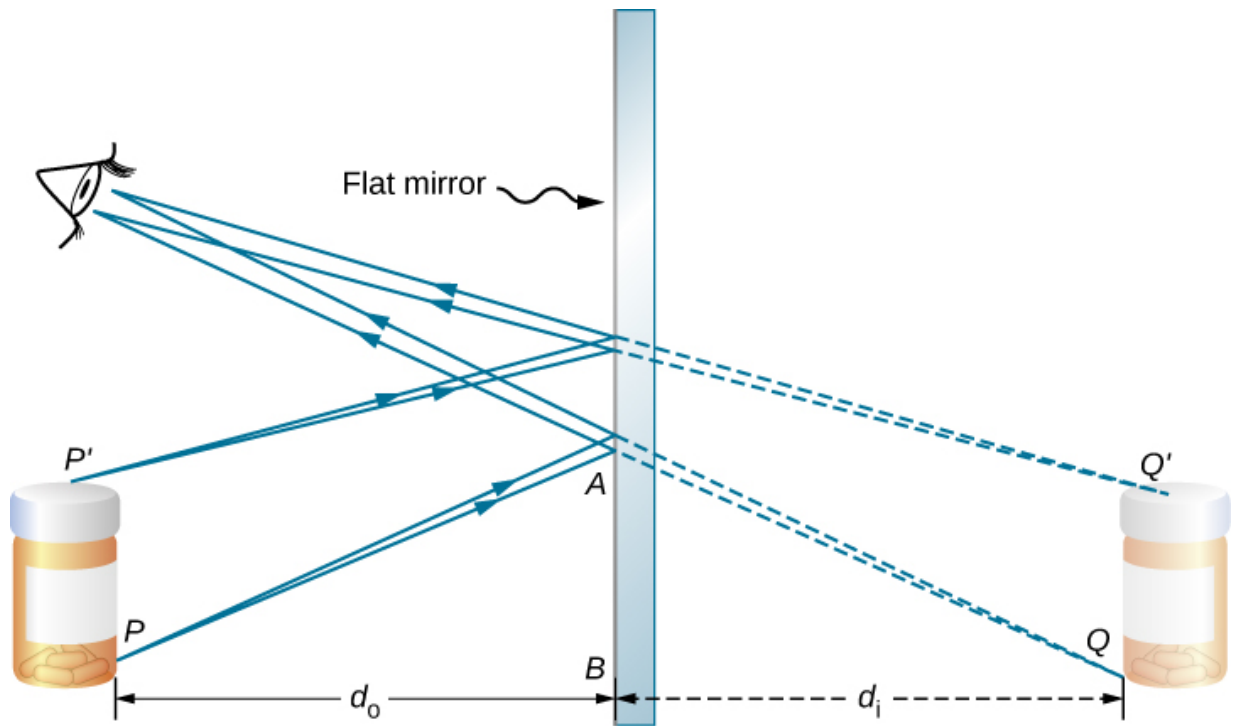
Images Formed by Plane Mirrors

By the end of this section, you will be able to:

- Describe how an image is formed by a plane mirror.
- Distinguish between real and virtual images.
- Find the location and characterize the orientation of an image created by a plane mirror.

You only have to look as far as the nearest bathroom to find an example of an image formed by a mirror. Images in a **plane mirror** are the same size as the object, are located behind the mirror, and are oriented in the same direction as the object (i.e., “upright”).

To understand how this happens, consider [\[link\]](#). Two rays emerge from point P , strike the mirror, and reflect into the observer’s eye. Note that we use the law of reflection to construct the reflected rays. If the reflected rays are extended backward behind the mirror (see dashed lines in [\[link\]](#)), they seem to originate from point Q . This is where the image of point P is located. If we repeat this process for point P' , we obtain its image at point Q' . You should convince yourself by using basic geometry that the image height (the distance from Q to Q') is the same as the object height (the distance from P to P'). By forming images of all points of the object, we obtain an upright image of the object behind the mirror.



Two light rays originating from point P on an object are reflected by a flat mirror into the eye of an observer. The reflected rays are obtained by using the law of reflection. Extending these reflected rays backward, they seem to come from point Q behind the mirror, which is where the virtual image is located. Repeating this process for point P' gives the image point Q' . The image height is thus the same as the object height, the image is upright, and the object distance d_o is the same as the image distance d_i . (credit: modification of work by Kevin Dufendach)

Notice that the reflected rays appear to the observer to come directly from the image behind the mirror. In reality, these rays come from the points on the mirror where they are reflected. The image behind the mirror is called a **virtual image** because it cannot be projected onto a screen—the rays only appear to originate from a common point behind the mirror. If you walk behind the mirror, you cannot see the image, because the rays do not go there. However, in front of the mirror, the rays behave exactly as if they come from behind the mirror, so that is where the virtual image is located.

Later in this chapter, we discuss real images; a **real image** can be projected onto a screen because the rays physically go through the image. You can certainly see both real and virtual images. The difference is that a virtual image cannot be projected onto a screen, whereas a real image can.

Locating an Image in a Plane Mirror

The law of reflection tells us that the angle of incidence is the same as the angle of reflection. Applying this to triangles PAB and QAB in [\[link\]](#) and using basic geometry shows that they are congruent triangles. This means that the distance PB from the object to the mirror is the same as the distance BQ from the mirror to the image. The **object distance** (denoted d_o) is the distance from the mirror to the object (or, more generally, from the center of the optical element that creates its image). Similarly, the **image distance** (denoted d_i) is the distance from the mirror to the image (or, more generally, from the center of the optical element that creates it). If we measure distances from the mirror, then the object and image are in opposite directions, so for a plane mirror, the object and image distances should have the opposite signs:

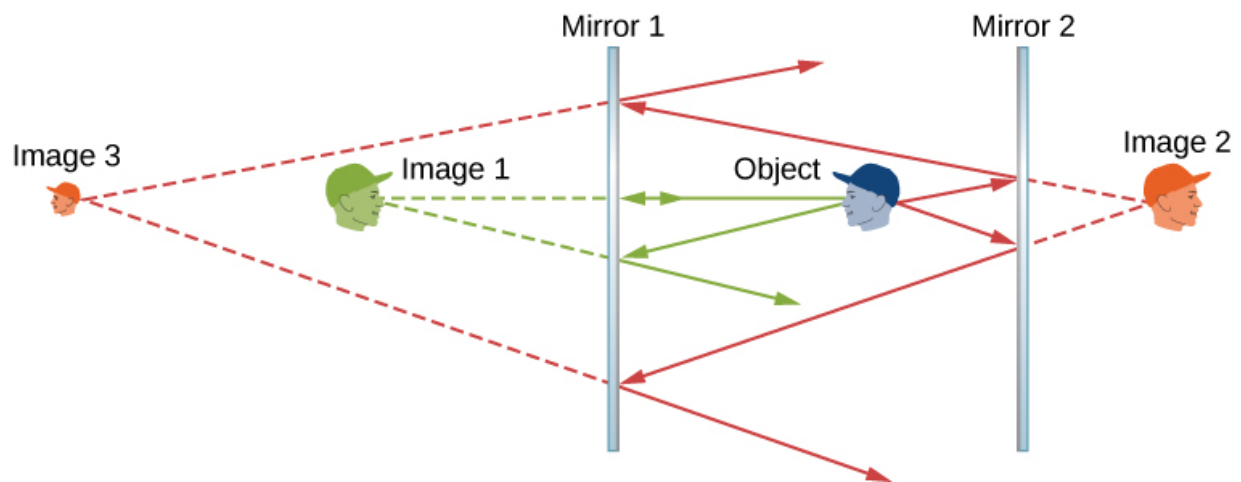
Note:
Equation:

$$d_o = -d_i.$$

An extended object such as the container in [\[link\]](#) can be treated as a collection of points, and we can apply the method above to locate the image of each point on the extended object, thus forming the extended image.

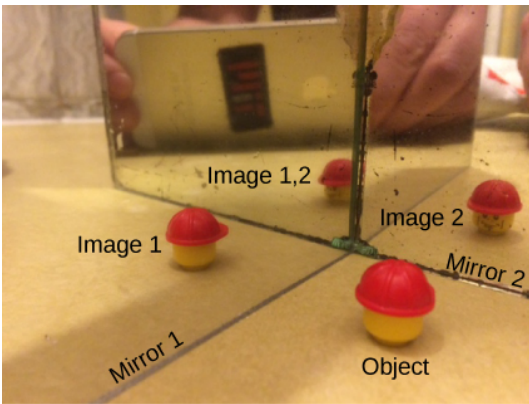
Multiple Images

If an object is situated in front of two mirrors, you may see images in both mirrors. In addition, the image in the first mirror may act as an object for the second mirror, so the second mirror may form an image of the image. If the mirrors are placed parallel to each other and the object is placed at a point other than the midpoint between them, then this process of image-of-an-image continues without end, as you may have noticed when standing in a hallway with mirrors on each side. This is shown in [\[link\]](#), which shows three images produced by the blue object. Notice that each reflection reverses front and back, just like pulling a right-hand glove inside out produces a left-hand glove (this is why a reflection of your right hand is a left hand). Thus, the fronts and backs of images 1 and 2 are both inverted with respect to the object, and the front and back of image 3 is inverted with respect to image 2, which is the object for image 3.

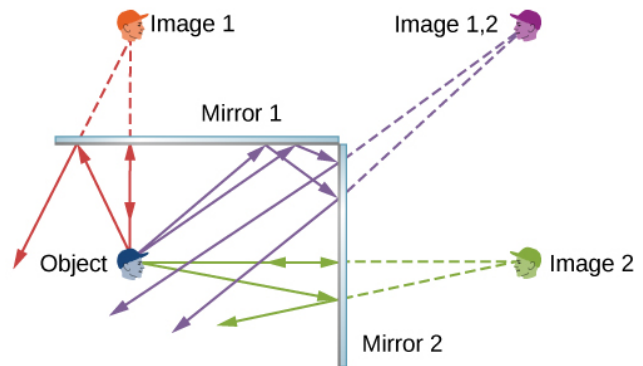


Two parallel mirrors can produce, in theory, an infinite number of images of an object placed off center between the mirrors. Three of these images are shown here. The front and back of each image is inverted with respect to its object. Note that the colors are only to identify the images. For normal mirrors, the color of an image is essentially the same as that of its object.

Infinite reflections may terminate. For instance, two mirrors at right angles form three images, as shown in part (a) of [\[link\]](#). Images 1 and 2 result from rays that reflect from only a single mirror, but image 1,2 is formed by rays that reflect from both mirrors. This is shown in the ray-tracing diagram in part (b) of [\[link\]](#). To find image 1,2, you have to look behind the corner of the two mirrors.



(a)



(b)

Two mirrors can produce multiple images. (a) Three images of a plastic head are visible in the two mirrors at a right angle. (b) A single object reflecting from two mirrors at a right angle can produce three images, as shown by the green, purple, and red images.

Summary

- A plane mirror always forms a virtual image (behind the mirror).
- The image and object are the same distance from a flat mirror, the image size is the same as the object size, and the image is upright.

Conceptual Questions

Exercise:

Problem:

What are the differences between real and virtual images? How can you tell (by looking) whether an image formed by a single lens or mirror is real or virtual?

Solution:

Virtual image cannot be projected on a screen. You cannot distinguish a real image from a virtual image simply by judging from the image perceived with your eye.

Exercise:

Problem: Can you see a virtual image? Explain your response.

Exercise:

Problem: Can you photograph a virtual image?

Solution:

Yes, you can photograph a virtual image. For example, if you photograph your reflection from a plane mirror, you get a photograph of a virtual image. The camera focuses the light that enters its lens to form an image; whether the source of the light is a real object or a reflection from mirror (i.e., a virtual image) does not matter.

Exercise:

Problem: Can you project a virtual image onto a screen?

Exercise:

Problem: Is it necessary to project a real image onto a screen to see it?

Solution:

No, you can see the real image the same way you can see the virtual image. The retina of your eye effectively serves as a screen.

Exercise:

Problem:

Devise an arrangement of mirrors allowing you to see the back of your head. What is the minimum number of mirrors needed for this task?

Exercise:

Problem:

If you wish to see your entire body in a flat mirror (from head to toe), how tall should the mirror be? Does its size depend upon your distance away from the mirror? Provide a sketch.

Solution:

The mirror should be half your size and its top edge should be at the level of your eyes. The size does not depend on your distance from the mirror.

Problems

Exercise:

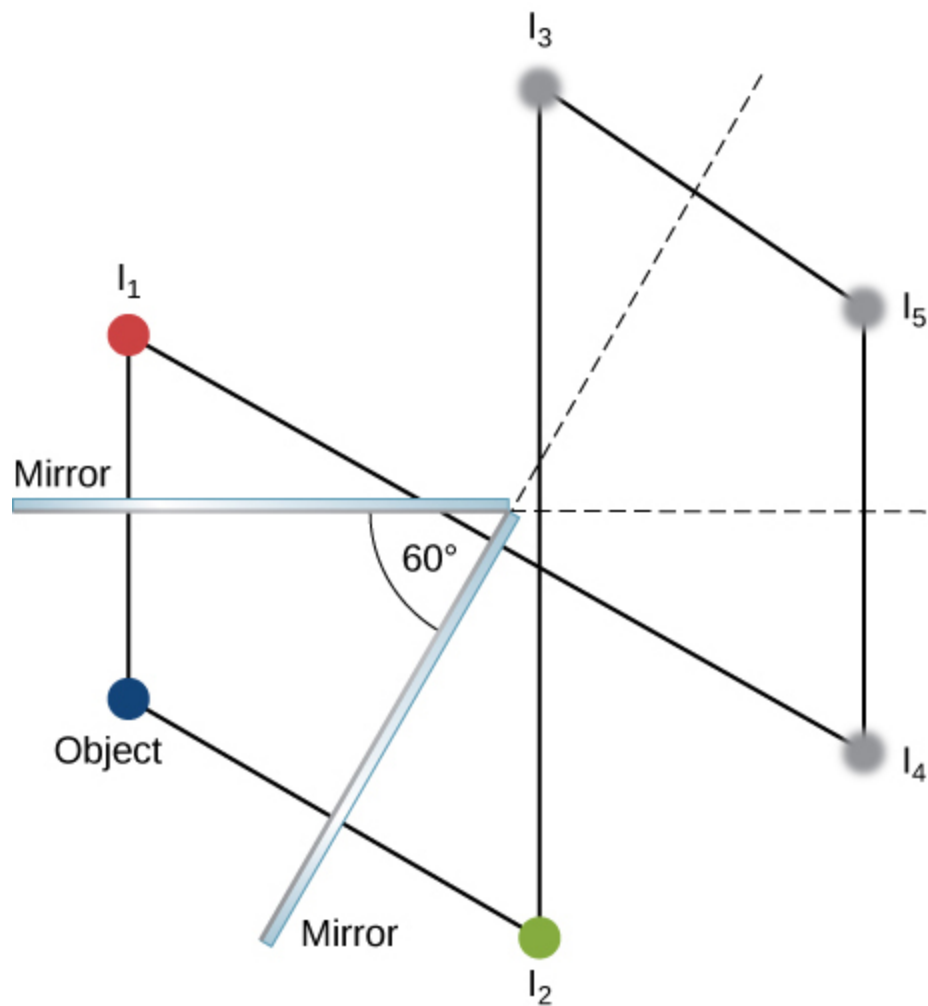
Problem:

Consider a pair of flat mirrors that are positioned so that they form an angle of 120° . An object is placed on the bisector between the mirrors. Construct a ray diagram as in [\[link\]](#) to show how many images are formed.

Exercise:

Problem:

Consider a pair of flat mirrors that are positioned so that they form an angle of 60° . An object is placed on the bisector between the mirrors. Construct a ray diagram as in [\[link\]](#) to show how many images are formed.

Solution:**Exercise:**

Problem:

By using more than one flat mirror, construct a ray diagram showing how to create an inverted image.

Glossary

plane mirror

plane (flat) reflecting surface

image distance

distance of the image from the central axis of the optical element that produces the image

object distance

distance of the object from the central axis of the optical element that produces its image

real image

image that can be projected onto a screen because the rays physically go through the image

virtual image

image that cannot be projected on a screen because the rays do not physically go through the image, they only appear to originate from the image

Spherical Mirrors

By the end of this section, you will be able to:

- Describe image formation by spherical mirrors.
- Use ray diagrams and the mirror equation to calculate the properties of an image in a spherical mirror.

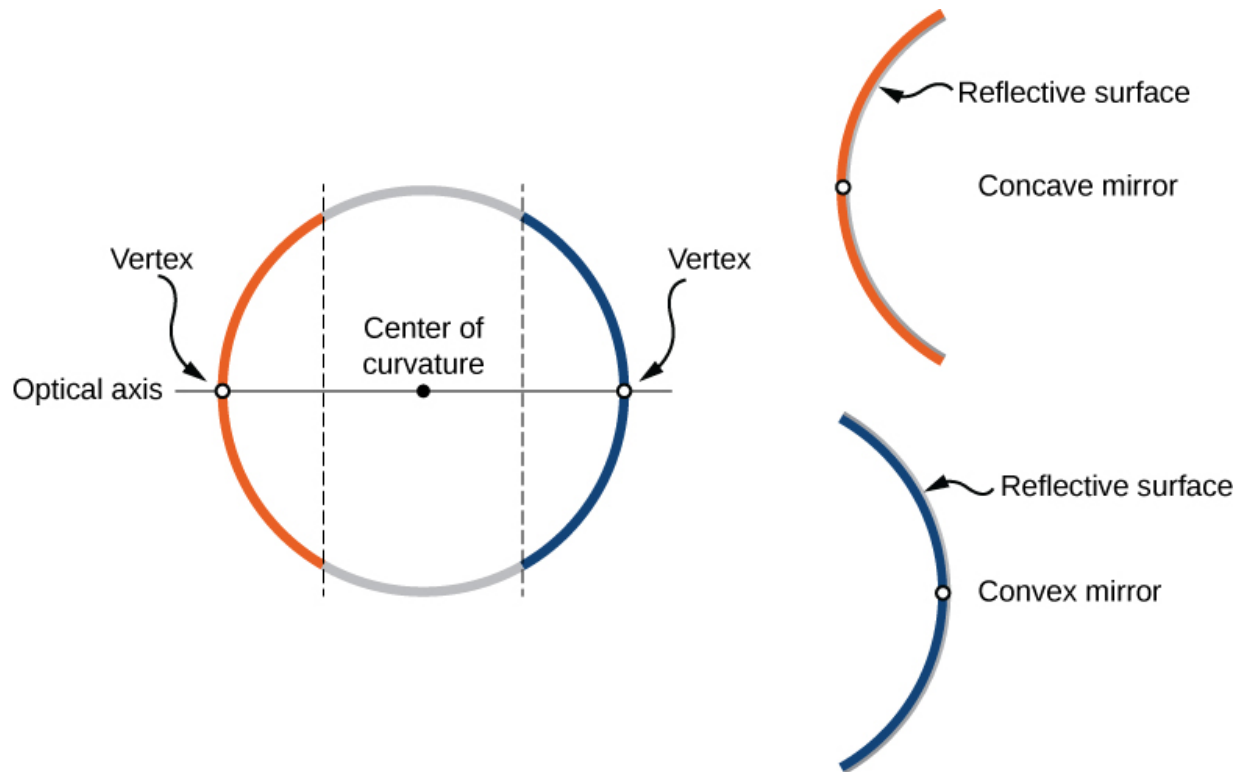
The image in a plane mirror has the same size as the object, is upright, and is the same distance behind the mirror as the object is in front of the mirror. A **curved mirror**, on the other hand, can form images that may be larger or smaller than the object and may form either in front of the mirror or behind it. In general, any curved surface will form an image, although some images may be so distorted as to be unrecognizable (think of fun house mirrors).

Because curved mirrors can create such a rich variety of images, they are used in many optical devices that find many uses. We will concentrate on spherical mirrors for the most part, because they are easier to manufacture than mirrors such as parabolic mirrors and so are more common.

Curved Mirrors

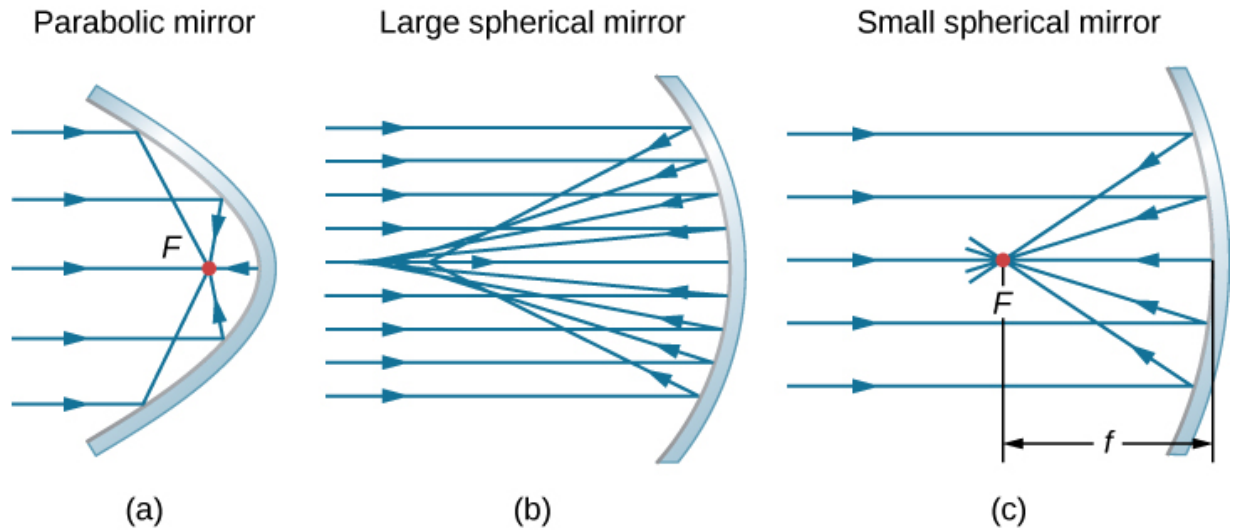
We can define two general types of spherical mirrors. If the reflecting surface is the outer side of the sphere, the mirror is called a **convex mirror**. If the inside surface is the reflecting surface, it is called a **concave mirror**.

Symmetry is one of the major hallmarks of many optical devices, including mirrors and lenses. The symmetry axis of such optical elements is often called the principal axis or **optical axis**. For a spherical mirror, the optical axis passes through the mirror's center of curvature and the mirror's **vertex**, as shown in [\[link\]](#).



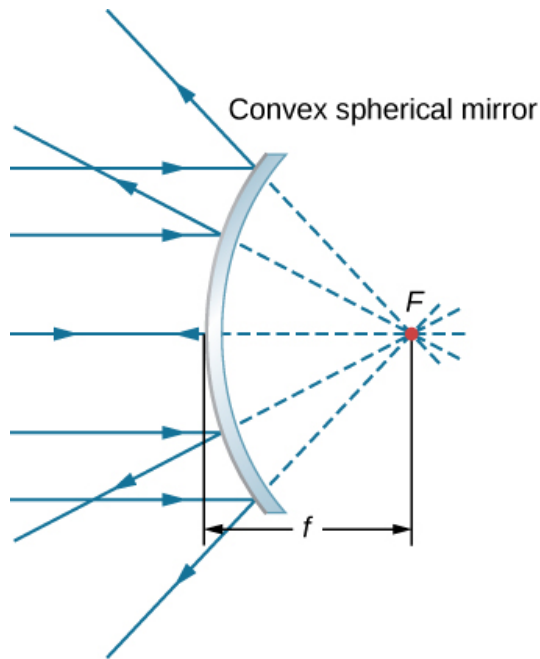
A spherical mirror is formed by cutting out a piece of a sphere and silvering either the inside or outside surface. A concave mirror has silvering on the interior surface (think “cave”), and a convex mirror has silvering on the exterior surface.

Consider rays that are parallel to the optical axis of a parabolic mirror, as shown in part (a) of [\[link\]](#). Following the law of reflection, these rays are reflected so that they converge at a point, called the **focal point**. Part (b) of this figure shows a spherical mirror that is large compared with its radius of curvature. For this mirror, the reflected rays do not cross at the same point, so the mirror does not have a well-defined focal point. This is called spherical aberration and results in a blurred image of an extended object. Part (c) shows a spherical mirror that is small compared to its radius of curvature. This mirror is a good approximation of a parabolic mirror, so rays that arrive parallel to the optical axis are reflected to a well-defined focal point. The distance along the optical axis from the mirror to the focal point is called the **focal length** of the mirror.



(a) Parallel rays reflected from a parabolic mirror cross at a single point called the focal point F . (b) Parallel rays reflected from a large spherical mirror do not cross at a common point. (c) If a spherical mirror is small compared with its radius of curvature, it better approximates the central part of a parabolic mirror, so parallel rays essentially cross at a common point. The distance along the optical axis from the mirror to the focal point is the focal length f of the mirror.

A convex spherical mirror also has a focal point, as shown in [\[link\]](#). Incident rays parallel to the optical axis are reflected from the mirror and seem to originate from point F at focal length f behind the mirror. Thus, the focal point is virtual because no real rays actually pass through it; they only appear to originate from it.



(a)



(b)

(a) Rays reflected by a convex spherical mirror: Incident rays of light parallel to the optical axis are reflected from a convex spherical mirror and seem to originate from a well-defined focal point at focal distance f on the opposite side of the mirror. The focal point is virtual because no real rays pass through it. (b) Photograph of a virtual image formed by a convex mirror. (credit b: modification of work by Jenny

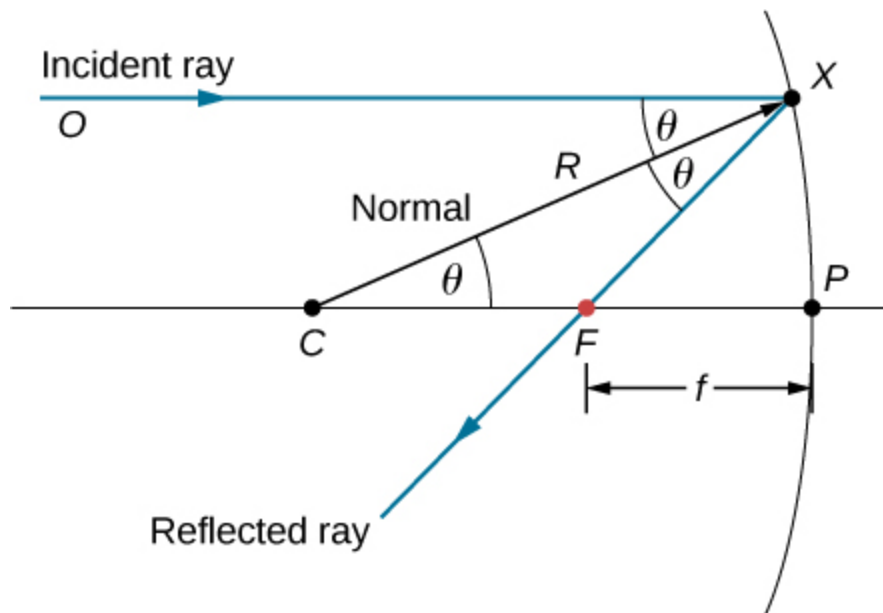
Downing)

How does the focal length of a mirror relate to the mirror's radius of curvature? [\[link\]](#) shows a single ray that is reflected by a spherical concave mirror. The incident ray is parallel to the optical axis. The point at which the reflected ray crosses the optical axis is the focal point. Note that all incident rays that are parallel to the optical axis are reflected through the focal point—we only show one ray for simplicity. We want to find how the focal length FP (denoted by f) relates to the radius of curvature of the mirror, R , whose length is $R = CF + FP$. The law of reflection tells us that angles OXC and CXF are the same, and because the incident ray is parallel to the optical axis, angles OXC and XCP are also the same. Thus, triangle CXF is an isosceles triangle with $CF = FX$. If the angle θ is small

(so that $\sin \theta \approx \theta$; this is called the “small-angle approximation”), then $FX \approx FP$ or $CF \approx FP$. Inserting this into the equation for the radius R , we get

Equation:

$$R = CF + FP = FP + FP = 2FP = 2f$$



Reflection in a concave mirror. In the small-angle approximation, a ray that is parallel to the optical axis CP is reflected through the focal point F of the mirror.

In other words, in the small-angle approximation, the focal length f of a concave spherical mirror is half of its radius of curvature, R :

Note:

Equation:

$$f = \frac{R}{2}.$$

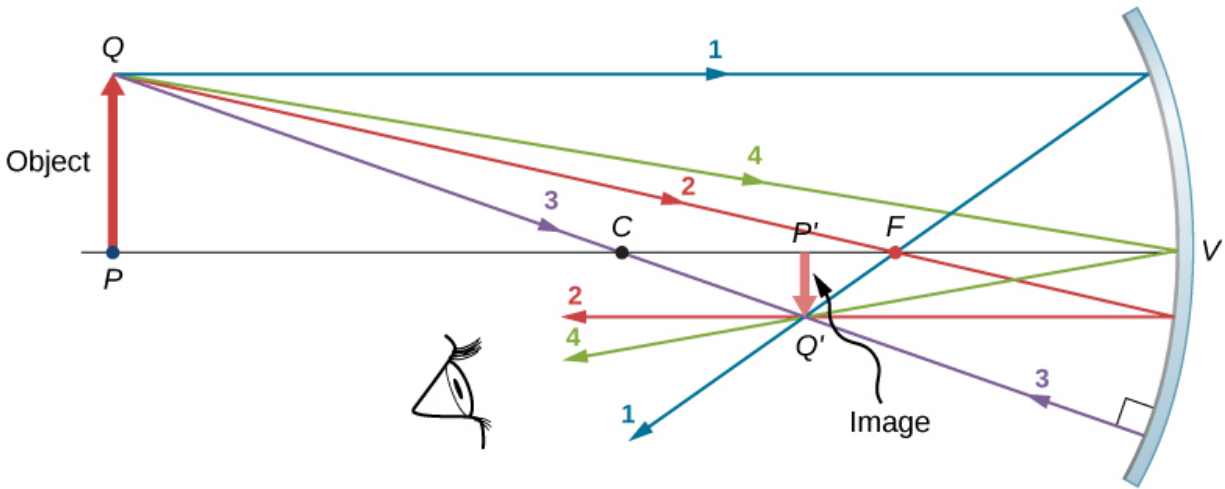
In this chapter, we assume that the **small-angle approximation** (also called the paraxial approximation) is always valid. In this approximation, all rays are paraxial rays, which means that they make a small angle with the optical axis and are at a distance much less than the radius of curvature from the optical axis. In this case, their angles θ of reflection are small angles, so $\sin \theta \approx \tan \theta \approx \theta$.

Using Ray Tracing to Locate Images

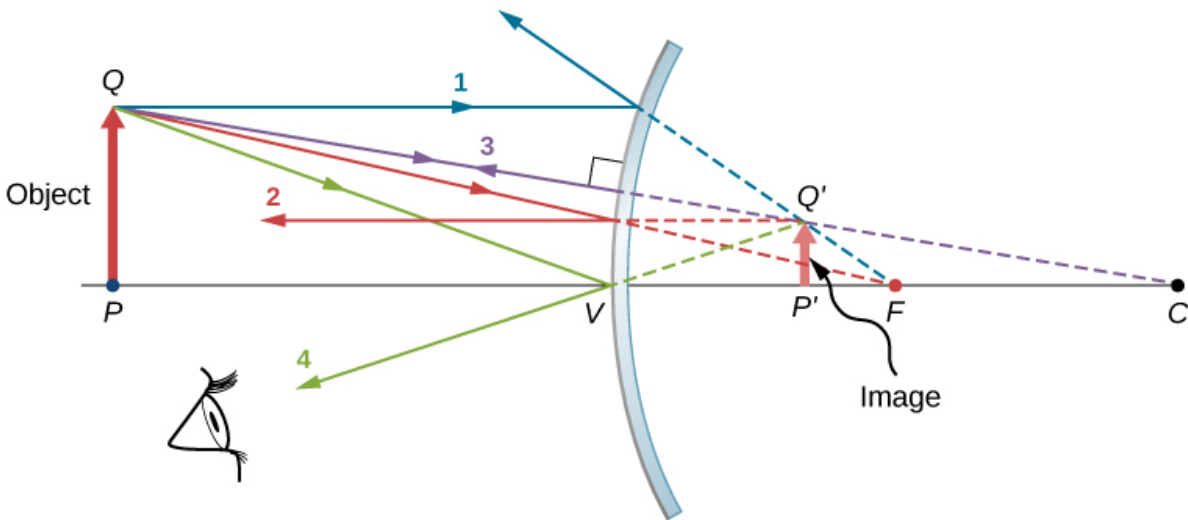
To find the location of an image formed by a spherical mirror, we first use ray tracing, which is the technique of drawing rays and using the law of reflection to determine the reflected rays (later, for lenses, we use the law of refraction to determine refracted rays). Combined with some basic geometry, we can use ray tracing to find the focal point, the image location, and other information about how a mirror manipulates light. In fact, we already used ray tracing above to locate the focal point of spherical mirrors, or the image distance of flat mirrors. To locate the image of an object, you must locate at least two points of the image. Locating each point requires drawing at least two rays from a point on the object and constructing their reflected rays. The point at which the reflected rays intersect, either in real space or in virtual space, is where the corresponding point of the image is located. To make ray tracing easier, we concentrate on four “principal” rays whose reflections are easy to construct.

[\[link\]](#) shows a concave mirror and a convex mirror, each with an arrow-shaped object in front of it. These are the objects whose images we want to locate by ray tracing. To do so, we draw rays from point Q that is on the object but not on the optical axis. We choose to draw our ray from the tip of the object. Principal ray 1 goes from point Q and travels parallel to the optical axis. The reflection of this ray must pass through the focal point, as discussed above. Thus, for the concave mirror, the reflection of principal

ray 1 goes through focal point F , as shown in part (b) of the figure. For the convex mirror, the backward extension of the reflection of principal ray 1 goes through the focal point (i.e., a virtual focus). Principal ray 2 travels first on the line going through the focal point and then is reflected back along a line parallel to the optical axis. Principal ray 3 travels toward the center of curvature of the mirror, so it strikes the mirror at normal incidence and is reflected back along the line from which it came. Finally, principal ray 4 strikes the vertex of the mirror and is reflected symmetrically about the optical axis.



(a)



(b)

The four principal rays shown for both (a) a concave mirror and (b) a convex mirror. The image forms where the rays intersect (for real images) or where their backward extensions intersect (for virtual images).

The four principal rays intersect at point Q' , which is where the image of point Q is located. To locate point Q' , drawing any two of these principle rays would suffice. We are thus free to choose whichever of the principal

rays we desire to locate the image. Drawing more than two principal rays is sometimes useful to verify that the ray tracing is correct.

To completely locate the extended image, we need to locate a second point in the image, so that we know how the image is oriented. To do this, we trace the principal rays from the base of the object. In this case, all four principal rays run along the optical axis, reflect from the mirror, and then run back along the optical axis. The difficulty is that, because these rays are collinear, we cannot determine a unique point where they intersect. All we know is that the base of the image is on the optical axis. However, because the mirror is symmetrical from top to bottom, it does not change the vertical orientation of the object. Thus, because the object is vertical, the image must be vertical. Therefore, the image of the base of the object is on the optical axis directly above the image of the tip, as drawn in the figure.

For the concave mirror, the extended image in this case forms between the focal point and the center of curvature of the mirror. It is inverted with respect to the object, is a real image, and is smaller than the object. Were we to move the object closer to or farther from the mirror, the characteristics of the image would change. For example, we show, as a later exercise, that an object placed between a concave mirror and its focal point leads to a virtual image that is upright and larger than the object. For the convex mirror, the extended image forms between the focal point and the mirror. It is upright with respect to the object, is a virtual image, and is smaller than the object.

Summary of Ray-Tracing Rules

Ray tracing is very useful for mirrors. The rules for ray tracing are summarized here for reference:

- A ray travelling parallel to the optical axis of a spherical mirror is reflected along a line that goes through the focal point of the mirror (ray 1 in [\[link\]](#)).
- A ray travelling along a line that goes through the focal point of a spherical mirror is reflected along a line parallel to the optical axis of the mirror (ray 2 in [\[link\]](#)).

- A ray travelling along a line that goes through the center of curvature of a spherical mirror is reflected back along the same line (ray 3 in [\[link\]](#)).
- A ray that strikes the vertex of a spherical mirror is reflected symmetrically about the optical axis of the mirror (ray 4 in [\[link\]](#)).

We use ray tracing to illustrate how images are formed by mirrors and to obtain numerical information about optical properties of the mirror. If we assume that a mirror is small compared with its radius of curvature, we can also use algebra and geometry to derive a mirror equation, which we do in the next section. Combining ray tracing with the mirror equation is a good way to analyze mirror systems.

Image Formation by Reflection—The Mirror Equation

For a plane mirror, we showed that the image formed has the same height and orientation as the object, and it is located at the same distance behind the mirror as the object is in front of the mirror. Although the situation is a bit more complicated for curved mirrors, using geometry leads to simple formulas relating the object and image distances to the focal lengths of concave and convex mirrors.

Consider the object OP shown in [\[link\]](#). The center of curvature of the mirror is labeled C and is a distance R from the vertex of the mirror, as marked in the figure. The object and image distances are labeled d_o and d_i , and the object and image heights are labeled h_o and h_i , respectively. Because the angles ϕ and ϕ' are alternate interior angles, we know that they have the same magnitude. However, they must differ in sign if we measure angles from the optical axis, so $\phi = -\phi'$. An analogous scenario holds for the angles θ and θ' . The law of reflection tells us that they have the same magnitude, but their signs must differ if we measure angles from the optical axis. Thus, $\theta = -\theta'$. Taking the tangent of the angles θ and θ' , and using the property that $\tan(-\theta) = -\tan \theta$, gives us

Equation:

$$\left. \begin{aligned} \tan \theta &= \frac{h_o}{d_o} \\ \tan \theta' &= -\tan \theta = \frac{h_i}{d_i} \end{aligned} \right\} \frac{h_o}{d_o} = -\frac{h_i}{d_i} \text{ or } -\frac{h_o}{h_i} = \frac{d_o}{d_i}.$$

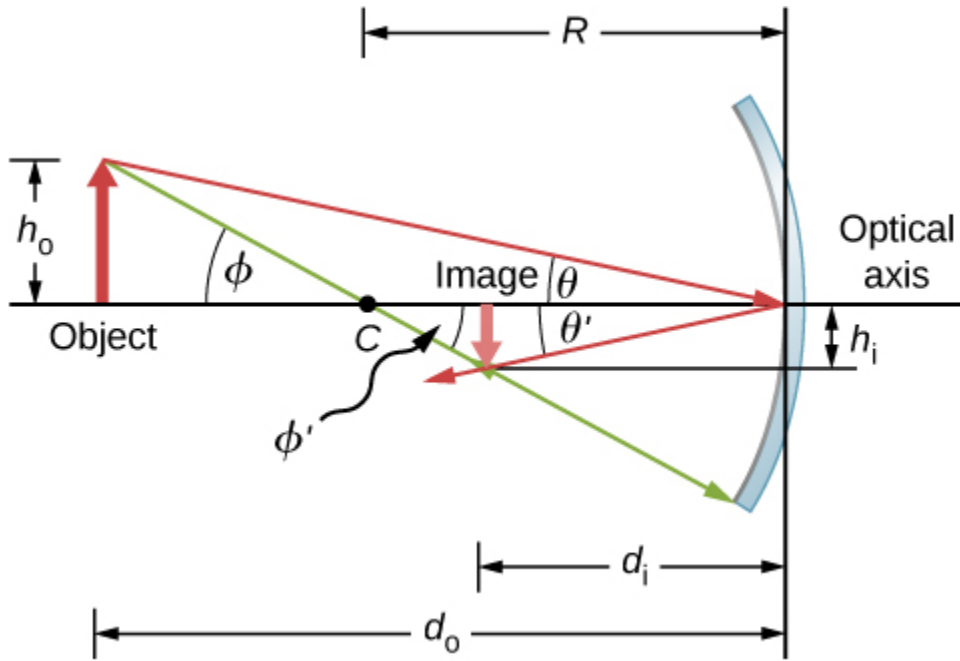


Image formed by a concave mirror.

Similarly, taking the tangent of ϕ and ϕ' gives

Equation:

$$\left. \begin{aligned} \tan \phi &= \frac{h_o}{d_o - R} \\ \tan \phi' &= -\tan \phi = \frac{h_i}{R - d_i} \end{aligned} \right\} \frac{h_o}{d_o - R} = -\frac{h_i}{R - d_i} \text{ or } -\frac{h_o}{h_i} = \frac{d_o - R}{R - d_i}.$$

Combining these two results gives

Equation:

$$\frac{d_o}{d_i} = \frac{d_o - R}{R - d_i}.$$

After a little algebra, this becomes

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{2}{R}.$$

No approximation is required for this result, so it is exact. However, as discussed above, in the small-angle approximation, the focal length of a spherical mirror is one-half the radius of curvature of the mirror, or $f = R/2$. Inserting this into [\[link\]](#) gives the *mirror equation*:

Note:

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}.$$

The mirror equation relates the image and object distances to the focal distance and is valid only in the small-angle approximation. Although it was derived for a concave mirror, it also holds for convex mirrors (proving this is left as an exercise). We can extend the mirror equation to the case of a plane mirror by noting that a plane mirror has an infinite radius of curvature. This means the focal point is at infinity, so the mirror equation simplifies to

Equation:

$$d_o = -d_i$$

which is the same as [\[link\]](#) obtained earlier.

Notice that we have been very careful with the signs in deriving the mirror equation. For a plane mirror, the image distance has the opposite sign of the object distance. Also, the real image formed by the concave mirror in [\[link\]](#) is on the opposite side of the optical axis with respect to the object. In this case, the image height should have the opposite sign of the object height. To keep track of the signs of the various quantities in the mirror equation, we now introduce a sign convention.

Sign convention for spherical mirrors

Using a consistent sign convention is very important in geometric optics. It assigns positive or negative values for the quantities that characterize an optical system. Understanding the sign convention allows you to describe an image without constructing a ray diagram. This text uses the following sign convention:

1. The focal length f is positive for concave mirrors and negative for convex mirrors.
2. The image distance d_i is positive for real images and negative for virtual images.

Notice that rule 1 means that the radius of curvature of a spherical mirror can be positive or negative. What does it mean to have a negative radius of curvature? This means simply that the radius of curvature for a convex mirror is defined to be negative.

Image magnification

Let's use the sign convention to further interpret the derivation of the mirror equation. In deriving this equation, we found that the object and image heights are related by

Equation:

$$-\frac{h_o}{h_i} = \frac{d_o}{d_i}.$$

See [\[link\]](#). Both the object and the image formed by the mirror in [\[link\]](#) are real, so the object and image distances are both positive. The highest point of the object is above the optical axis, so the object height is positive. The image, however, is below the optical axis, so the image height is negative. Thus, this sign convention is consistent with our derivation of the mirror equation.

[\[link\]](#) in fact describes the **linear magnification** (often simply called “**magnification**”) of the image in terms of the object and image distances. We thus define the dimensionless magnification m as follows:

Equation:

$$m = \frac{h_i}{h_o}.$$

If m is positive, the image is upright, and if m is negative, the image is inverted. If $|m| > 1$, the image is larger than the object, and if $|m| < 1$, the image is smaller than the object. With this definition of magnification, we get the following relation between the vertical and horizontal object and image distances:

Note:

Equation:

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}.$$

This is a very useful relation because it lets you obtain the magnification of the image from the object and image distances, which you can obtain from the mirror equation.

Example:

Solar Electric Generating System

One of the solar technologies used today for generating electricity involves a device (called a parabolic trough or concentrating collector) that concentrates sunlight onto a blackened pipe that contains a fluid. This heated fluid is pumped to a heat exchanger, where the thermal energy is transferred to another system that is used to generate steam and eventually generates electricity through a conventional steam cycle. [\[link\]](#) shows such a working system in southern California. The real mirror is a parabolic cylinder with its focus located at the pipe; however, we can approximate the mirror as exactly one-quarter of a circular cylinder.



Parabolic trough collectors are used to generate electricity in southern California. (credit: “kjkolb”/Wikimedia Commons)

- If we want the rays from the sun to focus at 40.0 cm from the mirror, what is the radius of the mirror?
- What is the amount of sunlight concentrated onto the pipe, per meter of pipe length, assuming the insolation (incident solar radiation) is 900 W/m^2 ?

- c. If the fluid-carrying pipe has a 2.00-cm diameter, what is the temperature increase of the fluid per meter of pipe over a period of 1 minute? Assume that all solar radiation incident on the reflector is absorbed by the pipe, and that the fluid is mineral oil.

Strategy

First identify the physical principles involved. Part (a) is related to the optics of spherical mirrors. Part (b) involves a little math, primarily geometry. Part (c) requires an understanding of heat and density.

Solution

- a. The sun is the object, so the object distance is essentially infinity: $d_o = \infty$. The desired image distance is $d_i = 40.0$ cm. We use the mirror equation to find the focal length of the mirror:

Equation:

$$\begin{aligned}\frac{1}{d_o} + \frac{1}{d_i} &= \frac{1}{f} \\ f &= \left(\frac{1}{d_o} + \frac{1}{d_i} \right)^{-1} \\ &= \left(\frac{1}{\infty} + \frac{1}{40.0 \text{ cm}} \right)^{-1} \\ &= 40.0 \text{ cm}\end{aligned}$$

Thus, the radius of the mirror is $R = 2f = 80.0$ cm.

- b. The insolation is 900 W/m^2 . You must find the cross-sectional area A of the concave mirror, since the power delivered is $900 \text{ W/m}^2 \times A$. The mirror in this case is estimated as a quarter-section of a cylinder, so the area for a length L of the mirror is $A = \frac{1}{4}(2\pi R)L$. The area for a length of 1.00 m is then

Equation:

$$A = \frac{\pi}{2} R(1.00 \text{ m}) = \frac{(3.14)}{2} (0.800 \text{ m})(1.00 \text{ m}) = 1.26 \text{ m}^2.$$

The insolation on the 1.00-m length of pipe is then

Equation:

$$\left(9.00 \times 10^2 \frac{\text{W}}{\text{m}^2}\right) (1.26 \text{ m}^2) = 1130 \text{ W}.$$

c. The increase in temperature is given by $Q = mc\Delta T$. The mass m of the mineral oil in the one-meter section of pipe is

Equation:

$$\begin{aligned} m &= \rho V = \rho \pi \left(\frac{d}{2}\right)^2 (1.00 \text{ m}) \\ &= \left(8.00 \times 10^2 \text{ kg/m}^3\right) (3.14) (0.0100 \text{ m})^2 (1.00 \text{ m}) \\ &= 0.251 \text{ kg} \end{aligned}$$

Therefore, the increase in temperature in one minute is

Equation:

$$\begin{aligned} \Delta T &= Q/mc \\ &= \frac{(1130 \text{ W})(60.0 \text{ s})}{(0.251 \text{ kg})(1670 \text{ J}\cdot\text{kg}/^\circ\text{C})} \\ &= 162^\circ\text{C} \end{aligned}$$

Significance

An array of such pipes in the California desert can provide a thermal output of 250 MW on a sunny day, with fluids reaching temperatures as high as 400°C . We are considering only one meter of pipe here and ignoring heat losses along the pipe.

Example:

Image in a Convex Mirror

A keratometer is a device used to measure the curvature of the cornea of the eye, particularly for fitting contact lenses. Light is reflected from the cornea, which acts like a convex mirror, and the keratometer measures the magnification of the image. The smaller the magnification, the smaller the radius of curvature of the cornea. If the light source is 12 cm from the

cornea and the image magnification is 0.032, what is the radius of curvature of the cornea?

Strategy

If you find the focal length of the convex mirror formed by the cornea, then you know its radius of curvature (it's twice the focal length). The object distance is $d_o = 12$ cm and the magnification is $m = 0.032$. First find the image distance d_i and then solve for the focal length f .

Solution

Start with the equation for magnification, $m = -d_i/d_o$. Solving for d_i and inserting the given values yields

Equation:

$$d_i = -md_o = -(0.032)(12 \text{ cm}) = -0.384 \text{ cm}$$

where we retained an extra significant figure because this is an intermediate step in the calculation. Solve the mirror equation for the focal length f and insert the known values for the object and image distances.

The result is

Equation:

$$\begin{aligned}\frac{1}{d_o} + \frac{1}{d_i} &= \frac{1}{f} \\ f &= \left(\frac{1}{d_o} + \frac{1}{d_i} \right)^{-1} \\ &= \left(\frac{1}{12 \text{ cm}} + \frac{1}{-0.384 \text{ cm}} \right)^{-1} \\ &= -0.40 \text{ cm}\end{aligned}$$

The radius of curvature is twice the focal length, so

Equation:

$$R = 2f = -0.80 \text{ cm}$$

Significance

The focal length is negative, so the focus is virtual, as expected for a concave mirror and a real object. The radius of curvature found here is reasonable for a cornea. The distance from cornea to retina in an adult eye is about 2.0 cm. In practice, corneas may not be spherical, which complicates the job of fitting contact lenses. Note that the image distance

here is negative, consistent with the fact that the image is behind the mirror. Thus, the image is virtual because no rays actually pass through it. In the problems and exercises, you will show that, for a fixed object distance, a smaller radius of curvature corresponds to a smaller the magnification.

Note:

Spherical Mirrors

Step 1. First make sure that image formation by a spherical mirror is involved.

Step 2. Determine whether ray tracing, the mirror equation, or both are required. A sketch is very useful even if ray tracing is not specifically required by the problem. Write symbols and known values on the sketch.

Step 3. Identify exactly what needs to be determined in the problem (identify the unknowns).

Step 4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns).

Step 5. If ray tracing is required, use the ray-tracing rules listed near the beginning of this section.

Step 6. Most quantitative problems require using the mirror equation. Use the examples as guides for using the mirror equation.

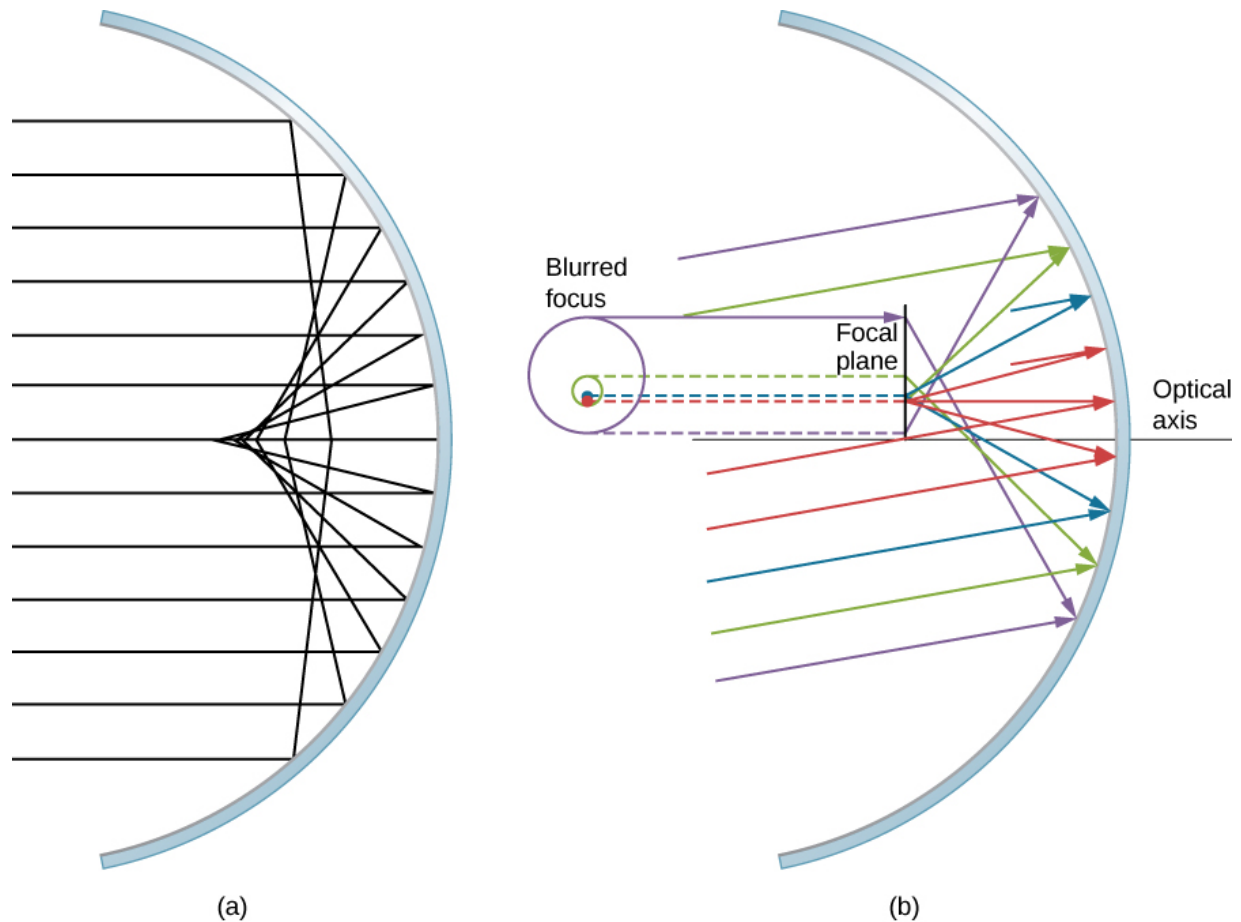
Step 7. Check to see whether the answer makes sense. Do the signs of object distance, image distance, and focal length correspond with what is expected from ray tracing? Is the sign of the magnification correct? Are the object and image distances reasonable?

Departure from the Small-Angle Approximation

The small-angle approximation is a cornerstone of the above discussion of image formation by a spherical mirror. When this approximation is violated, then the image created by a spherical mirror becomes distorted. Such distortion is called **aberration**. Here we briefly discuss two specific types of aberrations: spherical aberration and coma.

Spherical aberration

Consider a broad beam of parallel rays impinging on a spherical mirror, as shown in [\[link\]](#).



(a) With spherical aberration, the rays that are farther from the optical axis and the rays that are closer to the optical axis are focused at different points. Notice that the aberration gets worse for rays farther from the optical axis. (b) For comatic aberration, parallel rays that are not parallel to the optical axis are focused at different heights and at different focal lengths, so the image contains a “tail” like a comet (which is “coma” in Latin). Note that the colored rays are only to facilitate viewing; the colors do not indicate the color of the light.

The farther from the optical axis the rays strike, the worse the spherical mirror approximates a parabolic mirror. Thus, these rays are not focused at the same point as rays that are near the optical axis, as shown in the figure. Because of **spherical aberration**, the image of an extended object in a spherical mirror will be blurred. Spherical aberrations are characteristic of the mirrors and lenses that we consider in the following section of this chapter (more sophisticated mirrors and lenses are needed to eliminate spherical aberrations).

Coma or comatic aberration

Coma is similar to spherical aberration, but arises when the incoming rays are not parallel to the optical axis, as shown in part (b) of [\[link\]](#). Recall that the small-angle approximation holds for spherical mirrors that are small compared to their radius. In this case, spherical mirrors are good approximations of parabolic mirrors. Parabolic mirrors focus all rays that are parallel to the optical axis at the focal point. However, parallel rays that are *not* parallel to the optical axis are focused at different heights and at different focal lengths, as shown in part (b) of [\[link\]](#). Because a spherical mirror is symmetric about the optical axis, the various colored rays in this figure create circles of the corresponding color on the focal plane.

Although a spherical mirror is shown in part (b) of [\[link\]](#), comatic aberration occurs also for parabolic mirrors—it does not result from a breakdown in the small-angle approximation. Spherical aberration, however, occurs only for spherical mirrors and is a result of a breakdown in the small-angle approximation. We will discuss both coma and spherical aberration later in this chapter, in connection with telescopes.

Summary

- Spherical mirrors may be concave (converging) or convex (diverging).
- The focal length of a spherical mirror is one-half of its radius of curvature: $f = R/2$.

- The mirror equation and ray tracing allow you to give a complete description of an image formed by a spherical mirror.
- Spherical aberration occurs for spherical mirrors but not parabolic mirrors; comatic aberration occurs for both types of mirrors.

Conceptual Questions

Exercise:

Problem: At what distance is an image always located: at d_o , d_i , or f ?

Exercise:

Problem:

Under what circumstances will an image be located at the focal point of a spherical lens or mirror?

Solution:

when the object is at infinity; see the mirror equation

Exercise:

Problem:

What is meant by a negative magnification? What is meant by a magnification whose absolute value is less than one?

Exercise:

Problem:

Can an image be larger than the object even though its magnification is negative? Explain.

Solution:

Yes, negative magnification simply means that the image is upside down; this does not prevent the image from being larger than the

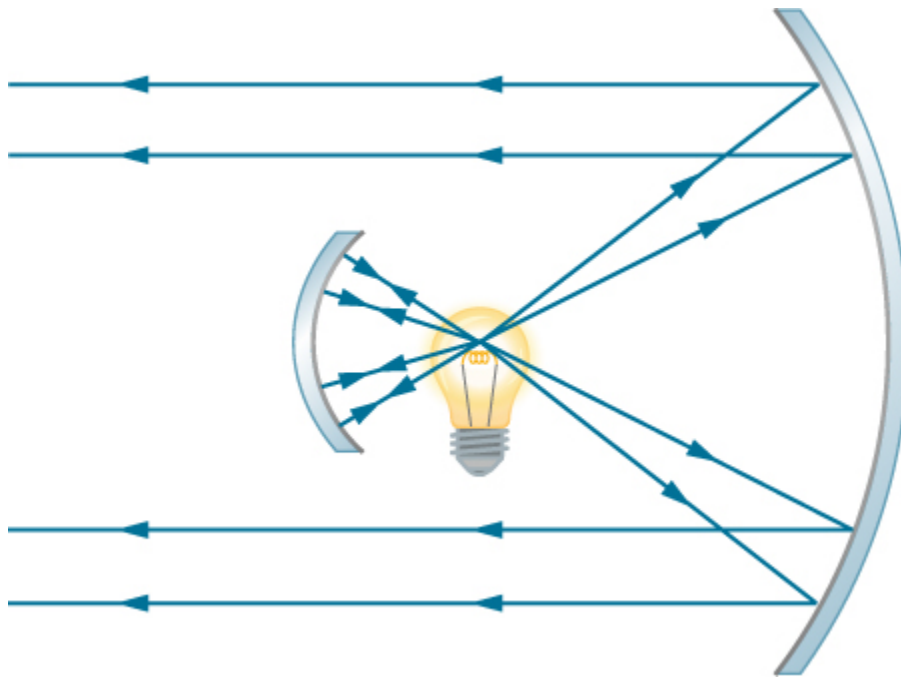
object. For instance, for a concave mirror, if distance to the object is larger than one focal distance but smaller than two focal distances the image will be inverted and magnified.

Problems

Exercise:

Problem:

The following figure shows a light bulb between two spherical mirrors. One mirror produces a beam of light with parallel rays; the other keeps light from escaping without being put into the beam. Where is the filament of the light in relation to the focal point or radius of curvature of each mirror?



Solution:

It is in the focal point of the big mirror and at the center of curvature of the small mirror.

Exercise:**Problem:**

Why are diverging mirrors often used for rearview mirrors in vehicles? What is the main disadvantage of using such a mirror compared with a flat one?

Exercise:**Problem:**

Some telephoto cameras use a mirror rather than a lens. What radius of curvature mirror is needed to replace a 800 mm-focal length telephoto lens?

Solution:

$$f = \frac{R}{2} \Rightarrow R = +1.60 \text{ m}$$

Exercise:**Problem:**

Calculate the focal length of a mirror formed by the shiny back of a spoon that has a 3.00 cm radius of curvature.

Exercise:**Problem:**

Electric room heaters use a concave mirror to reflect infrared (IR) radiation from hot coils. Note that IR radiation follows the same law of reflection as visible light. Given that the mirror has a radius of curvature of 50.0 cm and produces an image of the coils 3.00 m away from the mirror, where are the coils?

Solution:

$$d_o = 27.3 \text{ cm}$$

Exercise:

Problem:

Find the magnification of the heater element in the previous problem. Note that its large magnitude helps spread out the reflected energy.

Exercise:**Problem:**

What is the focal length of a makeup mirror that produces a magnification of 1.50 when a person's face is 12.0 cm away? Explicitly show how you follow the steps in the [\[link\]](#).

Solution:

Step 1: Image formation by a mirror is involved.

Step 2: Draw the problem set up when possible.

Step 3: Use thin-lens equations to solve this problem.

Step 4: Find f .

Step 5: Given: $m = 1.50$, $d_o = 0.120$ m.

Step 6: No ray tracing is needed.

Step 7: Using $m = \frac{d_i}{d_o}$, $d_i = -0.180$ m. Then, $f = 0.360$ m.

Step 8: The image is virtual because the image distance is negative. The focal length is positive, so the mirror is concave.

Exercise:**Problem:**

A shopper standing 3.00 m from a convex security mirror sees his image with a magnification of 0.250. (a) Where is his image? (b) What is the focal length of the mirror? (c) What is its radius of curvature?

Exercise:

Problem:

An object 1.50 cm high is held 3.00 cm from a person's cornea, and its reflected image is measured to be 0.167 cm high. (a) What is the magnification? (b) Where is the image? (c) Find the radius of curvature of the convex mirror formed by the cornea. (Note that this technique is used by optometrists to measure the curvature of the cornea for contact lens fitting. The instrument used is called a keratometer, or curve measurer.)

Solution:

- a. for a convex mirror $d_i < 0 \Rightarrow m > 0$. $m = +0.111$; b.
 $d_i = -0.334$ cm (behind the cornea);
c. $f = -0.376$ cm, so that $R = -0.752$ cm

Exercise:**Problem:**

Ray tracing for a flat mirror shows that the image is located a distance behind the mirror equal to the distance of the object from the mirror. This is stated as $d_i = -d_o$, since this is a negative image distance (it is a virtual image). What is the focal length of a flat mirror?

Exercise:**Problem:**

Show that, for a flat mirror, $h_i = h_o$, given that the image is the same distance behind the mirror as the distance of the object from the mirror.

Solution:

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o} = -\frac{-d_o}{d_o} = \frac{d_o}{d_o} = 1 \Rightarrow h_i = h_o$$

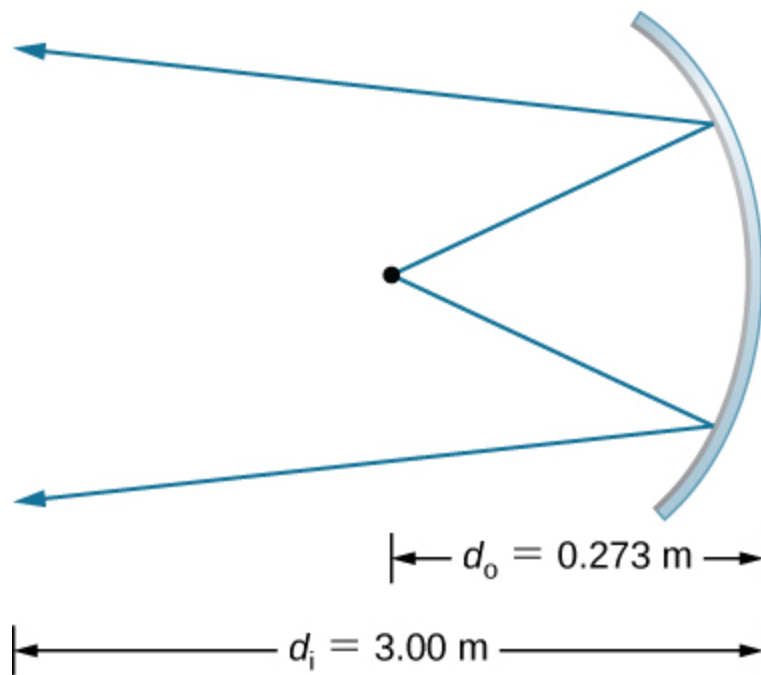
Exercise:

Problem:

Use the law of reflection to prove that the focal length of a mirror is half its radius of curvature. That is, prove that $f = R/2$. Note this is true for a spherical mirror only if its diameter is small compared with its radius of curvature.

Exercise:**Problem:**

Referring to the electric room heater considered in problem 5, calculate the intensity of IR radiation in W/m^2 projected by the concave mirror on a person 3.00 m away. Assume that the heating element radiates 1500 W and has an area of 100 cm^2 , and that half of the radiated power is reflected and focused by the mirror.

Solution:

$$m = -11.0$$

$$A' = 0.110 \text{ m}^2$$

$$I = 6.82 \text{ kW/m}^2$$

Exercise:

Problem:

Two mirrors are inclined at an angle of 60° and an object is placed at a point that is equidistant from the two mirrors. Use a protractor to draw rays accurately and locate all images. You may have to draw several figures so that that rays for different images do not clutter your drawing.

Exercise:

Problem:

Two parallel mirrors are facing each other and are separated by a distance of 3 cm. A point object is placed between the mirrors 1 cm from one of the mirrors. Find the coordinates of all the images.

Solution:

$$x_{2m} = -x_{2m-1}, \quad (m = 1, 2, 3, \dots),$$

$$x_{2m+1} = b - x_{2m}, \quad (m = 0, 1, 2, \dots), \text{ with } x_0 = a.$$

Glossary

aberration

distortion in an image caused by departures from the small-angle approximation

coma

similar to spherical aberration, but arises when the incoming rays are not parallel to the optical axis

concave mirror

spherical mirror with its reflecting surface on the inner side of the sphere; the mirror forms a “cave”

convex mirror

spherical mirror with its reflecting surface on the outer side of the sphere

curved mirror

mirror formed by a curved surface, such as spherical, elliptical, or parabolic

focal length

distance along the optical axis from the focal point to the optical element that focuses the light rays

focal point

for a converging lens or mirror, the point at which converging light rays cross; for a diverging lens or mirror, the point from which diverging light rays appear to originate

linear magnification

ratio of image height to object height

magnification

ratio of image size to object size

optical axis

axis about which the mirror is rotationally symmetric; you can rotate the mirror about this axis without changing anything

small-angle approximation

approximation that is valid when the size of a spherical mirror is significantly smaller than the mirror's radius; in this approximation, spherical aberration is negligible and the mirror has a well-defined focal point

spherical aberration

distortion in the image formed by a spherical mirror when rays are not all focused at the same point

vertex

point where the mirror's surface intersects with the optical axis

Images Formed by Refraction

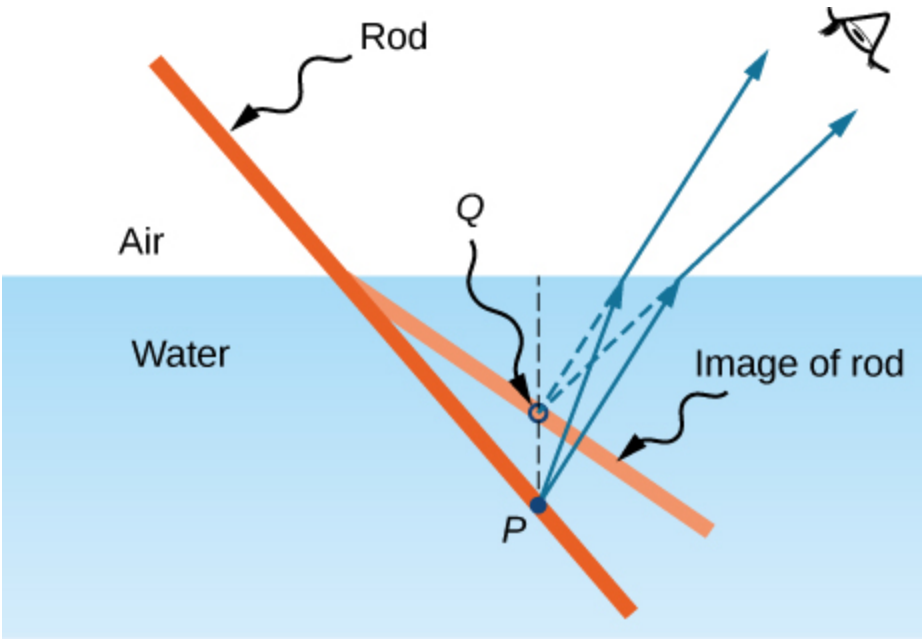
By the end of this section, you will be able to:

- Describe image formation by a single refracting surface
- Determine the location of an image and calculate its properties by using a ray diagram
- Determine the location of an image and calculate its properties by using the equation for a single refracting surface

When rays of light propagate from one medium to another, these rays undergo refraction, which is when light waves are bent at the interface between two media. The refracting surface can form an image in a similar fashion to a reflecting surface, except that the law of refraction (Snell's law) is at the heart of the process instead of the law of reflection.

Refraction at a Plane Interface—Apparent Depth

If you look at a straight rod partially submerged in water, it appears to bend at the surface ([link](#)). The reason behind this curious effect is that the image of the rod inside the water forms a little closer to the surface than the actual position of the rod, so it does not line up with the part of the rod that is above the water. The same phenomenon explains why a fish in water appears to be closer to the surface than it actually is.



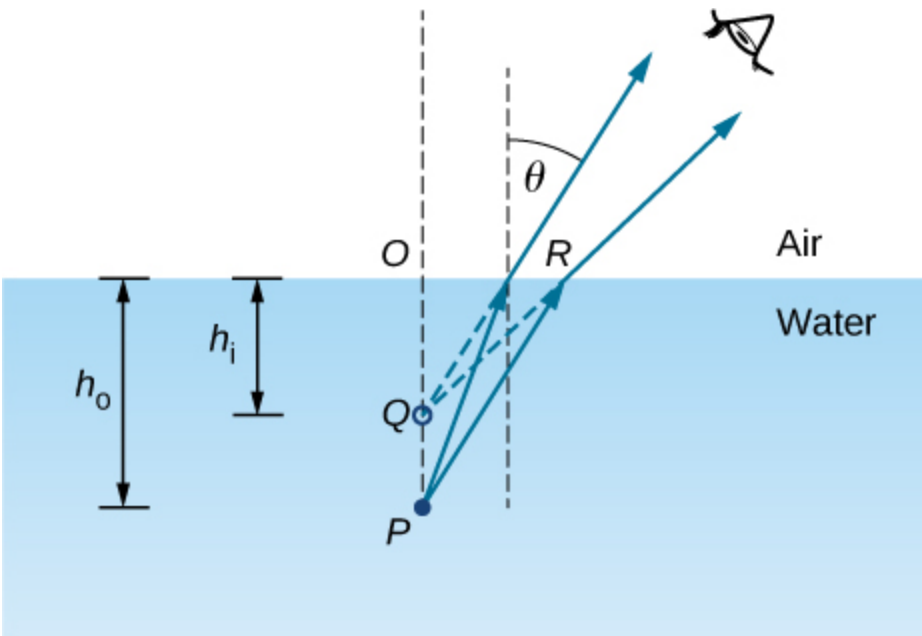
Bending of a rod at a water-air interface. Point P on the rod appears to be at point Q , which is where the image of point P forms due to refraction at the air-water interface.

To study image formation as a result of refraction, consider the following questions:

1. What happens to the rays of light when they enter or pass through a different medium?
2. Do the refracted rays originating from a single point meet at some point or diverge away from each other?

To be concrete, we consider a simple system consisting of two media separated by a plane interface ([\[link\]](#)). The object is in one medium and the observer is in the other. For instance, when you look at a fish from above the water surface, the fish is in medium 1 (the water) with refractive index 1.33, and your eye is in medium 2 (the air) with refractive index 1.00, and the surface of the water is the interface. The depth that you “see” is the

image height h_i and is called the **apparent depth**. The actual depth of the fish is the object height h_o .



Apparent depth due to refraction. The real object at point P creates an image at point Q . The image is not at the same depth as the object, so the observer sees the image at an “apparent depth.”

The apparent depth h_i depends on the angle at which you view the image. For a view from above (the so-called “normal” view), we can approximate the refraction angle θ to be small, and replace $\sin \theta$ in Snell’s law by $\tan \theta$. With this approximation, you can use the triangles $\triangle OPR$ and $\triangle OQR$ to show that the apparent depth is given by

Note:
Equation:

$$h_i = \left(\frac{n_2}{n_1} \right) h_o.$$

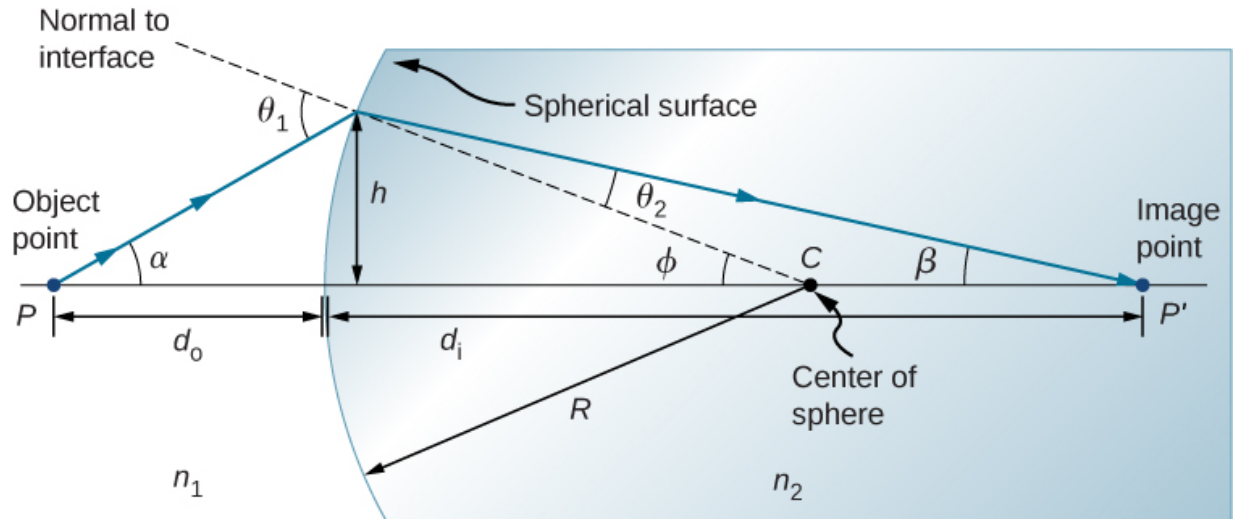
The derivation of this result is left as an exercise. Thus, a fish appears at 3/4 of the real depth when viewed from above.

Refraction at a Spherical Interface

Spherical shapes play an important role in optics primarily because high-quality spherical shapes are far easier to manufacture than other curved surfaces. To study refraction at a single spherical surface, we assume that the medium with the spherical surface at one end continues indefinitely (a “semi-infinite” medium).

Refraction at a convex surface

Consider a point source of light at point P in front of a convex surface made of glass (see [\[link\]](#)). Let R be the radius of curvature, n_1 be the refractive index of the medium in which object point P is located, and n_2 be the refractive index of the medium with the spherical surface. We want to know what happens as a result of refraction at this interface.



Refraction at a convex surface ($n_2 > n_1$).

Because of the symmetry involved, it is sufficient to examine rays in only one plane. The figure shows a ray of light that starts at the object point P , refracts at the interface, and goes through the image point P' . We derive a formula relating the object distance d_o , the image distance d_i , and the radius of curvature R .

Applying Snell's law to the ray emanating from point P gives $n_1 \sin \theta_1 = n_2 \sin \theta_2$. We work in the small-angle approximation, so $\sin \theta \approx \theta$ and Snell's law then takes the form

Equation:

$$n_1 \theta_1 \approx n_2 \theta_2.$$

From the geometry of the figure, we see that

Equation:

$$\theta_1 = \alpha + \phi, \quad \theta_2 = \phi - \beta.$$

Inserting these expressions into Snell's law gives

Equation:

$$n_1(\alpha + \phi) \approx n_2(\phi - \beta).$$

Using the diagram, we calculate the tangent of the angles α , β , and ϕ :

Equation:

$$\tan \alpha \approx \frac{h}{d_o}, \quad \tan \beta \approx \frac{h}{d_i}, \quad \tan \phi \approx \frac{h}{R}.$$

Again using the small-angle approximation, we find that $\tan \theta \approx \theta$, so the above relationships become

Equation:

$$\alpha \approx \frac{h}{d_o}, \quad \beta \approx \frac{h}{d_i}, \quad \phi \approx \frac{h}{R}.$$

Putting these angles into Snell's law gives

Equation:

$$n_1 \left(\frac{h}{d_o} + \frac{h}{R} \right) = n_2 \left(\frac{h}{R} - \frac{h}{d_i} \right).$$

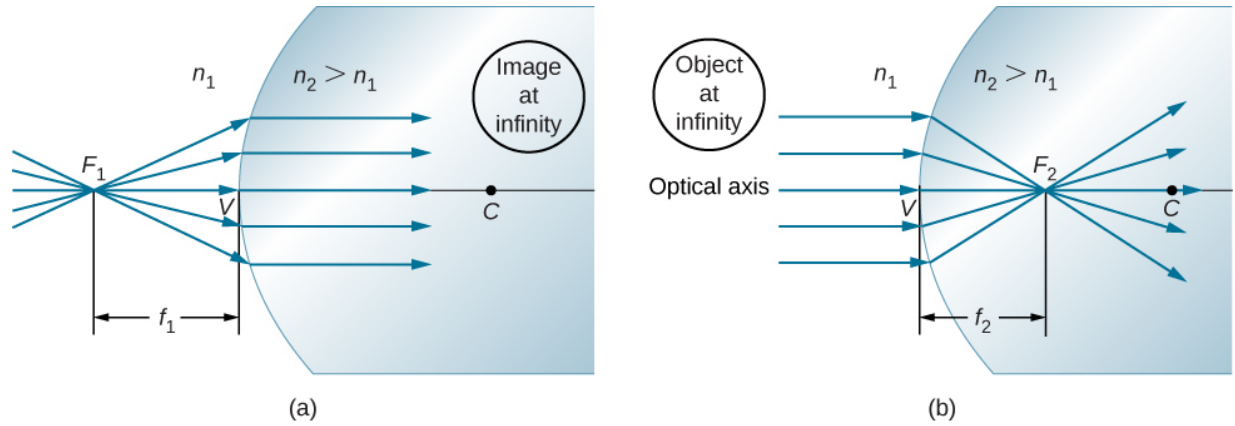
We can write this more conveniently as

Note:

Equation:

$$\frac{n_1}{d_o} + \frac{n_2}{d_i} = \frac{n_2 - n_1}{R}.$$

If the object is placed at a special point called the **first focus**, or the **object focus** F_1 , then the image is formed at infinity, as shown in part (a) of [\[link\]](#).



(a) First focus (called the “object focus”) for refraction at a convex surface. (b) Second focus (called “image focus”) for refraction at a convex surface.

We can find the location f_1 of the first focus F_1 by setting $d_i = \infty$ in the preceding equation.

Equation:

$$\frac{n_1}{f_1} + \frac{n_2}{\infty} = \frac{n_2 - n_1}{R}$$

Equation:

$$f_1 = \frac{n_1 R}{n_2 - n_1}$$

Similarly, we can define a **second focus** or **image focus** F_2 where the image is formed for an object that is far away [part (b)]. The location of the second focus F_2 is obtained from [\[link\]](#) by setting $d_o = \infty$:

Equation:

$$\frac{n_1}{\infty} + \frac{n_2}{f_2} = \frac{n_2 - n_1}{R}$$

Equation:

$$f_2 = \frac{n_2 R}{n_2 - n_1}.$$

Note that the object focus is at a different distance from the vertex than the image focus because $n_1 \neq n_2$.

Sign convention for single refracting surfaces

Although we derived this equation for refraction at a convex surface, the same expression holds for a concave surface, provided we use the following sign convention:

1. $R > 0$ if surface is convex toward object; otherwise, $R < 0$.
2. $d_i > 0$ if image is real and on opposite side from the object; otherwise, $d_i < 0$.

Summary

This section explains how a single refracting interface forms images.

- When an object is observed through a plane interface between two media, then it appears at an apparent distance h_i that differs from the actual distance h_o : $h_i = (n_2/n_1)h_o$.
- An image is formed by the refraction of light at a spherical interface between two media of indices of refraction n_1 and n_2 .
- Image distance depends on the radius of curvature of the interface, location of the object, and the indices of refraction of the media.

Conceptual Questions

Exercise:**Problem:**

Derive the formula for the apparent depth of a fish in a fish tank using Snell's law.

Exercise:**Problem:**

Use a ruler and a protractor to find the image by refraction in the following cases. Assume an air-glass interface. Use a refractive index of 1 for air and of 1.5 for glass. (*Hint: Use Snell's law at the interface.*)

- (a) A point object located on the axis of a concave interface located at a point within the focal length from the vertex.
- (b) A point object located on the axis of a concave interface located at a point farther than the focal length from the vertex.
- (c) A point object located on the axis of a convex interface located at a point within the focal length from the vertex.
- (d) A point object located on the axis of a convex interface located at a point farther than the focal length from the vertex.
- (e) Repeat (a)–(d) for a point object off the axis.

Solution:

answers may vary

Problems**Exercise:**

Problem:

An object is located in air 30 cm from the vertex of a concave surface made of glass with a radius of curvature 10 cm. Where does the image by refraction form and what is its magnification? Use $n_{\text{air}} = 1$ and $n_{\text{glass}} = 1.5$.

Exercise:**Problem:**

An object is located in air 30 cm from the vertex of a convex surface made of glass with a radius of curvature 80 cm. Where does the image by refraction form and what is its magnification?

Solution:

$$d_i = -55 \text{ cm}; m = +1.8$$

Exercise:**Problem:**

An object is located in water 15 cm from the vertex of a concave surface made of glass with a radius of curvature 10 cm. Where does the image by refraction form and what is its magnification? Use $n_{\text{water}} = 4/3$ and $n_{\text{glass}} = 1.5$.

Exercise:**Problem:**

An object is located in water 30 cm from the vertex of a convex surface made of Plexiglas with a radius of curvature of 80 cm. Where does the image form by refraction and what is its magnification? $n_{\text{water}} = 4/3$ and $n_{\text{Plexiglas}} = 1.65$.

Solution:

$$d_i = -41 \text{ cm}, m = 1.4$$

Exercise:**Problem:**

An object is located in air 5 cm from the vertex of a concave surface made of glass with a radius of curvature 20 cm. Where does the image form by refraction and what is its magnification? Use $n_{\text{air}} = 1$ and $n_{\text{glass}} = 1.5$.

Exercise:**Problem:**

Derive the spherical interface equation for refraction at a concave surface. (*Hint: Follow the derivation in the text for the convex surface.*)

Solution:

proof

Glossary

apparent depth

depth at which an object is perceived to be located with respect to an interface between two media

first focus or object focus

object located at this point will result in an image created at infinity on the opposite side of a spherical interface between two media

second focus or image focus

for a converging interface, the point where a bundle of parallel rays refracting at a spherical interface; for a diverging interface, the point at which the backward continuation of the refracted rays will converge between two media will focus

The Eye

By the end of this section, you will be able to:

- Understand the basic physics of how images are formed by the human eye
- Recognize several conditions of impaired vision as well as the optics principles for treating these conditions

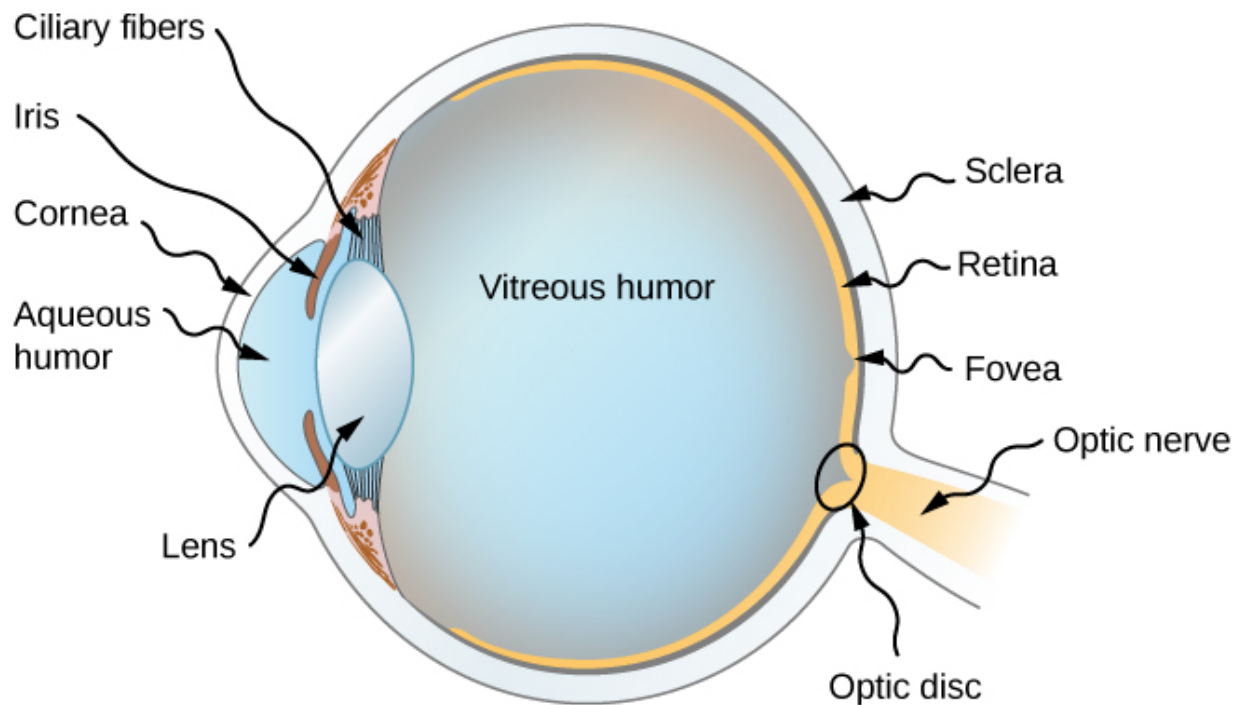
The human eye is perhaps the most interesting and important of all optical instruments. Our eyes perform a vast number of functions: They allow us to sense direction, movement, colors, and distance. In this section, we explore the geometric optics of the eye.

Physics of the Eye

The eye is remarkable in how it forms images and in the richness of detail and color it can detect. However, our eyes often need some correction to reach what is called “normal” vision. Actually, normal vision should be called “ideal” vision because nearly one-half of the human population requires some sort of eyesight correction, so requiring glasses is by no means “abnormal.” Image formation by our eyes and common vision correction can be analyzed with the optics discussed earlier in this chapter.

[\[link\]](#) shows the basic anatomy of the eye. The cornea and lens form a system that, to a good approximation, acts as a single thin lens. For clear vision, a real image must be projected onto the light-sensitive retina, which lies a fixed distance from the lens. The flexible lens of the eye allows it to adjust the radius of curvature of the lens to produce an image on the retina for objects at different distances. The center of the image falls on the fovea, which has the greatest density of light receptors and the greatest acuity (sharpness) in the visual field. The variable opening (i.e., the pupil) of the eye, along with chemical adaptation, allows the eye to detect light intensities from the lowest observable to 10^{10} times greater (without damage). This is an incredible range of detection. Processing of visual nerve impulses begins with interconnections in the retina and continues in

the brain. The optic nerve conveys the signals received by the eye to the brain.



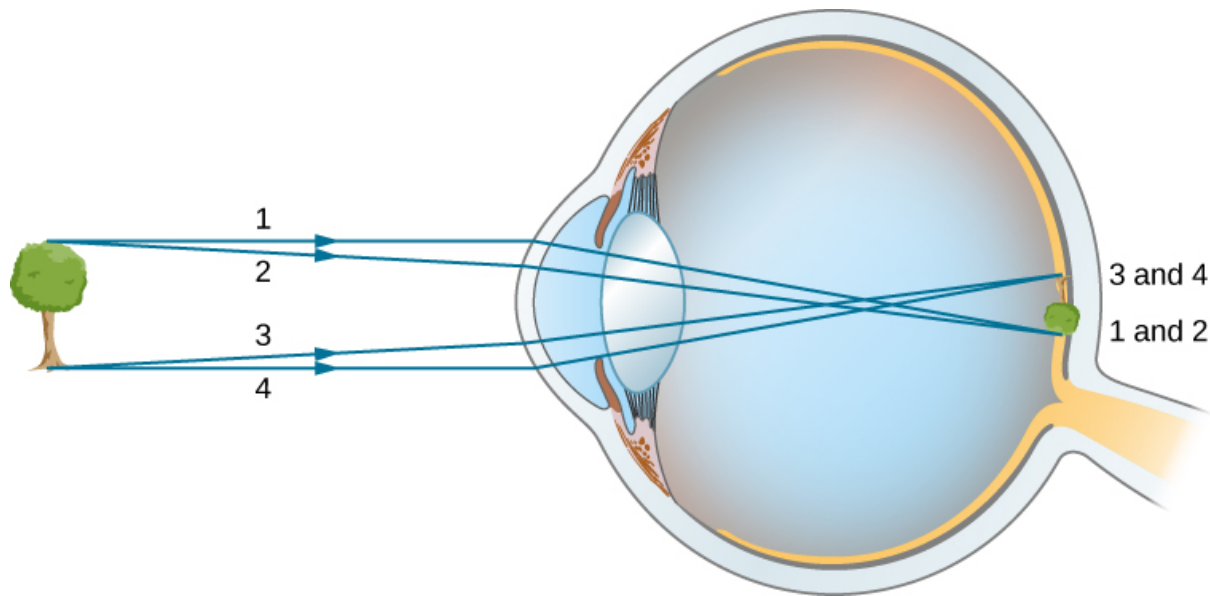
The cornea and lens of the eye act together to form a real image on the light-sensing retina, which has its densest concentration of receptors in the fovea and a blind spot over the optic nerve. The radius of curvature of the lens of an eye is adjustable to form an image on the retina for different object distances. Layers of tissues with varying indices of refraction in the lens are shown here. However, they have been omitted from other pictures for clarity.

The indices of refraction in the eye are crucial to its ability to form images. [\[link\]](#) lists the indices of refraction relevant to the eye. The biggest change in the index of refraction, which is where the light rays are most bent, occurs at the air-cornea interface rather than at the aqueous humor-lens interface. The ray diagram in [\[link\]](#) shows image formation by the cornea and lens of the eye. The cornea, which is itself a converging lens with a

focal length of approximately 2.3 cm, provides most of the focusing power of the eye. The lens, which is a converging lens with a focal length of about 6.4 cm, provides the finer focus needed to produce a clear image on the retina. The cornea and lens can be treated as a single thin lens, even though the light rays pass through several layers of material (such as cornea, aqueous humor, several layers in the lens, and vitreous humor), changing direction at each interface. The image formed is much like the one produced by a single convex lens (i.e., a real, inverted image). Although images formed in the eye are inverted, the brain inverts them once more to make them seem upright.

Material	Index of Refraction
Water	1.33
Air	1.0
Cornea	1.38
Aqueous humor	1.34
Lens	1.41*
Vitreous humor	1.34

Refractive Indices Relevant to the Eye*This is an average value. The actual index of refraction varies throughout the lens and is greatest in center of the lens.



In the human eye, an image forms on the retina. Rays from the top and bottom of the object are traced to show how a real, inverted image is produced on the retina. The distance to the object is not to scale.

As noted, the image must fall precisely on the retina to produce clear vision—that is, the image distance d_i must equal the lens-to-retina distance. Because the lens-to-retina distance does not change, the image distance d_i must be the same for objects at all distances. The ciliary muscles adjust the shape of the eye lens for focusing on nearby or far objects. By changing the shape of the eye lens, the eye changes the focal length of the lens. This mechanism of the eye is called **accommodation**.

The nearest point an object can be placed so that the eye can form a clear image on the retina is called the **near point** of the eye. Similarly, the **far point** is the farthest distance at which an object is clearly visible. A person with normal vision can see objects clearly at distances ranging from 25 cm to essentially infinity. The near point increases with age, becoming several meters for some older people. In this text, we consider the near point to be 25 cm.

We can use the thin-lens equations to quantitatively examine image formation by the eye. First, we define the **optical power** of a lens as

Note:

Equation:

$$P = \frac{1}{f}$$

with the focal length f given in meters. The units of optical power are called “diopters” (D). That is, $1 \text{ D} = \frac{1}{\text{m}}$, or 1 m^{-1} . Optometrists prescribe common eyeglasses and contact lenses in units of diopters. With this definition of optical power, we can rewrite the thin-lens equations as

Equation:

$$P = \frac{1}{d_o} + \frac{1}{d_i}.$$

Working with optical power is convenient because, for two or more lenses close together, the effective optical power of the lens system is approximately the sum of the optical power of the individual lenses:

Note:

Equation:

$$P_{\text{total}} = P_{\text{lens 1}} + P_{\text{lens 2}} + P_{\text{lens 3}} + \cdots$$

Example:

Effective Focal Length of the Eye

The cornea and eye lens have focal lengths of 2.3 and 6.4 cm, respectively. Find the net focal length and optical power of the eye.

Strategy

The optical powers of the closely spaced lenses add, so

$$P_{\text{eye}} = P_{\text{cornea}} + P_{\text{lens}}.$$

Solution

Writing the equation for power in terms of the focal lengths gives

Equation:

$$\frac{1}{f_{\text{eye}}} = \frac{1}{f_{\text{cornea}}} + \frac{1}{f_{\text{lens}}} = \frac{1}{2.3 \text{ cm}} + \frac{1}{6.4 \text{ cm}}.$$

Hence, the focal length of the eye (cornea and lens together) is

Equation:

$$f_{\text{eye}} = 1.69 \text{ cm}.$$

The optical power of the eye is

Equation:

$$P_{\text{eye}} = \frac{1}{f_{\text{eye}}} = \frac{1}{0.0169 \text{ m}} = 59 \text{ D}.$$

For clear vision, the image distance d_i must equal the lens-to-retina distance. Normal vision is possible for objects at distances $d_o = 25 \text{ cm}$ to infinity. The following example shows how to calculate the image distance for an object placed at the near point of the eye.

Example:

Image of an object placed at the near point

The net focal length of a particular human eye is 1.7 cm. An object is placed at the near point of the eye. How far behind the lens is a focused

image formed?

Strategy

The near point is 25 cm from the eye, so the object distance is $d_o = 25$ cm. We determine the image distance from the lens equation:

Equation:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o}.$$

Solution

Equation:

$$\begin{aligned} d_i &= \left(\frac{1}{f} - \frac{1}{d_o} \right)^{-1} \\ &= \left(\frac{1}{1.7 \text{ cm}} - \frac{1}{25 \text{ cm}} \right)^{-1} \\ &= 1.8 \text{ cm} \end{aligned}$$

Therefore, the image is formed 1.8 cm behind the lens.

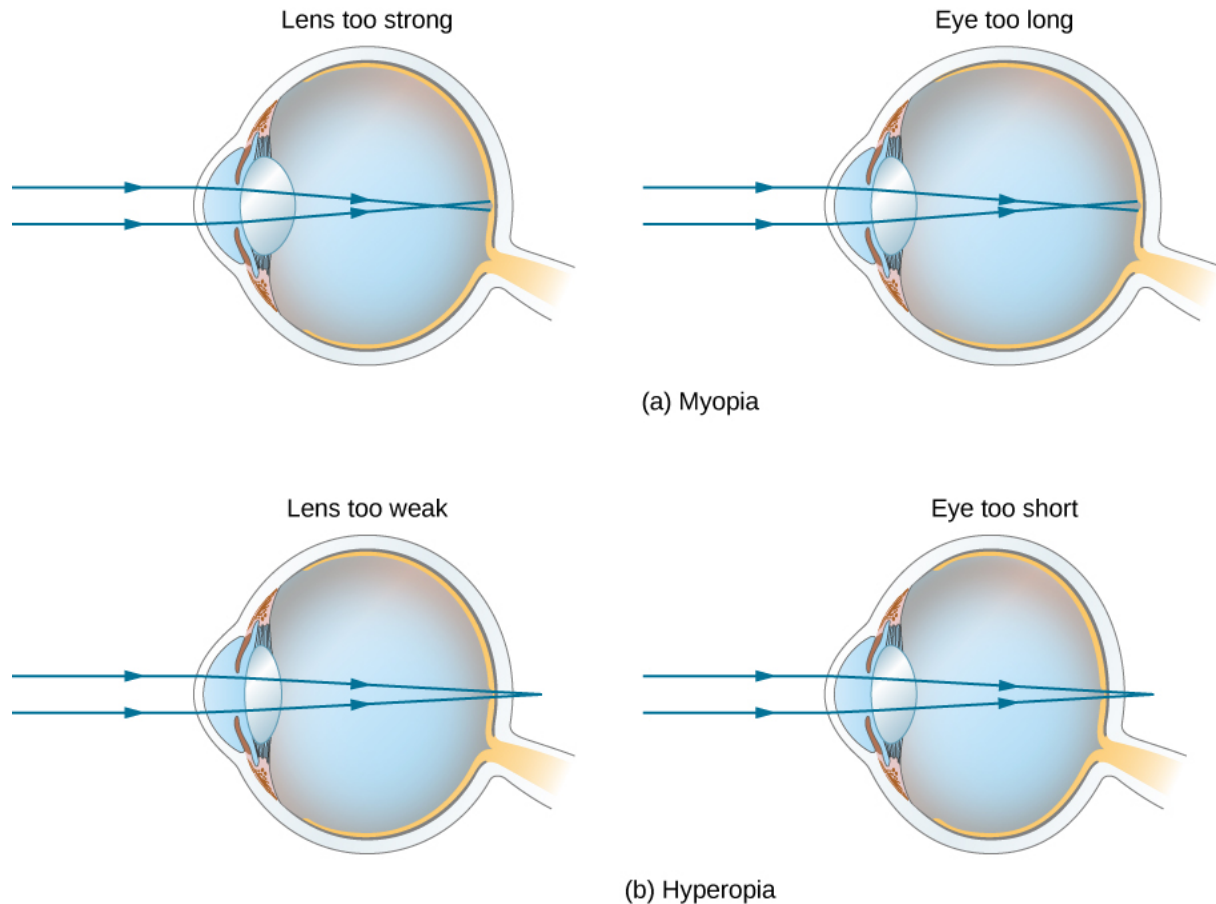
Significance

From the magnification formula, we find $m = -\frac{1.8 \text{ cm}}{25 \text{ cm}} = -0.073$. Since $m < 0$, the image is inverted in orientation with respect to the object. From the absolute value of m we see that the image is much smaller than the object; in fact, it is only 7% of the size of the object.

Vision Correction

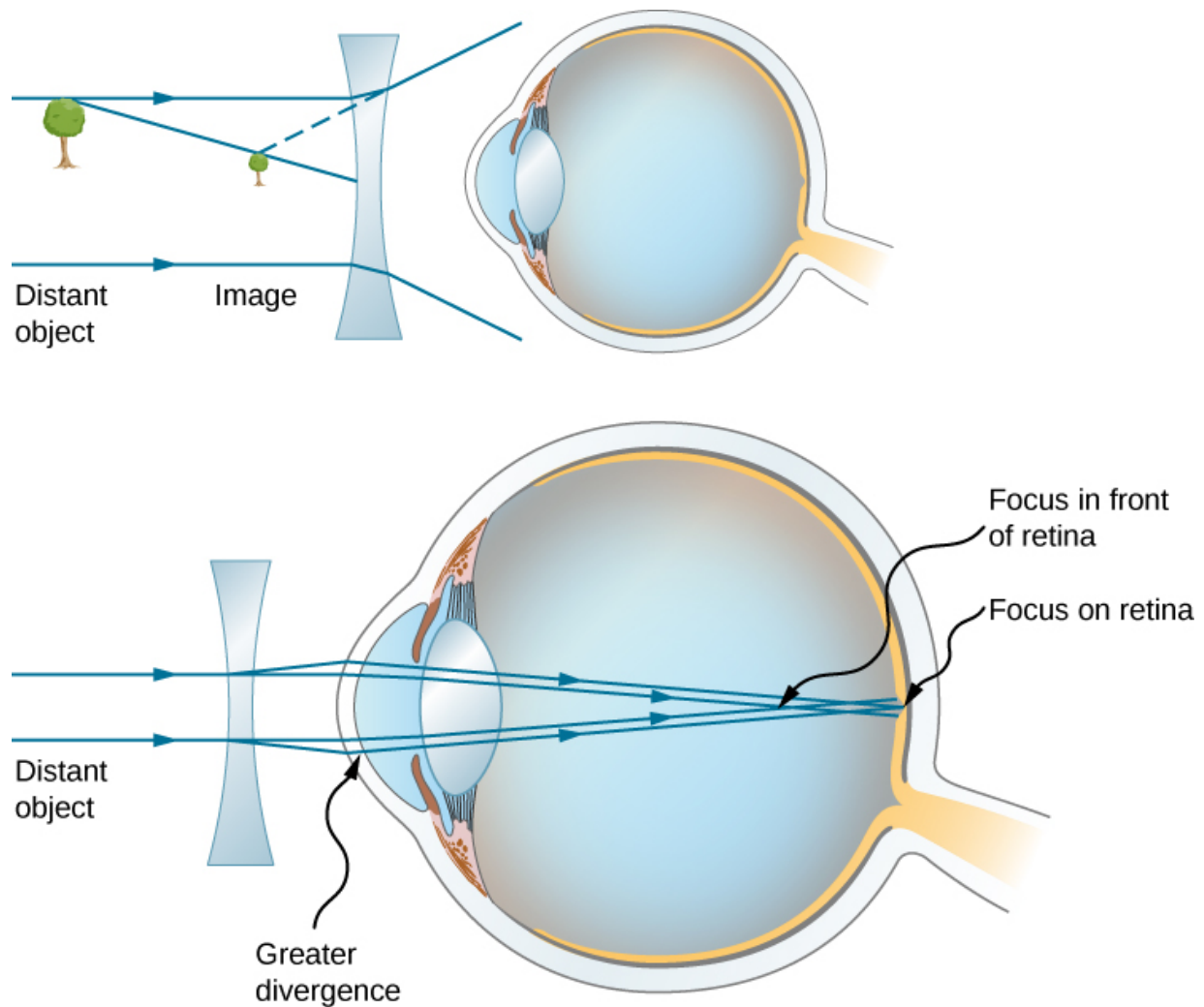
The need for some type of vision correction is very common. Typical vision defects are easy to understand with geometric optics, and some are simple to correct. [\[link\]](#) illustrates two common vision defects. **Nearsightedness**, or **myopia**, is the ability to see near objects, whereas distant objects are blurry. The eye overconverges the nearly parallel rays from a distant object, and the rays cross in front of the retina. More divergent rays from a close object are converged on the retina for a clear image. The distance to the farthest object that can be seen clearly is called the far point of the eye (normally the far point is at infinity). **Farsightedness**, or **hyperopia**, is the

ability to see far objects clearly, whereas near objects are blurry. A farsighted eye does not sufficiently converge the rays from a near object to make the rays meet on the retina.



(a) The nearsighted (myopic) eye converges rays from a distant object in front of the retina, so they have diverged when they strike the retina, producing a blurry image. An eye lens that is too powerful can cause nearsightedness, or the eye may be too long. (b) The farsighted (hyperopic) eye is unable to converge the rays from a close object on the retina, producing blurry near-field vision. An eye lens with insufficient optical power or an eye that is too short can cause farsightedness.

Since the nearsighted eye overconverges light rays, the correction for nearsightedness consists of placing a diverging eyeglass lens in front of the eye, as shown in [\[link\]](#). This reduces the optical power of an eye that is too powerful (recall that the focal length of a diverging lens is negative, so its optical power is negative). Another way to understand this correction is that a diverging lens will cause the incoming rays to diverge more to compensate for the excessive convergence caused by the lens system of the eye. The image produced by the diverging eyeglass lens serves as the (optical) object for the eye, and because the eye cannot focus on objects beyond its far point, the diverging lens must form an image of distant (physical) objects at a point that is closer than the far point.



Correction of nearsightedness requires a diverging lens that compensates for overconvergence by the eye. The diverging lens produces an image closer to the eye than the physical object. This image serves as the optical object for the eye, and the nearsighted person can see it clearly because it is closer than their far point.

Example:**Correcting Nearsightedness**

What optical power of eyeglass lens is needed to correct the vision of a nearsighted person whose far point is 30.0 cm? Assume the corrective lens is fixed 1.50 cm away from the eye.

Strategy

You want this nearsighted person to be able to see distant objects clearly, which means that the eyeglass lens must produce an image 30.0 cm from the eye for an object at infinity. An image 30.0 cm from the eye will be $30.0 \text{ cm} - 1.50 \text{ cm} = 28.5 \text{ cm}$ from the eyeglass lens. Therefore, we must have $d_i = -28.5 \text{ cm}$ when $d_o = \infty$. The image distance is negative because it is on the same side of the eyeglass lens as the object.

Solution

Since d_i and d_o are known, we can find the optical power of the eyeglass lens by using [\[link\]](#):

Equation:

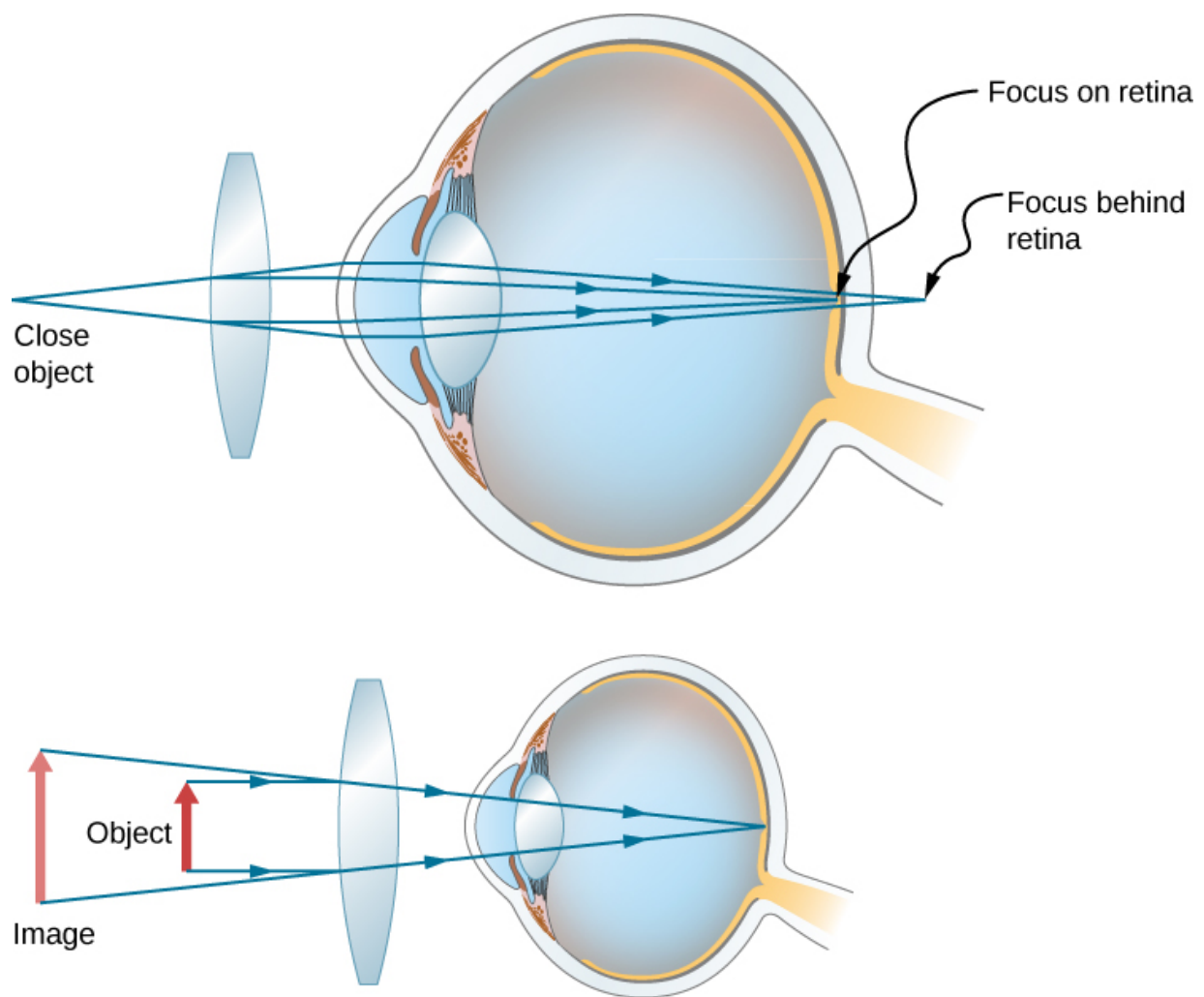
$$P = \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{\infty} + \frac{1}{-0.285 \text{ m}} = -3.51\text{D}.$$

Significance

The negative optical power indicates a diverging (or concave) lens, as expected. If you examine eyeglasses for nearsighted people, you will find the lenses are thinnest in the center. Additionally, if you examine a prescription for eyeglasses for nearsighted people, you will find that the prescribed optical power is negative and given in units of diopters.

Correcting farsightedness consists simply of using the opposite type of lens as for nearsightedness (i.e., a converging lens), as shown in [\[link\]](#).

Such a lens will produce an image of physical objects that are closer than the near point at a distance that is between the near point and the far point, so that the person can see the image clearly. To determine the optical power needed for correction, you must therefore know the person's near point, as explained in [\[link\]](#).



Correction of farsightedness uses a converging lens that compensates for the underconvergence by the eye. The converging lens produces an

image farther from the eye than the object, so that the farsighted person can see it clearly.

Example:

Correcting Farsightedness

What optical power of eyeglass lens is needed to allow a farsighted person, whose near point is 1.00 m, to see an object clearly that is 25.0 cm from the eye? Assume the corrective lens is fixed 1.5 cm from the eye.

Strategy

When an object is 25.0 cm from the person's eyes, the eyeglass lens must produce an image 1.00 m away (the near point), so that the person can see it clearly. An image 1.00 m from the eye will be

$100\text{ cm} - 1.5\text{ cm} = 98.5\text{ cm}$ from the eyeglass lens because the eyeglass lens is 1.5 cm from the eye. Therefore, $d_i = -98.5\text{ cm}$, where the minus sign indicates that the image is on the same side of the lens as the object.

The object is $25.0\text{ cm} - 1.5\text{ cm} = 23.5\text{ cm}$ from the eyeglass lens, so $d_o = 23.5\text{ cm}$.

Solution

Since d_i and d_o are known, we can find the optical power of the eyeglass lens by using [\[link\]](#):

Equation:

$$P = \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{0.235\text{ m}} + \frac{1}{-0.985\text{ m}} = +3.24\text{ D}.$$

Significance

The positive optical power indicates a converging (convex) lens, as expected. If you examine eyeglasses of farsighted people, you will find the lenses to be thickest in the center. In addition, prescription eyeglasses for farsighted people have a prescribed optical power that is positive.

Summary

- Image formation by the eye is adequately described by the thin-lens equation.
- The eye produces a real image on the retina by adjusting its focal length in a process called accommodation.
- Nearsightedness, or myopia, is the inability to see far objects and is corrected with a diverging lens to reduce the optical power of the eye.
- Farsightedness, or hyperopia, is the inability to see near objects and is corrected with a converging lens to increase the optical power of the eye.
- In myopia and hyperopia, the corrective lenses produce images at distances that fall between the person's near and far points so that images can be seen clearly.

Conceptual Questions

Exercise:

Problem:

If the lens of a person's eye is removed because of cataracts (as has been done since ancient times), why would you expect an eyeglass lens of about 16 D to be prescribed?

Exercise:

Problem:

When laser light is shone into a relaxed normal-vision eye to repair a tear by spot-welding the retina to the back of the eye, the rays entering the eye must be parallel. Why?

Solution:

A relaxed, normal-vision eye will focus parallel rays of light onto the retina.

Exercise:

Problem:

Why is your vision so blurry when you open your eyes while swimming under water? How does a face mask enable clear vision?

Exercise:**Problem:**

It has become common to replace the cataract-clouded lens of the eye with an internal lens. This intraocular lens can be chosen so that the person has perfect distant vision. Will the person be able to read without glasses? If the person was nearsighted, is the power of the intraocular lens greater or less than the removed lens?

Solution:

A person with an internal lens will need glasses to read because their muscles cannot distort the lens as they do with biological lenses, so they cannot focus on near objects. To correct nearsightedness, the power of the intraocular lens must be less than that of the removed lens.

Exercise:**Problem:**

If the cornea is to be reshaped (this can be done surgically or with contact lenses) to correct myopia, should its curvature be made greater or smaller? Explain.

Problems

Unless otherwise stated, the lens-to-retina distance is 2.00 cm.

Exercise:**Problem:**

What is the power of the eye when viewing an object 50.0 cm away?

Solution:

$$P = 52.0 \text{ D}$$

Exercise:**Problem:**

Calculate the power of the eye when viewing an object 3.00 m away.

Exercise:**Problem:**

The print in many books averages 3.50 mm in height. How high is the image of the print on the retina when the book is held 30.0 cm from the eye?

Solution:

$$\frac{h_i}{h_o} = -\frac{d_i}{d_o} \Rightarrow h_i = -h_o \left(\frac{d_i}{d_o} \right) = -(3.50 \text{ mm}) \left(\frac{2.00 \text{ cm}}{30.0 \text{ cm}} \right) = -0.233 \text{ mm}$$

Exercise:**Problem:**

Suppose a certain person's visual acuity is such that he can see objects clearly that form an image 4.00 μm high on his retina. What is the maximum distance at which he can read the 75.0-cm-high letters on the side of an airplane?

Exercise:

Problem:

People who do very detailed work close up, such as jewelers, often can see objects clearly at much closer distance than the normal 25 cm. (a) What is the power of the eyes of a woman who can see an object clearly at a distance of only 8.00 cm? (b) What is the image size of a 1.00-mm object, such as lettering inside a ring, held at this distance? (c) What would the size of the image be if the object were held at the normal 25.0 cm distance?

Solution:

- a. $P = +62.5 \text{ D}$;
- b. $\frac{h_i}{h_o} = -\frac{d_i}{d_o} \Rightarrow h_i = -0.250 \text{ mm}$;
- c. $h_i = -0.0800 \text{ mm}$

Exercise:**Problem:**

What is the far point of a person whose eyes have a relaxed power of 50.5 D?

Exercise:**Problem:**

What is the near point of a person whose eyes have an accommodated power of 53.5 D?

Solution:

$$P = \frac{1}{d_o} + \frac{1}{d_i} \Rightarrow d_o = 28.6 \text{ cm}$$

Exercise:

Problem:

(a) A laser reshaping the cornea of a myopic patient reduces the power of his eye by 9.00 D, with a $\pm 5.0\%$ uncertainty in the final correction. What is the range of diopters for eyeglass lenses that this person might need after this procedure? (b) Was the person nearsighted or farsighted before the procedure? How do you know?

Exercise:**Problem:**

The power for normal close vision is 54.0 D. In a vision-correction procedure, the power of a patient's eye is increased by 3.00 D. Assuming that this produces normal close vision, what was the patient's near point before the procedure?

Solution:

Originally, the close vision was 51.0 D. Therefore,

$$P = \frac{1}{d_o} + \frac{1}{d_i} \Rightarrow d_o = 1.00 \text{ m}$$

Exercise:**Problem:**

For normal distant vision, the eye has a power of 50.0 D. What was the previous far point of a patient who had laser vision correction that reduced the power of her eye by 7.00 D, producing normal distant vision?

Exercise:**Problem:**

The power for normal distant vision is 50.0 D. A severely myopic patient has a far point of 5.00 cm. By how many diopters should the power of his eye be reduced in laser vision correction to obtain normal distant vision for him?

Solution:

originally, $P = 70.0 \text{ D}$; because the power for normal distant vision is 50.0 D , the power should be decreased by 20.0 D

Exercise:**Problem:**

A student's eyes, while reading the blackboard, have a power of 51.0 D . How far is the board from his eyes?

Exercise:**Problem:**

The power of a physician's eyes is 53.0 D while examining a patient. How far from her eyes is the object that is being examined?

Solution:

$$P = \frac{1}{d_o} + \frac{1}{d_i} \Rightarrow d_o = 0.333 \text{ m}$$

Exercise:**Problem:**

The normal power for distant vision is 50.0 D . A young woman with normal distant vision has a 10.0% ability to accommodate (that is, increase) the power of her eyes. What is the closest object she can see clearly?

Exercise:**Problem:**

The far point of a myopic administrator is 50.0 cm . (a) What is the relaxed power of his eyes? (b) If he has the normal 8.00% ability to accommodate, what is the closest object he can see clearly?

Solution:

a. $P = 52.0 \text{ D}$;
 $P' = 56.16 \text{ D}$

b. $\frac{1}{d_o} + \frac{1}{d_i} = P \Rightarrow d_o = 16.2 \text{ cm}$

Exercise:

Problem:

A very myopic man has a far point of 20.0 cm. What power contact lens (when on the eye) will correct his distant vision?

Exercise:

Problem:

Repeat the previous problem for eyeglasses held 1.50 cm from the eyes.

Solution:

We need $d_i = -18.5 \text{ cm}$ when $d_o = \infty$, so
 $P = -5.41 \text{ D}$

Exercise:

Problem:

A myopic person sees that her contact lens prescription is -4.00 D . What is her far point?

Exercise:

Problem:

Repeat the previous problem for glasses that are 1.75 cm from the eyes.

Solution:

Let x = far point

$$\Rightarrow P = \frac{1}{-(x-0.0175 \text{ m})} + \frac{1}{\infty} \Rightarrow -xP + (0.0175 \text{ m})P = 1$$

$$\Rightarrow x = 26.8 \text{ cm}$$

Exercise:

Problem:

The contact lens prescription for a mildly farsighted person is 0.750 D, and the person has a near point of 29.0 cm. What is the power of the tear layer between the cornea and the lens if the correction is ideal, taking the tear layer into account?

Glossary

accommodation

use of the ciliary muscles to adjust the shape of the eye lens for focusing on near or far objects

far point

furthest point an eye can see in focus

farsightedness (or hyperopia)

visual defect in which near objects appear blurred because their images are focused behind the retina rather than on the retina; a farsighted person can see far objects clearly but near objects appear blurred

near point

closest point an eye can see in focus

nearsightedness (or myopia)

visual defect in which far objects appear blurred because their images are focused in front of the retina rather than on the retina; a nearsighted person can see near objects clearly but far objects appear blurred

optical power

(P) inverse of the focal length of a lens, with the focal length expressed in meters. The optical power P of a lens is expressed in units of diopters D; that is, $1\text{D} = 1/\text{m} = 1\text{ m}^{-1}$

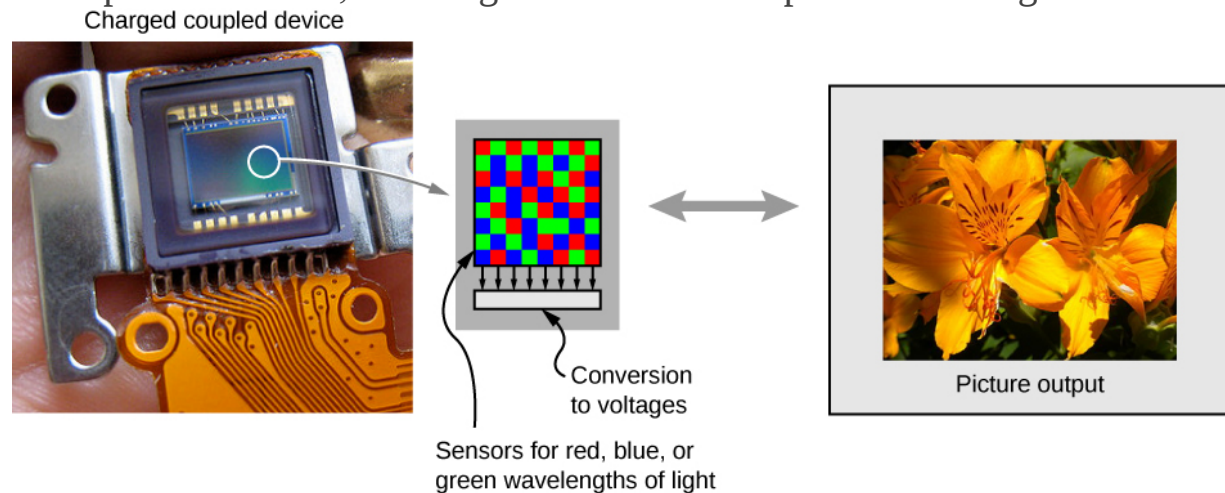
The Camera

By the end of this section, you will be able to:

- Describe the optics of a camera
- Characterize the image created by a camera

Cameras are very common in our everyday life. Between 1825 and 1827, French inventor Nicéphore Niépce successfully photographed an image created by a primitive camera. Since then, enormous progress has been achieved in the design of cameras and camera-based detectors.

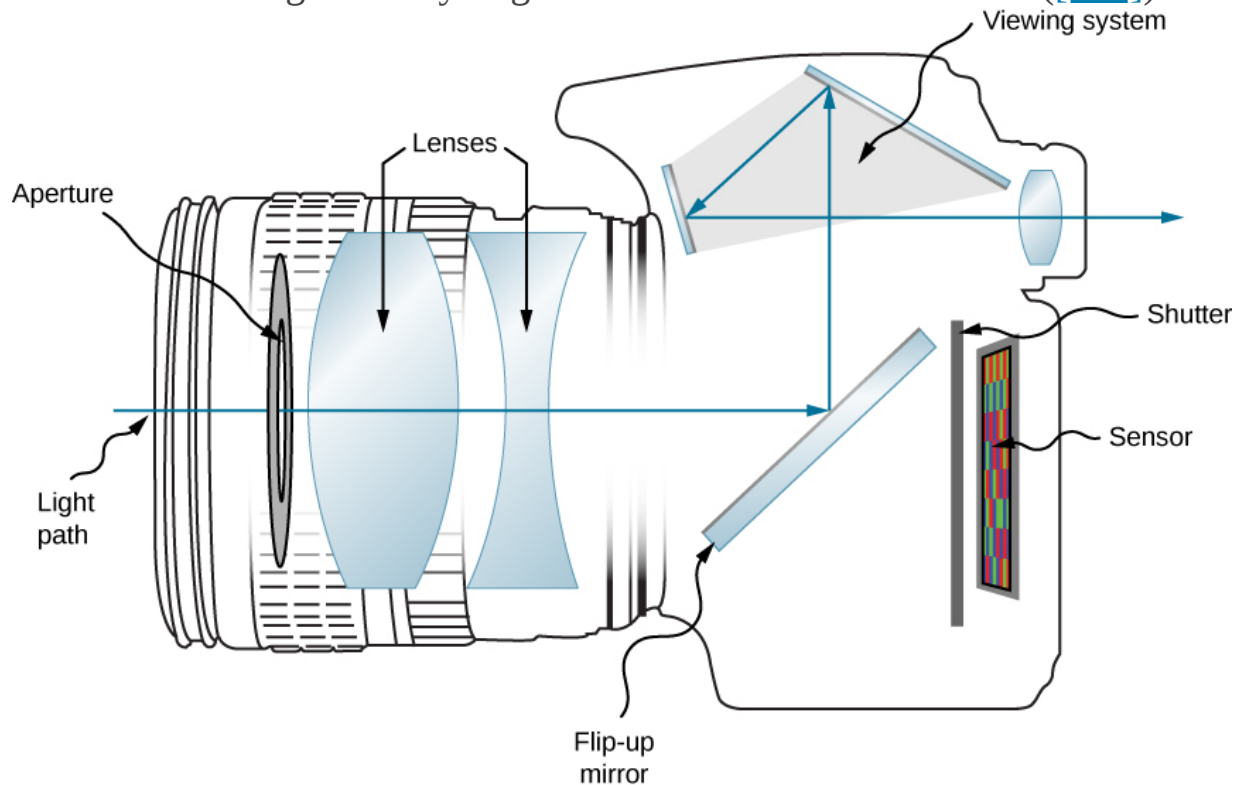
Initially, photographs were recorded by using the light-sensitive reaction of silver-based compounds such as silver chloride or silver bromide. Silver-based photographic paper was in common use until the advent of digital photography in the 1980s, which is intimately connected to **charge-coupled device (CCD)** detectors. In a nutshell, a CCD is a semiconductor chip that records images as a matrix of tiny pixels, each pixel located in a “bin” in the surface. Each pixel is capable of detecting the intensity of light impinging on it. Color is brought into play by putting red-, blue-, and green-colored filters over the pixels, resulting in colored digital images ([\[link\]](#)). At its best resolution, one CCD pixel corresponds to one pixel of the image. To reduce the resolution and decrease the size of the file, we can “bin” several CCD pixels into one, resulting in a smaller but “pixelated” image.



A charge-coupled device (CCD) converts light signals into electronic signals, enabling electronic processing and storage of visual images.

This is the basis for electronic imaging in all digital cameras, from cell phones to movie cameras. (credit left: modification of work by Bruce Turner)

Clearly, electronics is a big part of a digital camera; however, the underlying physics is basic optics. As a matter of fact, the optics of a camera are pretty much the same as those of a single lens with an object distance that is significantly larger than the lens's focal distance ([\[link\]](#)).



Modern digital cameras have several lenses to produce a clear image with minimal aberration and use red, blue, and green filters to produce a color image.

For instance, let us consider the camera in a smartphone. An average smartphone camera is equipped with a stationary wide-angle lens with a focal length of about 4–5 mm. (This focal length is about equal to the

thickness of the phone.) The image created by the lens is focused on the CCD detector mounted at the opposite side of the phone. In a cell phone, the lens and the CCD cannot move relative to each other. So how do we make sure that both the images of a distant and a close object are in focus?

Recall that a human eye can accommodate for distant and close images by changing its focal distance. A cell phone camera cannot do that because the distance from the lens to the detector is fixed. Here is where the small focal distance becomes important. Let us assume we have a camera with a 5-mm focal distance. What is the image distance for a selfie? The object distance for a selfie (the length of the hand holding the phone) is about 50 cm. Using the thin-lens equation, we can write

Equation:

$$\frac{1}{5 \text{ mm}} = \frac{1}{500 \text{ mm}} + \frac{1}{d_i}$$

We then obtain the image distance:

Equation:

$$\frac{1}{d_i} = \frac{1}{5 \text{ mm}} - \frac{1}{500 \text{ mm}}$$

Note that the object distance is 100 times larger than the focal distance. We can clearly see that the $1/(500 \text{ mm})$ term is significantly smaller than $1/(5 \text{ mm})$, which means that the image distance is pretty much equal to the lens's focal length. An actual calculation gives us the image distance $d_i = 5.05 \text{ mm}$. This value is extremely close to the lens's focal distance.

Now let us consider the case of a distant object. Let us say that we would like to take a picture of a person standing about 5 m from us. Using the thin-lens equation again, we obtain the image distance of 5.005 mm. The farther the object is from the lens, the closer the image distance is to the focal distance. At the limiting case of an infinitely distant object, we obtain the image distance exactly equal to the focal distance of the lens.

As you can see, the difference between the image distance for a selfie and the image distance for a distant object is just about 0.05 mm or 50 microns. Even a short object distance such as the length of your hand is two orders of magnitude larger than the lens's focal length, resulting in minute variations of the image distance. (The 50-micron difference is smaller than the thickness of an average sheet of paper.) Such a small difference can be easily accommodated by the same detector, positioned at the focal distance of the lens. Image analysis software can help improve image quality.

Conventional point-and-shoot cameras often use a movable lens to change the lens-to-image distance. Complex lenses of the more expensive mirror reflex cameras allow for superb quality photographic images. The optics of these camera lenses is beyond the scope of this textbook.

Summary

- Cameras use combinations of lenses to create an image for recording.
- Digital photography is based on charge-coupled devices (CCDs) that break an image into tiny “pixels” that can be converted into electronic signals.

Glossary

charge-coupled device (CCD)

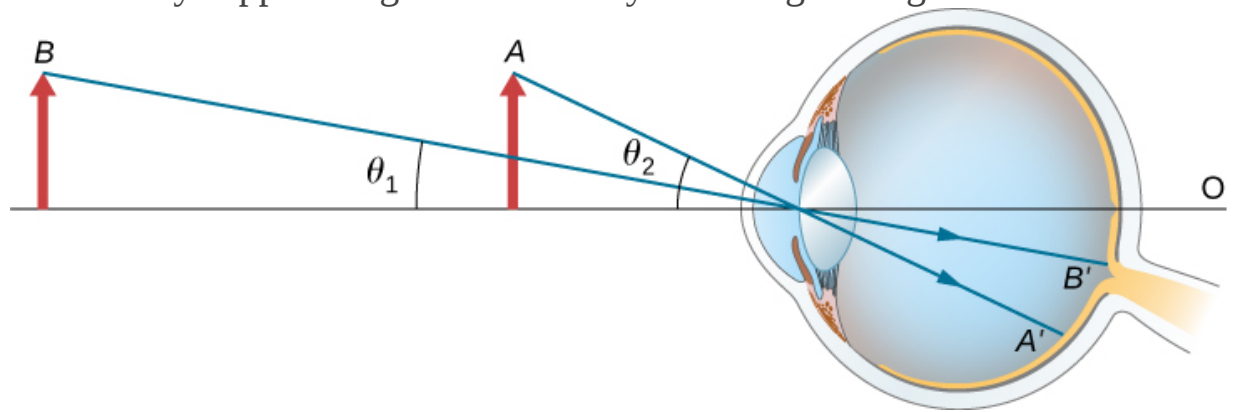
semiconductor chip that converts a light image into tiny pixels that can be converted into electronic signals of color and intensity

The Simple Magnifier

By the end of this section, you will be able to:

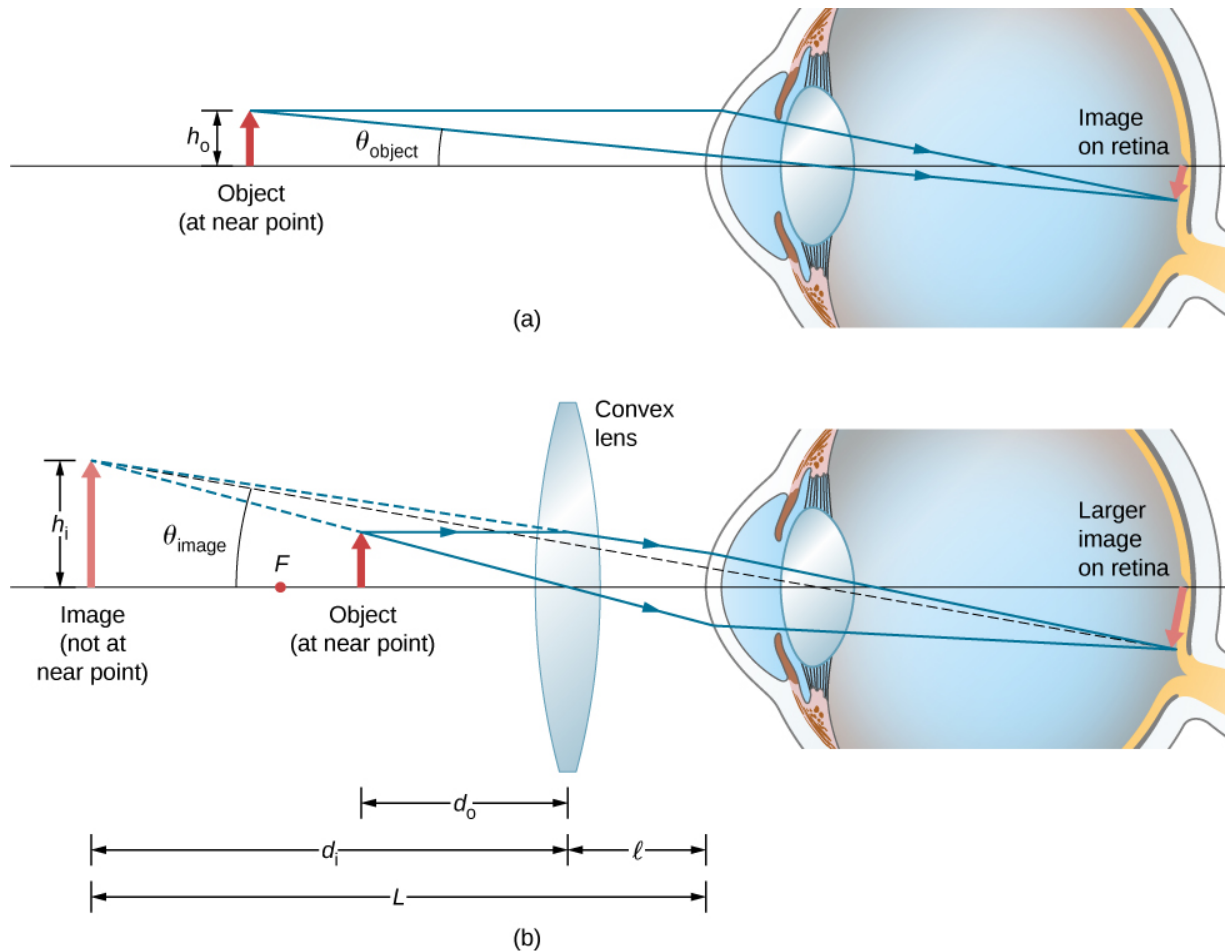
- Understand the optics of a simple magnifier
- Characterize the image created by a simple magnifier

The apparent size of an object perceived by the eye depends on the angle the object subtends from the eye. As shown in [\[link\]](#), the object at A subtends a larger angle from the eye than when it is positioned at point B . Thus, the object at A forms a larger image on the retina (see OA') than when it is positioned at B (see OB'). Thus, objects that subtend large angles from the eye appear larger because they form larger images on the retina.



Size perceived by an eye is determined by the angle subtended by the object. An image formed on the retina by an object at A is larger than an image formed on the retina by the same object positioned at B (compared image heights OA' to OB').

We have seen that, when an object is placed within a focal length of a convex lens, its image is virtual, upright, and larger than the object (see part (b) of [\[link\]](#)). Thus, when such an image produced by a convex lens serves as the object for the eye, as shown in [\[link\]](#), the image on the retina is enlarged, because the image produced by the lens subtends a larger angle in the eye than does the object. A convex lens used for this purpose is called a **magnifying glass** or a **simple magnifier**.



The simple magnifier is a convex lens used to produce an enlarged image of an object on the retina. (a) With no convex lens, the object subtends an angle θ_{object} from the eye. (b) With the convex lens in place, the image produced by the convex lens subtends an angle θ_{image} from the eye, with $\theta_{\text{image}} > \theta_{\text{object}}$. Thus, the image on the retina is larger with the convex lens in place.

To account for the magnification of a magnifying lens, we compare the angle subtended by the image (created by the lens) with the angle subtended by the object (viewed with no lens), as shown in [\[link\]](#). We assume that the object is situated at the near point of the eye, because this is the object distance at which the unaided eye can form the largest image on the retina. We will compare the magnified images created by a lens with this

maximum image size for the unaided eye. The magnification of an image when observed by the eye is the **angular magnification** M , which is defined by the ratio of the angle θ_{image} subtended by the image to the angle θ_{object} subtended by the object:

Note:

Equation:

$$M = \frac{\theta_{\text{image}}}{\theta_{\text{object}}}.$$

Consider the situation shown in [\[link\]](#). The magnifying lens is held a distance ℓ from the eye, and the image produced by the magnifier forms a distance L from the eye. We want to calculate the angular magnification for any arbitrary L and ℓ . In the small-angle approximation, the angular size θ_{image} of the image is h_i/L . The angular size θ_{object} of the object at the near point is $\theta_{\text{object}} = h_o/25 \text{ cm}$. The angular magnification is then

Equation:

$$M = \frac{\theta_{\text{image}}}{\theta_{\text{object}}} = \frac{h_i(25 \text{ cm})}{Lh_o}.$$

Using [\[link\]](#) for linear magnification

Equation:

$$m = -\frac{d_i}{d_o} = \frac{h_i}{h_o}$$

and the thin-lens equation

Equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

in [\[link\]](#), we arrive at the following expression for the angular magnification of a magnifying lens:

Equation:

$$\begin{aligned} M &= \left(-\frac{d_i}{d_o} \right) \left(\frac{25 \text{ cm}}{L} \right) \\ &= -d_i \left(\frac{1}{f} - \frac{1}{d_i} \right) \left(\frac{25 \text{ cm}}{L} \right) \\ &= \left(1 - \frac{d_i}{f} \right) \left(\frac{25 \text{ cm}}{L} \right) \end{aligned}$$

From part (b) of the figure, we see that the absolute value of the image distance is $|d_i| = L - \ell$. Note that $d_i < 0$ because the image is virtual, so we can dispense with the absolute value by explicitly inserting the minus sign: $-d_i = L - \ell$. Inserting this into [\[link\]](#) gives us the final equation for the angular magnification of a magnifying lens:

Note:

Equation:

$$M = \left(\frac{25 \text{ cm}}{L} \right) \left(1 + \frac{L - \ell}{f} \right).$$

Note that all the quantities in this equation have to be expressed in centimeters. Often, we want the image to be at the near-point distance ($L = 25 \text{ cm}$) to get maximum magnification, and we hold the magnifying lens close to the eye ($\ell = 0$). In this case, [\[link\]](#) gives

Equation:

$$M = 1 + \frac{25 \text{ cm}}{f}$$

which shows that the greatest magnification occurs for the lens with the shortest focal length. In addition, when the image is at the near-point distance and the lens is held close to the eye ($\ell = 0$), then $L = d_i = 25 \text{ cm}$ and [\[link\]](#) becomes

Equation:

$$M = \frac{h_i}{h_o} = m$$

where m is the linear magnification ([\[link\]](#)) derived for spherical mirrors and thin lenses. Another useful situation is when the image is at infinity ($L = \infty$). [\[link\]](#) then takes the form

Equation:

$$M(L = \infty) = \frac{25 \text{ cm}}{f}.$$

The resulting magnification is simply the ratio of the near-point distance to the focal length of the magnifying lens, so a lens with a shorter focal length gives a stronger magnification. Although this magnification is smaller by 1 than the magnification obtained with the image at the near point, it provides for the most comfortable viewing conditions, because the eye is relaxed when viewing a distant object.

By comparing [\[link\]](#) with [\[link\]](#), we see that the range of angular magnification of a given converging lens is

Note:

Equation:

$$\frac{25 \text{ cm}}{f} \leq M \leq 1 + \frac{25 \text{ cm}}{f}.$$

Example:

Magnifying a Diamond

A jeweler wishes to inspect a 3.0-mm-diameter diamond with a magnifier. The diamond is held at the jeweler's near point (25 cm), and the jeweler holds the magnifying lens close to his eye.

(a) What should the focal length of the magnifying lens be to see a 15-mm-diameter image of the diamond?

(b) What should the focal length of the magnifying lens be to obtain $10\times$ magnification?

Strategy

We need to determine the requisite magnification of the magnifier. Because the jeweler holds the magnifying lens close to his eye, we can use [\[link\]](#) to find the focal length of the magnifying lens.

Solution

- a. The required linear magnification is the ratio of the desired image diameter to the diamond's actual diameter ([\[link\]](#)). Because the jeweler holds the magnifying lens close to his eye and the image forms at his near point, the linear magnification is the same as the angular magnification, so

Equation:

$$M = m = \frac{h_i}{h_o} = \frac{15 \text{ mm}}{3.0 \text{ mm}} = 5.0.$$

The focal length f of the magnifying lens may be calculated by solving [\[link\]](#) for f , which gives

Equation:

$$\begin{aligned} M &= 1 + \frac{25 \text{ cm}}{f} \\ f &= \frac{25 \text{ cm}}{M-1} = \frac{25 \text{ cm}}{5.0-1} = 6.3 \text{ cm} \end{aligned}$$

b. To get an image magnified by a factor of ten, we again solve [\[link\]](#) for f , but this time we use $M = 10$. The result is

Equation:

$$f = \frac{25 \text{ cm}}{M - 1} = \frac{25 \text{ cm}}{10 - 1} = 2.8 \text{ cm}.$$

Significance

Note that a greater magnification is achieved by using a lens with a smaller focal length. We thus need to use a lens with radii of curvature that are less than a few centimeters and hold it very close to our eye. This is not very convenient. A compound microscope, explored in the following section, can overcome this drawback.

Summary

- A simple magnifier is a converging lens and produces a magnified virtual image of an object located within the focal length of the lens.
- Angular magnification accounts for magnification of an image created by a magnifier. It is equal to the ratio of the angle subtended by the image to that subtended by the object when the object is observed by the unaided eye.
- Angular magnification is greater for magnifying lenses with smaller focal lengths.
- Simple magnifiers can produce as great as tenfold ($10 \times$) magnification.

Problems

Exercise:

Problem:

If the image formed on the retina subtends an angle of 30° and the object subtends an angle of 5° , what is the magnification of the image?

Solution:

$$M = 6 \times$$

Exercise:**Problem:**

What is the magnification of a magnifying lens with a focal length of 10 cm if it is held 3.0 cm from the eye and the object is 12 cm from the eye?

Exercise:**Problem:**

How far should you hold a 2.1 cm-focal length magnifying glass from an object to obtain a magnification of $10 \times$? Assume you place your eye 5.0 cm from the magnifying glass.

Solution:

$$M = \left(\frac{25 \text{ cm}}{L} \right) \left(1 + \frac{L - \ell}{f} \right)$$

$$L - \ell = d_o$$

$$d_o = 13 \text{ cm}$$

Exercise:**Problem:**

You hold a 5.0 cm-focal length magnifying glass as close as possible to your eye. If you have a normal near point, what is the magnification?

Exercise:**Problem:**

You view a mountain with a magnifying glass of focal length $f = 10 \text{ cm}$. What is the magnification?

Solution:

$$M = 2.5 \times$$

Exercise:**Problem:**

You view an object by holding a 2.5 cm-focal length magnifying glass 10 cm away from it. How far from your eye should you hold the magnifying glass to obtain a magnification of $10 \times$?

Exercise:**Problem:**

A magnifying glass forms an image 10 cm on the opposite side of the lens from the object, which is 10 cm away. What is the magnification of this lens for a person with a normal near point if their eye 12 cm from the object?

Solution:

$$M = -2.1 \times$$

Exercise:**Problem:**

An object viewed with the naked eye subtends a 2° angle. If you view the object through a $10 \times$ magnifying glass, what angle is subtended by the image formed on your retina?

Exercise:**Problem:**

For a normal, relaxed eye, a magnifying glass produces an angular magnification of 4.0. What is the largest magnification possible with this magnifying glass?

Solution:

$$M = \frac{25 \text{ cm}}{f}$$

$$M_{\text{max}} = 5$$

Exercise:

Problem:

What range of magnification is possible with a 7.0 cm-focal length converging lens?

Exercise:

Problem:

A magnifying glass produces an angular magnification of 4.5 when used by a young person with a near point of 18 cm. What is the maximum angular magnification obtained by an older person with a near point of 45 cm?

Solution:

$$M_{\text{max}}^{\text{young}} = 1 + \frac{18 \text{ cm}}{f} \Rightarrow f = \frac{18 \text{ cm}}{M_{\text{max}}^{\text{young}} - 1}$$

$$M_{\text{max}}^{\text{old}} = 9.8 \times$$

Glossary

angular magnification

ratio of the angle subtended by an object observed with a magnifier to that observed by the naked eye

simple magnifier (or magnifying glass)

converging lens that produces a virtual image of an object that is within the focal length of the lens

Microscopes and Telescopes

By the end of this section, you will be able to:

- Explain the physics behind the operation of microscopes and telescopes
- Describe the image created by these instruments and calculate their magnifications

Microscopes and telescopes are major instruments that have contributed hugely to our current understanding of the micro- and macroscopic worlds. The invention of these devices led to numerous discoveries in disciplines such as physics, astronomy, and biology, to name a few. In this section, we explain the basic physics that make these instruments work.

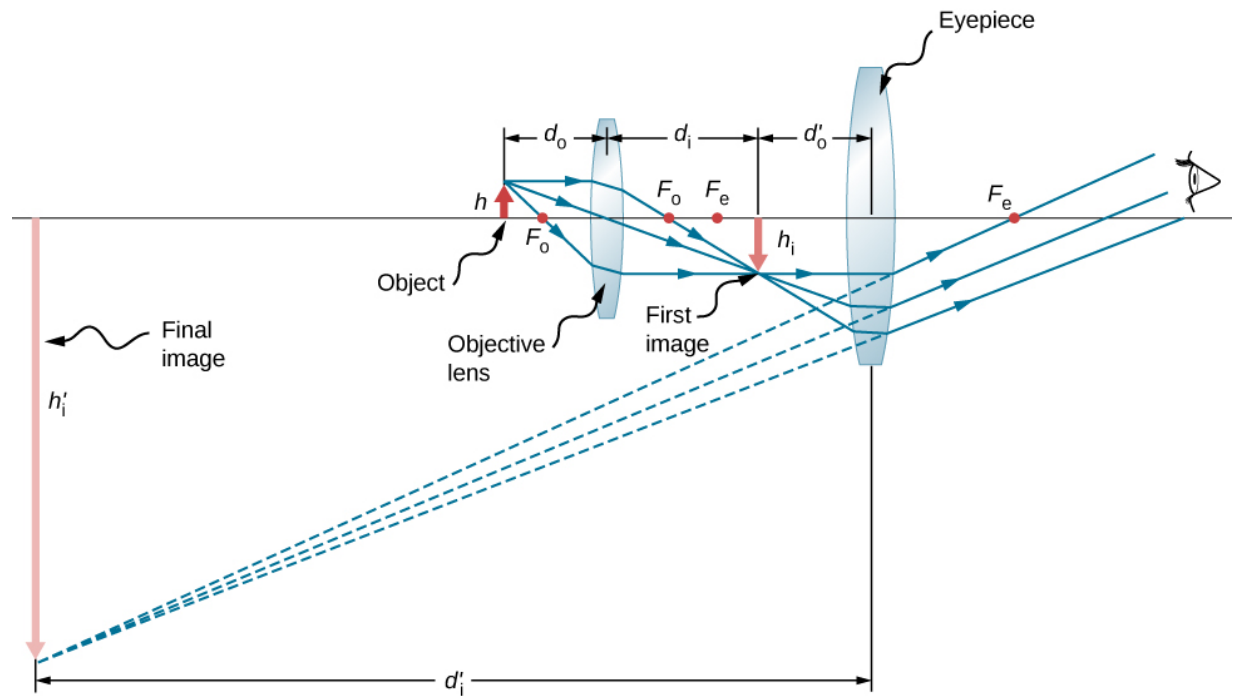
Microscopes

Although the eye is marvelous in its ability to see objects large and small, it obviously is limited in the smallest details it can detect. The desire to see beyond what is possible with the naked eye led to the use of optical instruments. We have seen that a simple convex lens can create a magnified image, but it is hard to get large magnification with such a lens. A magnification greater than $5\times$ is difficult without distorting the image. To get higher magnification, we can combine the simple magnifying glass with one or more additional lenses. In this section, we examine microscopes that enlarge the details that we cannot see with the naked eye.

Microscopes were first developed in the early 1600s by eyeglass makers in The Netherlands and Denmark. The simplest **compound microscope** is constructed from two convex lenses ([\[link\]](#)). The **objective** lens is a convex lens of short focal length (i.e., high power) with typical magnification from $5\times$ to $100\times$. The **eyepiece**, also referred to as the ocular, is a convex lens of longer focal length.

The purpose of a microscope is to create magnified images of small objects, and both lenses contribute to the final magnification. Also, the final enlarged image is produced sufficiently far from the observer to be easily

viewed, since the eye cannot focus on objects or images that are too close (i.e., closer than the near point of the eye).



A compound microscope is composed of two lenses: an objective and an eyepiece. The objective forms the first image, which is larger than the object. This first image is inside the focal length of the eyepiece and serves as the object for the eyepiece. The eyepiece forms the final image that is further magnified. The d_o and d_i shown will be discussed with superscripts "obj" below to denote they are measured from the objective lens, while the eye piece variables will have superscripts of "eye" to denote this lens.

To see how the microscope in [\[link\]](#) forms an image, consider its two lenses in succession. The object is just beyond the focal length f^{obj} of the objective lens, producing a real, inverted image that is larger than the object. This first image serves as the object for the second lens, or eyepiece. The eyepiece is positioned so that the first image is within its focal length

f^{eye} , so that it can further magnify the image. In a sense, it acts as a magnifying glass that magnifies the intermediate image produced by the objective. The image produced by the eyepiece is a magnified virtual image. The final image remains inverted but is farther from the observer than the object, making it easy to view.

The eye views the virtual image created by the eyepiece, which serves as the object for the lens in the eye. The virtual image formed by the eyepiece is well outside the focal length of the eye, so the eye forms a real image on the retina.

The magnification of the microscope is the product of the linear magnification m^{obj} by the objective and the angular magnification M^{eye} by the eyepiece. These are given by

Equation:

$$m^{\text{obj}} = -\frac{d_i^{\text{obj}}}{d_o^{\text{obj}}} \approx -\frac{d_i^{\text{obj}}}{f^{\text{obj}}} \text{ (linear magnification by objective)}$$

$$M^{\text{eye}} = 1 + \frac{25 \text{ cm}}{f^{\text{eye}}} \text{ (angular magnification by eyepiece)}$$

Here, f^{obj} and f^{eye} are the focal lengths of the objective and the eyepiece, respectively. We assume that the final image is formed at the near point of the eye, providing the largest magnification. Note that the angular magnification of the eyepiece is the same as obtained earlier for the simple magnifying glass. This should not be surprising, because the eyepiece is essentially a magnifying glass, and the same physics applies here. The **net magnification** M_{net} of the compound microscope is the product of the linear magnification of the objective and the angular magnification of the eyepiece:

Note:

Equation:

$$M_{\text{net}} = m^{\text{obj}} M^{\text{eye}} = -\frac{d_i^{\text{obj}} (f^{\text{eye}} + 25 \text{ cm})}{f^{\text{obj}} f^{\text{eye}}}.$$

Example:

Microscope Magnification

Calculate the magnification of an object placed 6.20 mm from a compound microscope that has a 6.00 mm-focal length objective and a 50.0 mm-focal length eyepiece. The objective and eyepiece are separated by 23.0 cm.

Strategy

This situation is similar to that shown in [\[link\]](#). To find the overall magnification, we must know the linear magnification of the objective and the angular magnification of the eyepiece. We can use [\[link\]](#), but we need to use the thin-lens equation to find the image distance d_i^{obj} of the objective.

Solution

Solving the thin-lens equation for d_i^{obj} gives

Equation:

$$\begin{aligned} d_i^{\text{obj}} &= \left(\frac{1}{f^{\text{obj}}} - \frac{1}{d_o^{\text{obj}}} \right)^{-1} \\ &= \left(\frac{1}{6.00 \text{ mm}} - \frac{1}{6.20 \text{ mm}} \right)^{-1} = 186 \text{ mm} = 18.6 \text{ cm} \end{aligned}$$

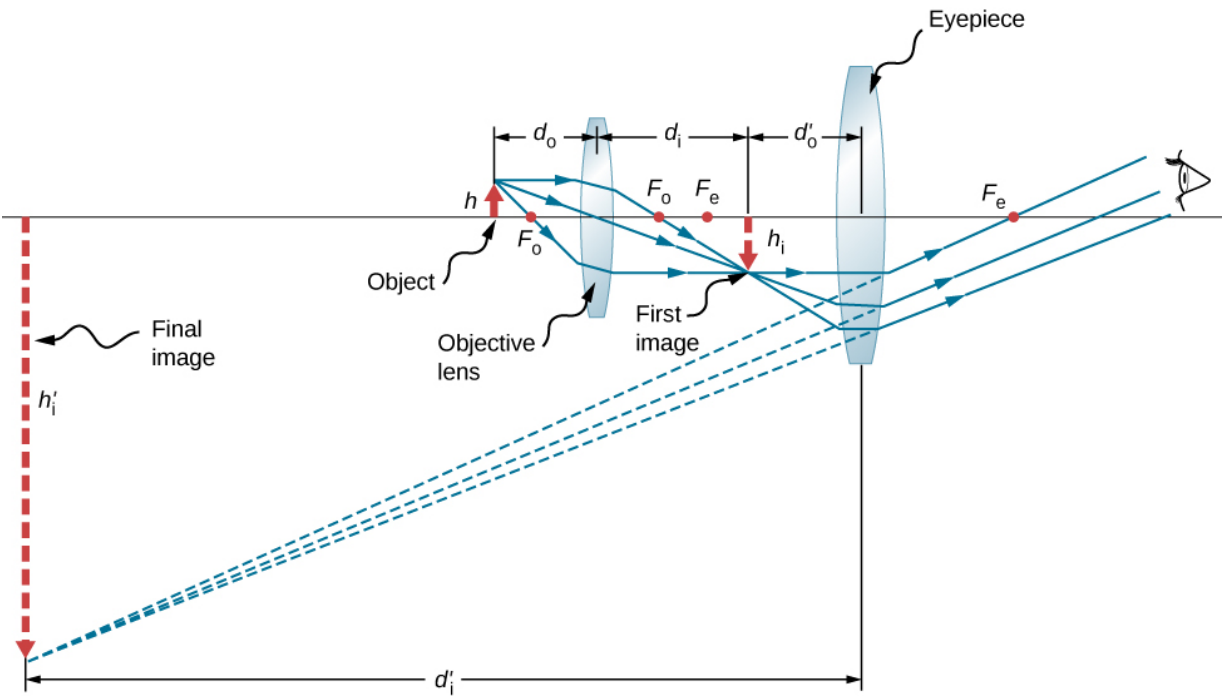
Inserting this result into [\[link\]](#) along with the known values $f^{\text{obj}} = 6.00 \text{ mm} = 0.600 \text{ cm}$ and $f^{\text{eye}} = 50.0 \text{ mm} = 5.00 \text{ cm}$ gives

Equation:

$$\begin{aligned} M_{\text{net}} &= -\frac{d_i^{\text{obj}} (f^{\text{eye}} + 25 \text{ cm})}{f^{\text{obj}} f^{\text{eye}}} \\ &= -\frac{(18.6 \text{ cm})(5.00 \text{ cm} + 25 \text{ cm})}{(0.600 \text{ cm})(5.00 \text{ cm})} \\ &= -186 \end{aligned}$$

Significance

Both the objective and the eyepiece contribute to the overall magnification, which is large and negative, consistent with [\[link\]](#), where the image is seen to be large and inverted. In this case, the image is virtual and inverted, which cannot happen for a single element (see [\[link\]](#)).



A compound microscope with the image created at infinity.

We now calculate the magnifying power of a microscope when the image is at infinity, as shown in [\[link\]](#), because this makes for the most relaxed viewing. The magnifying power of the microscope is the product of linear magnification m^{obj} of the objective and the angular magnification M^{eye} of the eyepiece. The magnification of the objective can be obtained from the thin-lens equation for magnification, which is

Equation:

$$m^{\text{obj}} = -\frac{d_i^{\text{obj}}}{d_o^{\text{obj}}}$$

If the final image is at infinity, then the image created by the objective must be located at the focal point of the eyepiece. This may be seen by considering the thin-lens equation with $d_i = \infty$ or by recalling that rays that pass through the focal point exit the lens parallel to each other, which is equivalent to focusing at infinity. For many microscopes, the distance between the image-side focal point of the objective and the object-side focal point of the eyepiece is standardized at $L = 16$ cm. This distance is called the tube length of the microscope. If the length of the compound microscope L is roughly the focal length of the objective, we can substitute L in for d_i^{obj} to get

Equation:

$$m^{\text{obj}} = \frac{L}{f^{\text{obj}}} = \frac{16 \text{ cm}}{f^{\text{obj}}}.$$

We now need to calculate the angular magnification of the eyepiece with the image at infinity. To do so, we take the ratio of the angle θ_{image} subtended by the image to the angle θ_{object} subtended by the object at the near point of the eye (this is the closest that the unaided eye can view the object, and thus this is the position where the object will form the largest image on the retina of the unaided eye). Using [\[link\]](#) and working in the small-angle approximation, we have $\theta_{\text{image}} \approx h_i^{\text{obj}} / f^{\text{eye}}$ and $\theta_{\text{object}} \approx h_i^{\text{obj}} / 25 \text{ cm}$, where h_i^{obj} is the height of the image formed by the objective, which is the object of the eyepiece. Thus, the angular magnification of the eyepiece is

Equation:

$$M^{\text{eye}} = \frac{\theta_{\text{image}}}{\theta_{\text{object}}} = \frac{h_i^{\text{obj}}}{f^{\text{eye}}} \frac{25 \text{ cm}}{h_i^{\text{obj}}} = \frac{25 \text{ cm}}{f^{\text{eye}}}.$$

The net magnifying power of the compound microscope with the image at infinity is therefore

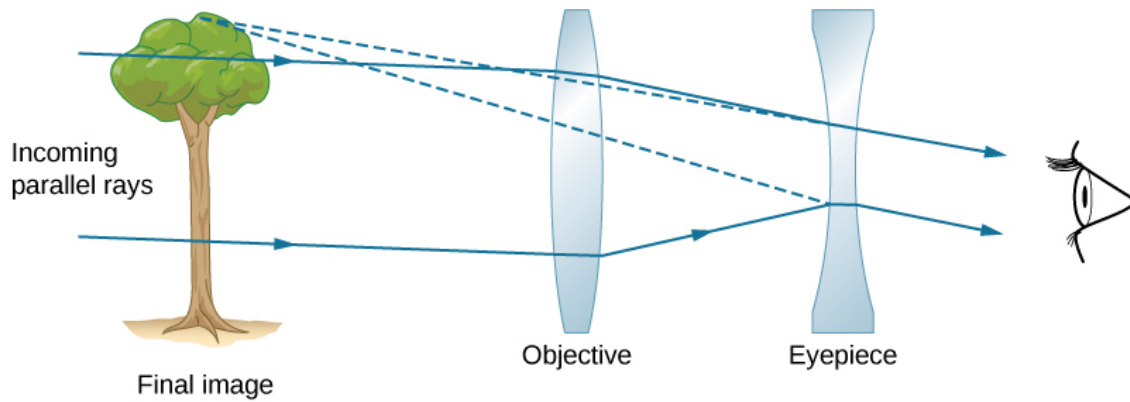
Equation:

$$M_{\text{net}} = m^{\text{obj}} M^{\text{eye}} = - \frac{(16 \text{ cm})(25 \text{ cm})}{f^{\text{obj}} f^{\text{eye}}}.$$

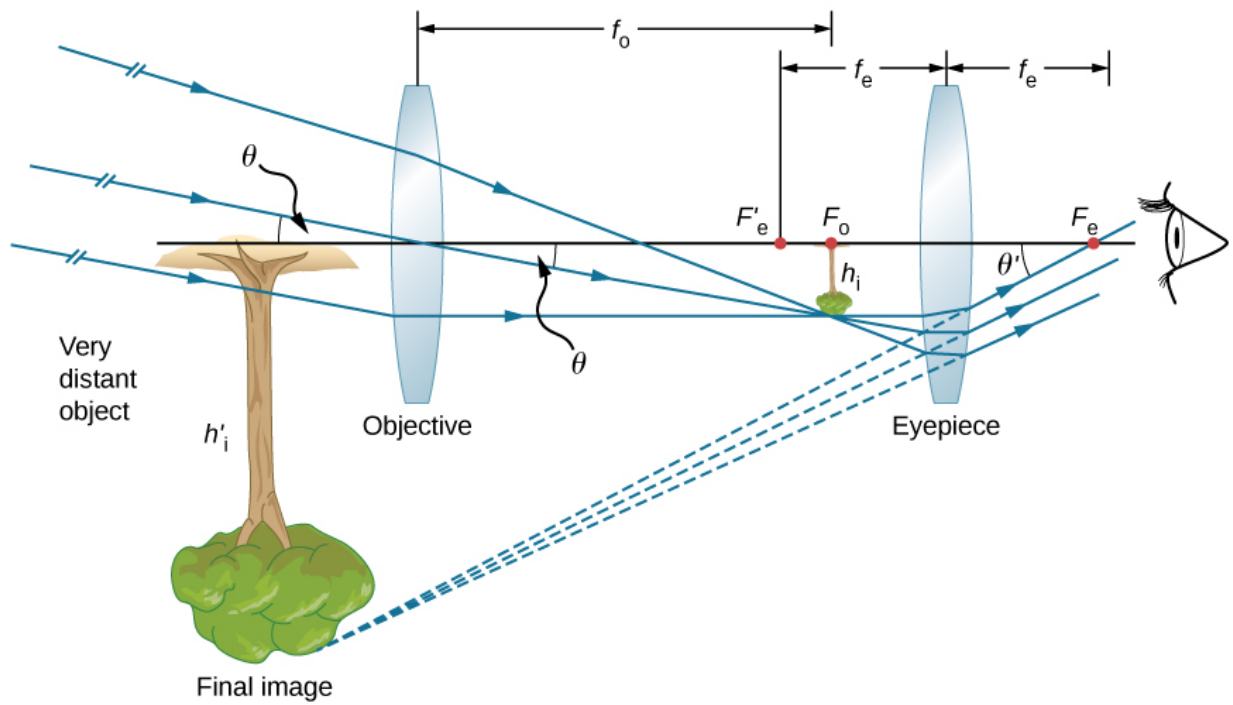
The focal distances must be in centimeters. The minus sign indicates that the final image is inverted. Note that the only variables in the equation are the focal distances of the eyepiece and the objective, which makes this equation particularly useful.

Telescopes

Telescopes are meant for viewing distant objects and produce an image that is larger than the image produced in the unaided eye. Telescopes gather far more light than the eye, allowing dim objects to be observed with greater magnification and better resolution. Telescopes were invented around 1600, and Galileo was the first to use them to study the heavens, with monumental consequences. He observed the moons of Jupiter, the craters and mountains on the moon, the details of sunspots, and the fact that the Milky Way is composed of a vast number of individual stars.



(a)



(b)

(a) Galileo made telescopes with a convex objective and a concave eyepiece. These produce an upright image and are used in spyglasses.

(b) Most simple refracting telescopes have two convex lenses. The objective forms a real, inverted image at (or just within) the focal plane of the eyepiece. This image serves as the object for the eyepiece. The eyepiece forms a virtual, inverted image that is magnified.

Part (a) of [\[link\]](#) shows a refracting telescope made of two lenses. The first lens, called the objective, forms a real image within the focal length of the second lens, which is called the eyepiece. The image of the objective lens serves as the object for the eyepiece, which forms a magnified virtual image that is observed by the eye. This design is what Galileo used to observe the heavens.

Although the arrangement of the lenses in a refracting telescope looks similar to that in a microscope, there are important differences. In a telescope, the real object is far away and the intermediate image is smaller than the object. In a microscope, the real object is very close and the intermediate image is larger than the object. In both the telescope and the microscope, the eyepiece magnifies the intermediate image; in the telescope, however, this is the only magnification.

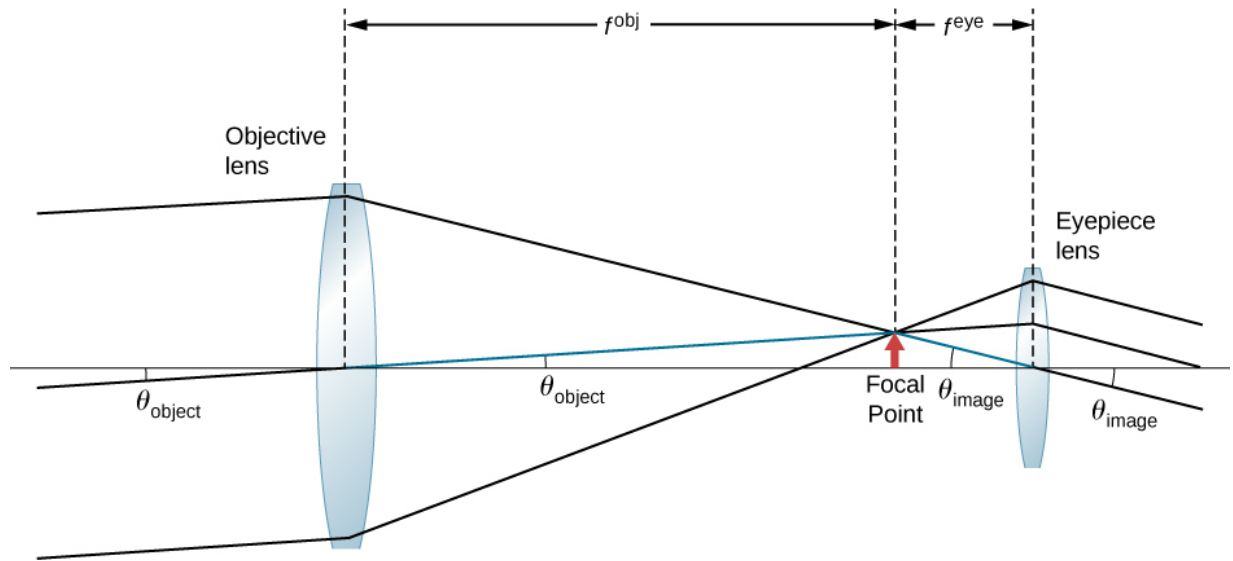
The most common two-lens telescope is shown in part (b) of the figure. The object is so far from the telescope that it is essentially at infinity compared with the focal lengths of the lenses ($d_o^{\text{obj}} \approx \infty$), so the incoming rays are essentially parallel and focus on the focal plane. Thus, the first image is produced at $d_i^{\text{obj}} = f^{\text{obj}}$, as shown in the figure, and is not large compared with what you might see by looking directly at the object. However, the eyepiece of the telescope eyepiece (like the microscope eyepiece) allows you to get nearer than your near point to this first image and so magnifies it (because you are near to it, it subtends a larger angle from your eye and so forms a larger image on your retina). As for a simple magnifier, the angular magnification of a telescope is the ratio of the angle subtended by the image [θ_{image} in part (b)] to the angle subtended by the real object [θ_{object} in part (b)]:

Equation:

$$M = \frac{\theta_{\text{image}}}{\theta_{\text{object}}}.$$

To obtain an expression for the magnification that involves only the lens parameters, note that the focal plane of the objective lens lies very close to

the focal plan of the eyepiece. If we assume that these planes are superposed, we have the situation shown in [\[link\]](#).



The focal plane of the objective lens of a telescope is very near to the focal plane of the eyepiece. The angle θ_{image} subtended by the image viewed through the eyepiece is larger than the angle θ_{object} subtended by the object when viewed with the unaided eye.

We further assume that the angles θ_{object} and θ_{image} are small, so that the small-angle approximation holds ($\tan \theta \approx \theta$). If the image formed at the focal plane has height h , then

Equation:

$$\theta_{object} \approx \tan \theta_{object} = \frac{h}{f_{obj}}$$

$$\theta_{image} \approx \tan \theta_{image} = \frac{-h}{f_{eye}}$$

where the minus sign is introduced because the height is negative if we measure both angles in the counterclockwise direction. Inserting these

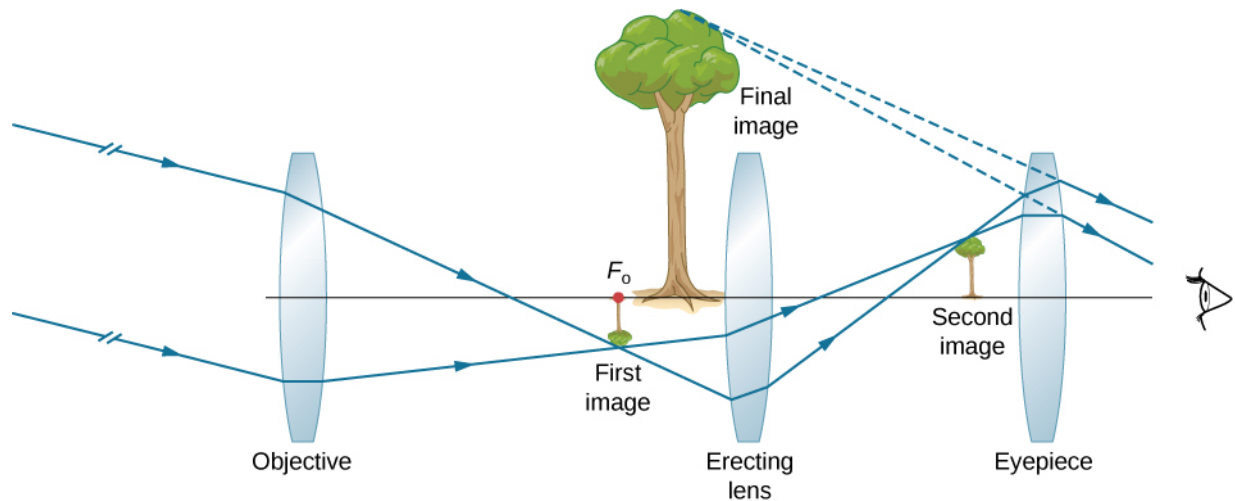
expressions into [\[link\]](#) gives

Equation:

$$M = \frac{-h_i}{f^{\text{eye}}} \frac{f^{\text{obj}}}{h_i} = -\frac{f^{\text{obj}}}{f^{\text{eye}}}.$$

Thus, to obtain the greatest angular magnification, it is best to have an objective with a long focal length and an eyepiece with a short focal length. The greater the angular magnification M , the larger an object will appear when viewed through a telescope, making more details visible. Limits to observable details are imposed by many factors, including lens quality and atmospheric disturbance. Typical eyepieces have focal lengths of 2.5 cm or 1.25 cm. If the objective of the telescope has a focal length of 1 meter, then these eyepieces result in magnifications of $40\times$ and $80\times$, respectively. Thus, the angular magnifications make the image appear 40 times or 80 times closer than the real object.

The minus sign in the magnification indicates the image is inverted, which is unimportant for observing the stars but is a real problem for other applications, such as telescopes on ships or telescopic gun sights. If an upright image is needed, Galileo's arrangement in part (a) of [\[link\]](#) can be used. But a more common arrangement is to use a third convex lens as an eyepiece, increasing the distance between the first two and inverting the image once again, as seen in [\[link\]](#).



This arrangement of three lenses in a telescope produces an upright final image. The first two lenses are far enough apart that the second lens inverts the image of the first. The third lens acts as a magnifier and keeps the image upright and in a location that is easy to view.

The largest refracting telescope in the world is the 40-inch diameter Yerkes telescope located at Lake Geneva, Wisconsin ([\[link\]](#)), and operated by the University of Chicago.

It is very difficult and expensive to build large refracting telescopes. You need large defect-free lenses, which in itself is a technically demanding task. A refracting telescope basically looks like a tube with a support structure to rotate it in different directions. A refracting telescope suffers from several problems. The aberration of lenses causes the image to be blurred. Also, as the lenses become thicker for larger lenses, more light is absorbed, making faint stars more difficult to observe. Large lenses are also very heavy and deform under their own weight. Some of these problems with refracting telescopes are addressed by avoiding refraction for collecting light and instead using a curved mirror in its place, as devised by Isaac Newton. These telescopes are called reflecting telescopes.



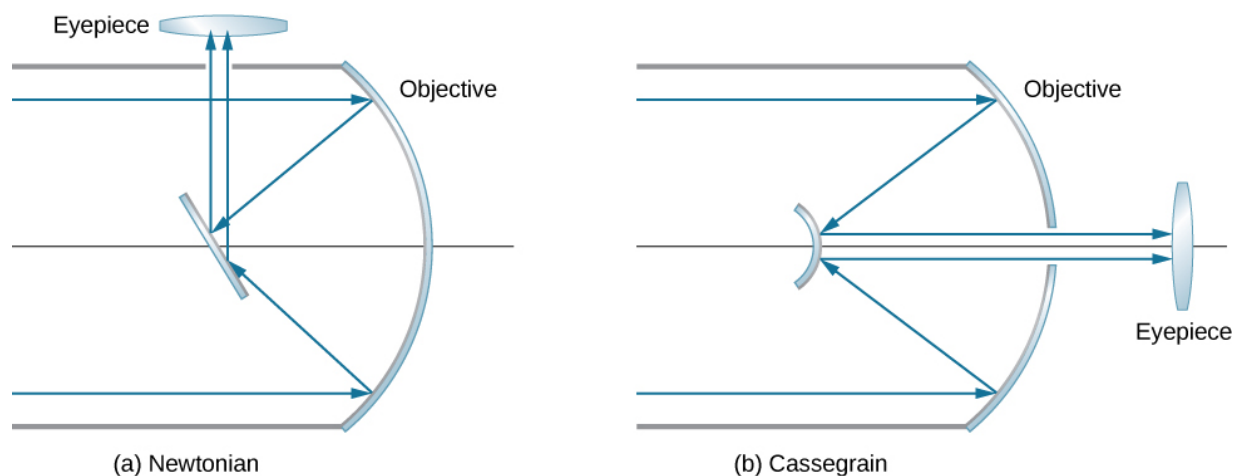
In 1897, the Yerkes Observatory in Wisconsin (USA) built a large refracting telescope with an objective lens that is 40 inches in diameter and has a tube length of 62 feet. (credit: Yerkes Observatory, University of Chicago)

Reflecting Telescopes

Isaac Newton designed the first reflecting telescope around 1670 to solve the problem of chromatic aberration that happens in all refracting telescopes. In chromatic aberration, light of different colors refracts by slightly different amounts in the lens. As a result, a rainbow appears around the image and the image appears blurred. In the reflecting telescope, light rays from a distant source fall upon the surface of a concave mirror fixed at the bottom end of the tube. The use of a mirror instead of a lens eliminates chromatic aberration. The concave mirror focuses the rays on its focal plane. The design problem is how to observe the focused image. Newton used a design in which the focused light from the concave mirror was

reflected to one side of the tube into an eyepiece [part (a) of [\[link\]](#)]. This arrangement is common in many amateur telescopes and is called the **Newtonian design**.

Some telescopes reflect the light back toward the middle of the concave mirror using a convex mirror. In this arrangement, the light-gathering concave mirror has a hole in the middle [part (b) of the figure]. The light then is incident on an eyepiece lens. This arrangement of the objective and eyepiece is called the **Cassegrain design**. Most big telescopes, including the Hubble space telescope, are of this design. Other arrangements are also possible. In some telescopes, a light detector is placed right at the spot where light is focused by the curved mirror.



Reflecting telescopes: (a) In the Newtonian design, the eyepiece is located at the side of the telescope; (b) in the Cassegrain design, the eyepiece is located past a hole in the primary mirror.

Most astronomical research telescopes are now of the reflecting type. One of the earliest large telescopes of this kind is the Hale 200-inch (or 5-meter) telescope built on Mount Palomar in southern California, which has a 200 inch-diameter mirror. One of the largest telescopes in the world is the 10-meter Keck telescope at the Keck Observatory on the summit of the

dormant Mauna Kea volcano in Hawaii. The Keck Observatory operates two 10-meter telescopes. Each is not a single mirror, but is instead made up of 36 hexagonal mirrors. Furthermore, the two telescopes on the Keck can work together, which increases their power to an effective 85-meter mirror. The Hubble telescope ([\[link\]](#)) is another large reflecting telescope with a 2.4 meter-diameter primary mirror. The Hubble was put into orbit around Earth in 1990.



The Hubble space telescope as seen from the Space Shuttle Discovery.
(credit: modification of work by NASA)

The angular magnification M of a reflecting telescope is also given by [\[link\]](#). For a spherical mirror, the focal length is half the radius of curvature,

so making a large objective mirror not only helps the telescope collect more light but also increases the magnification of the image.

Summary

- Many optical devices contain more than a single lens or mirror. These are analyzed by considering each element sequentially. The image formed by the first is the object for the second, and so on. The same ray-tracing and thin-lens techniques developed in the previous sections apply to each lens element.
- The overall magnification of a multiple-element system is the product of the linear magnifications of its individual elements times the angular magnification of the eyepiece. For a two-element system with an objective and an eyepiece, this is

Equation:

$$M = m^{\text{obj}} M^{\text{eye}}.$$

where m^{obj} is the linear magnification of the objective and M^{eye} is the angular magnification of the eyepiece.

- The microscope is a multiple-element system that contains more than a single lens or mirror. It allows us to see detail that we could not see with the unaided eye. Both the eyepiece and objective contribute to the magnification. The magnification of a compound microscope with the image at infinity is

Equation:

$$M_{\text{net}} = - \frac{(16 \text{ cm})(25 \text{ cm})}{f^{\text{obj}} f^{\text{eye}}}.$$

In this equation, 16 cm is the standardized distance between the image-side focal point of the objective lens and the object-side focal point of the eyepiece, 25 cm is the normal near point distance, f^{obj} and f^{eye} are the focal distances for the objective lens and the eyepiece, respectively.

- Simple telescopes can be made with two lenses. They are used for viewing objects at large distances.

- The angular magnification M for a telescope is given by
Equation:

$$M = -\frac{f^{\text{obj}}}{f^{\text{eye}}},$$

where f^{obj} and f^{eye} are the focal lengths of the objective lens and the eyepiece, respectively.

Key Equations

Image distance in a plane mirror	$d_o = -d_i$
Focal length for a spherical mirror	$f = \frac{R}{2}$
Mirror equation	$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$
Magnification of a spherical mirror	$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}$
Sign convention for mirrors	
Focal length f	+ for concave mirror – for convex mirror
Object distance d_o	+ for real object – for virtual object

Image distance d_i	+ for real image – for virtual image
Magnification m	+ for upright image – for inverted image
Apparent depth equation	$h_i = \left(\frac{n_2}{n_1} \right) h_o$
Spherical interface equation	$\frac{n_1}{d_o} + \frac{n_2}{d_i} = \frac{n_2 - n_1}{R}$
The thin-lens equation	$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$
The lens maker's equation	$\frac{1}{f} = \left(\frac{n_2}{n_1} - 1 \right) \left(\frac{1}{R_1} - \frac{1}{R_2} \right)$
The magnification m of an object	$m \equiv \frac{h_i}{h_o} = -\frac{d_i}{d_o}$
Optical power	$P = \frac{1}{f}$
Optical power of thin, closely spaced lenses	$P_{\text{total}} = P_{\text{lens1}} + P_{\text{lens2}} + P_{\text{lens3}} + \cdots$
Angular magnification M of a simple magnifier	$M = \frac{\theta_{\text{image}}}{\theta_{\text{object}}}$
Angular magnification of an object a distance L from the eye for a convex lens of focal length f held a distance ℓ from the eye	$M = \left(\frac{25 \text{ cm}}{L} \right) \left(1 + \frac{L - \ell}{f} \right)$

Range of angular magnification for a given lens for a person with a near point of 25 cm	$\frac{25 \text{ cm}}{f} \leq M \leq 1 + \frac{25 \text{ cm}}{f}$
Net magnification of compound microscope	$M_{\text{net}} = m^{\text{obj}} M^{\text{eye}} = - \frac{d_i^{\text{obj}} (f^{\text{eye}} + 25 \text{ cm})}{f^{\text{obj}} f^{\text{eye}}}$

Conceptual Questions

Exercise:

Problem:

Geometric optics describes the interaction of light with macroscopic objects. Why, then, is it correct to use geometric optics to analyze a microscope's image?

Solution:

Microscopes create images of macroscopic size, so geometric optics applies.

Exercise:

Problem:

The image produced by the microscope in [\[link\]](#) cannot be projected. Could extra lenses or mirrors project it? Explain.

Exercise:

Problem:

If you want your microscope or telescope to project a real image onto a screen, how would you change the placement of the eyepiece relative to the objective?

Solution:

The eyepiece would be moved slightly farther from the objective so that the image formed by the objective falls just beyond the focal length of the eyepiece.

Problems

Exercise:

Problem:

A microscope with an overall magnification of 800 has an objective that magnifies by 200. (a) What is the angular magnification of the eyepiece? (b) If there are two other objectives that can be used, having magnifications of 100 and 400, what other total magnifications are possible?

Exercise:

Problem:

(a) What magnification is produced by a 0.150 cm-focal length microscope objective that is 0.155 cm from the object being viewed? (b) What is the overall magnification if an $8\times$ eyepiece (one that produces an angular magnification of 8.00) is used?

Solution:

$$\text{a. } \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \Rightarrow d_i = 4.65 \text{ cm};$$
$$\Rightarrow m = -30.0$$

$$\text{b. } M_{\text{net}} = -240$$

Exercise:

Problem:

Where does an object need to be placed relative to a microscope for its 0.50 cm-focal length objective to produce a magnification of -400 ?

Exercise:

Problem:

An amoeba is 0.305 cm away from the 0.300 cm-focal length objective lens of a microscope. (a) Where is the image formed by the objective lens? (b) What is this image's magnification? (c) An eyepiece with a 2.00-cm focal length is placed 20.0 cm from the objective. Where is the final image? (d) What angular magnification is produced by the eyepiece? (e) What is the overall magnification? (See [\[link\]](#).)

Solution:

- a. $\frac{1}{d_o^{\text{obj}}} + \frac{1}{d_i^{\text{obj}}} = \frac{1}{f^{\text{obj}}} \Rightarrow d_i^{\text{obj}} = 18.3 \text{ cm}$ behind the objective lens;
b. $m^{\text{obj}} = -60.0$;
 $d_o^{\text{eye}} = 1.70 \text{ cm}$
c. $d_i^{\text{eye}} = -11.3 \text{ cm}$
in front of the eyepiece; d. $M^{\text{eye}} = 13.5$;
e. $M_{\text{net}} = -810$

Exercise:**Problem:**

Unreasonable Results Your friends show you an image through a microscope. They tell you that the microscope has an objective with a 0.500-cm focal length and an eyepiece with a 5.00-cm focal length. The resulting overall magnification is 250,000. Are these viable values for a microscope?

Unless otherwise stated, the lens-to-retina distance is 2.00 cm.

Exercise:**Problem:**

What is the angular magnification of a telescope that has a 100 cm-focal length objective and a 2.50 cm-focal length eyepiece?

Solution:

$$M = -40.0$$

Exercise:**Problem:**

Find the distance between the objective and eyepiece lenses in the telescope in the above problem needed to produce a final image very far from the observer, where vision is most relaxed. Note that a telescope is normally used to view very distant objects.

Exercise:**Problem:**

A large reflecting telescope has an objective mirror with a 10.0-m radius of curvature. What angular magnification does it produce when a 3.00 m-focal length eyepiece is used?

Solution:

$$f^{\text{obj}} = \frac{R}{2}, M = -1.67$$

Exercise:**Problem:**

A small telescope has a concave mirror with a 2.00-m radius of curvature for its objective. Its eyepiece is a 4.00 cm-focal length lens. (a) What is the telescope's angular magnification? (b) What angle is subtended by a 25,000 km-diameter sunspot? (c) What is the angle of its telescopic image?

Exercise:**Problem:**

A $7.5 \times$ binocular produces an angular magnification of -7.50 , acting like a telescope. (Mirrors are used to make the image upright.) If the binoculars have objective lenses with a 75.0-cm focal length, what is the focal length of the eyepiece lenses?

Solution:

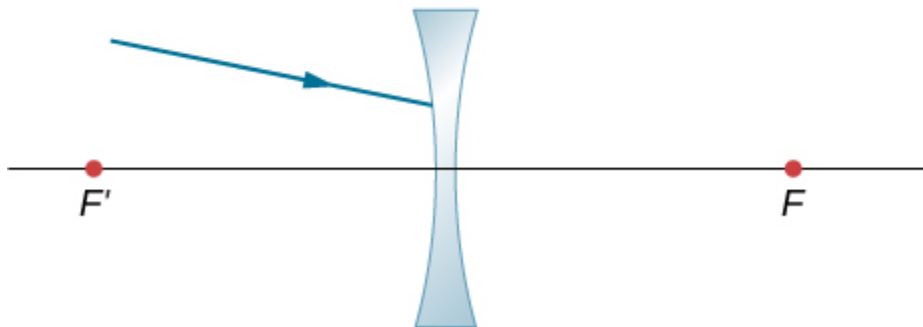
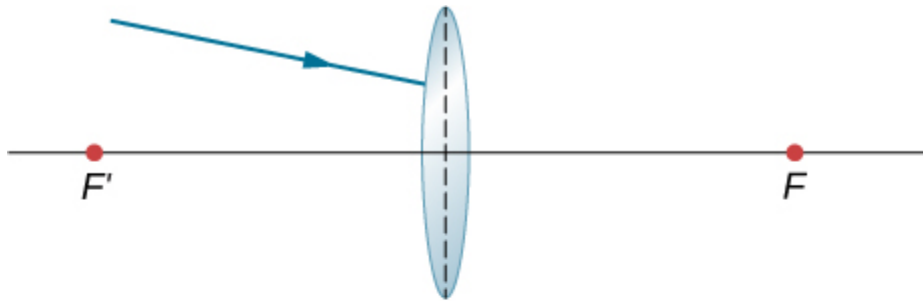
$$M = -\frac{f^{\text{obj}}}{f^{\text{eye}}}, f^{\text{eye}} = +10.0 \text{ cm}$$

Exercise:**Problem:**

Construct Your Own Problem Consider a telescope of the type used by Galileo, having a convex objective and a concave eyepiece as illustrated in part (a) of [\[link\]](#). Construct a problem in which you calculate the location and size of the image produced. Among the things to be considered are the focal lengths of the lenses and their relative placements as well as the size and location of the object. Verify that the angular magnification is greater than one. That is, the angle subtended at the eye by the image is greater than the angle subtended by the object.

Exercise:**Problem:**

Trace rays to find which way the given ray will emerge after refraction through the thin lens in the following figure. Assume thin-lens approximation. (*Hint:* Pick a point P on the given ray in each case. Treat that point as an object. Now, find its image Q . Use the rule: All rays on the other side of the lens will either go through Q or appear to be coming from Q .)



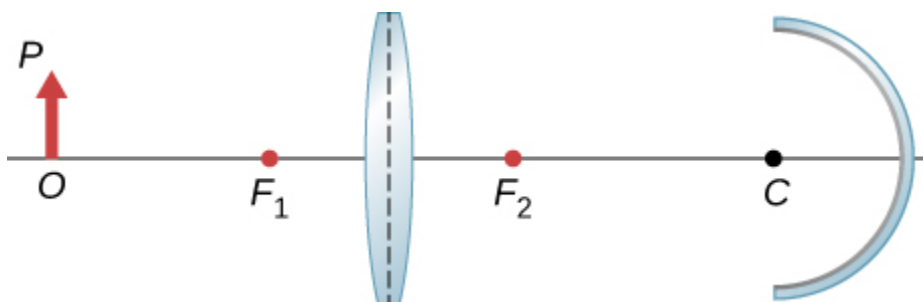
Solution:

Answers will vary.

Exercise:

Problem:

Copy and draw rays to find the final image in the following diagram.
(Hint: Find the intermediate image through lens alone. Use the intermediate image as the object for the mirror and work with the mirror alone to find the final image.)



Exercise:**Problem:**

A concave mirror of radius of curvature 10 cm is placed 30 cm from a thin convex lens of focal length 15 cm. Find the location and magnification of a small bulb sitting 50 cm from the lens by using the algebraic method.

Solution:

12 cm to the left of the mirror, $m = 3/5$

Exercise:**Problem:**

An object of height 3 cm is placed at 25 cm in front of a converging lens of focal length 20 cm. Behind the lens there is a concave mirror of focal length 20 cm. The distance between the lens and the mirror is 5 cm. Find the location, orientation and size of the final image.

Exercise:**Problem:**

An object of height 3 cm is placed at a distance of 25 cm in front of a converging lens of focal length 20 cm, to be referred to as the first lens. Behind the lens there is another converging lens of focal length 20 cm placed 10 cm from the first lens. There is a concave mirror of focal length 15 cm placed 50 cm from the second lens. Find the location, orientation, and size of the final image.

Solution:

27 cm in front of the mirror, $m = 0.6$, $h_i = 1.76$ cm, orientation upright

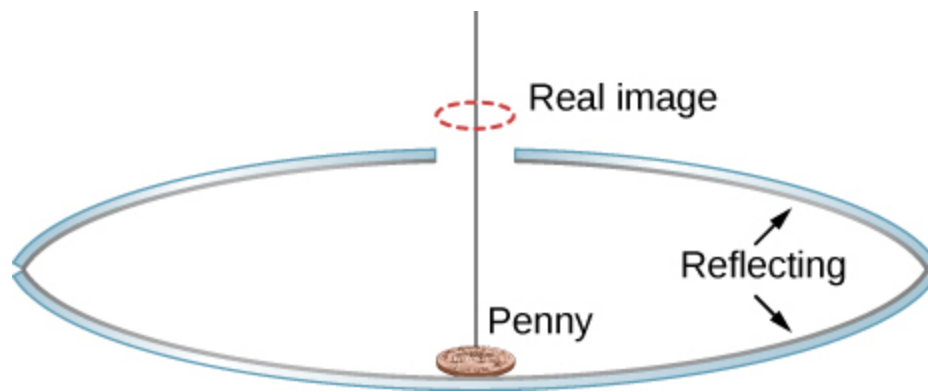
Exercise:

Problem:

An object of height 2 cm is placed at 50 cm in front of a converging lens of focal length 40 cm. Behind the lens, there is a convex mirror of focal length 15 cm placed 30 cm from the converging lens. Find the location, orientation, and size of the final image.

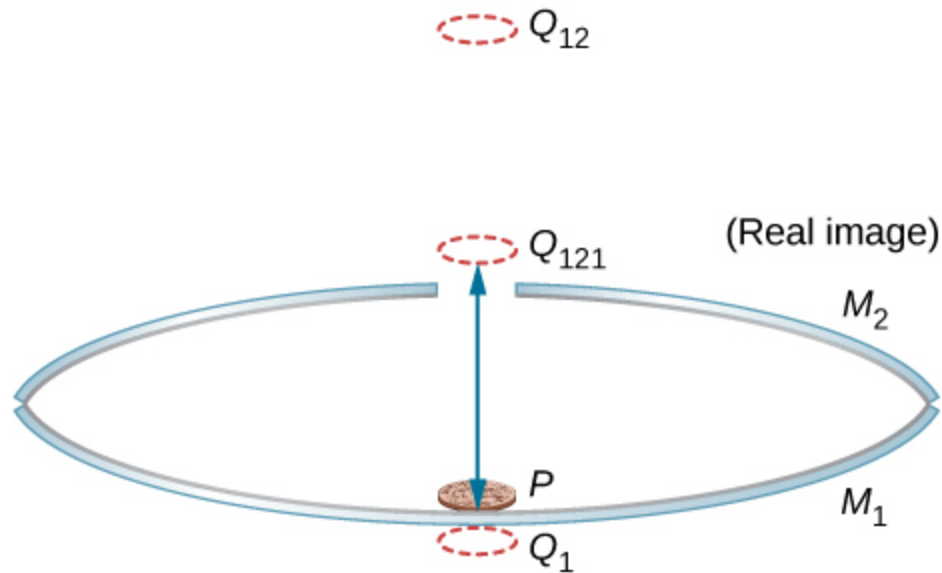
Exercise:**Problem:**

Two concave mirrors are placed facing each other. One of them has a small hole in the middle. A penny is placed on the bottom mirror (see the following figure). When you look from the side, a real image of the penny is observed above the hole. Explain how that could happen.



Solution:

The following figure shows three successive images beginning with the image Q_1 in mirror M_1 . Q_1 is the image in mirror M_1 , whose image in mirror M_2 is Q_{12} whose image in mirror M_1 is the real image Q_{121} .



Exercise:

Problem:

A lamp of height 5 cm is placed 40 cm in front of a converging lens of focal length 20 cm. There is a plane mirror 15 cm behind the lens. Where would you find the image when you look in the mirror?

Exercise:

Problem:

Parallel rays from a faraway source strike a converging lens of focal length 20 cm at an angle of 15 degrees with the horizontal direction. Find the vertical position of the real image observed on a screen in the focal plane.

Solution:

5.4 cm from the axis

Exercise:

Problem:

Parallel rays from a faraway source strike a diverging lens of focal length 20 cm at an angle of 10 degrees with the horizontal direction. As you look through the lens, where in the vertical plane the image would appear?

Exercise:**Problem:**

A light bulb is placed 10 cm from a plane mirror, which faces a convex mirror of radius of curvature 8 cm. The plane mirror is located at a distance of 30 cm from the vertex of the convex mirror. Find the location of two images in the convex mirror. Are there other images? If so, where are they located?

Solution:

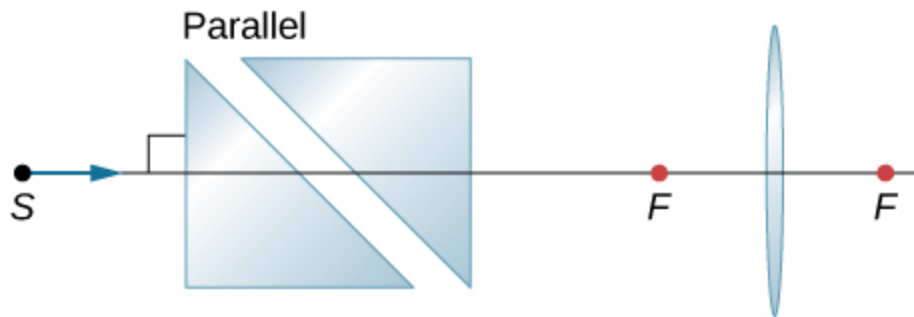
Let the vertex of the concave mirror be the origin of the coordinate system. Image 1 is at $-10/3$ cm (-3.3 cm), image 2 is at $-40/11$ cm (-3.6 cm). These serve as objects for subsequent images, which are at $-310/83$ cm (-3.7 cm), $-9340/2501$ cm (-3.7 cm), $-140,720/37,681$ cm (-3.7 cm). All remaining images are at approximately -3.7 cm.

Exercise:**Problem:**

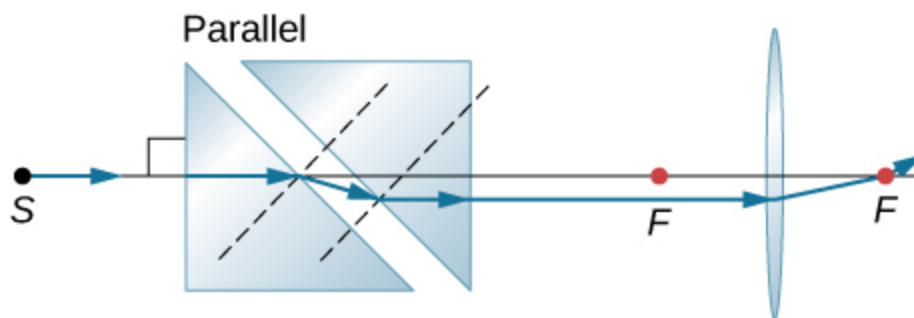
A point source of light is 50 cm in front of a converging lens of focal length 30 cm. A concave mirror with a focal length of 20 cm is placed 25 cm behind the lens. Where does the final image form, and what are its orientation and magnification?

Exercise:**Problem:**

Copy and trace to find how a horizontal ray from S comes out after the lens. Use $n_{\text{glass}} = 1.5$ for the prism material.



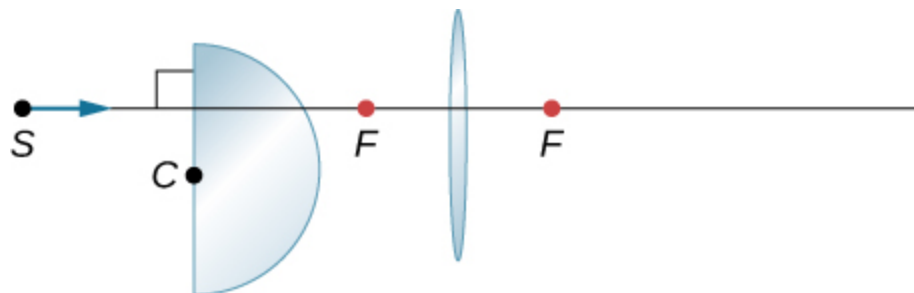
Solution:



Exercise:

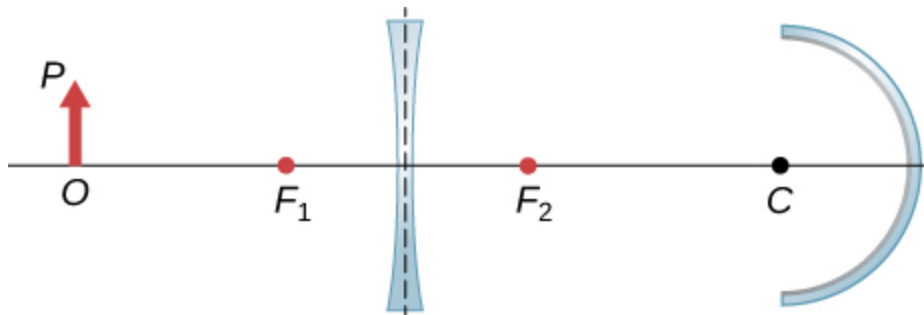
Problem:

Copy and trace how a horizontal ray from S comes out after the lens.
Use $n = 1.55$ for the glass.

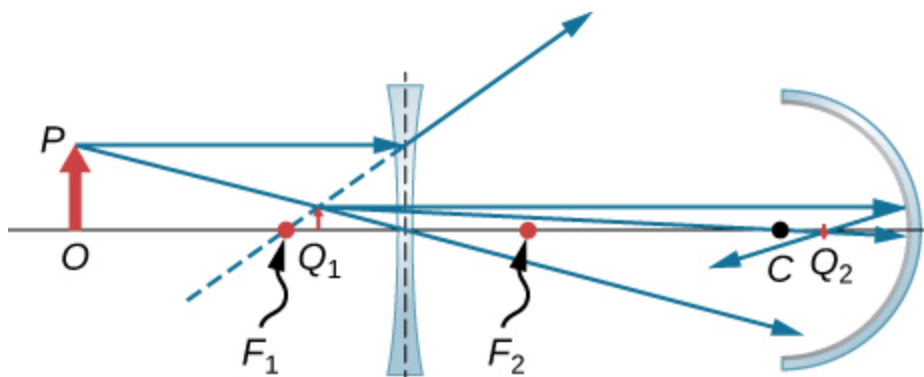


Exercise:

Problem: Copy and draw rays to figure out the final image.



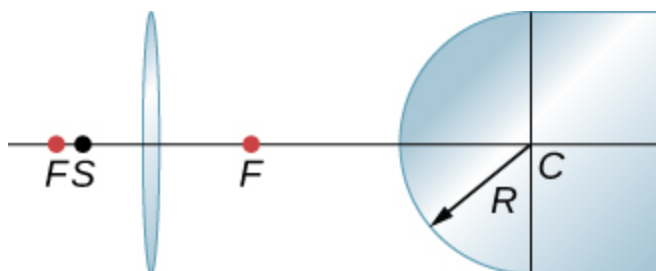
Solution:



Exercise:

Problem:

By ray tracing or by calculation, find the place inside the glass where rays from S converge as a result of refraction through the lens and the convex air-glass interface. Use a ruler to estimate the radius of curvature.



Exercise:

Problem:

A diverging lens has a focal length of 20 cm. What is the power of the lens in diopters?

Solution:

-5 D

Exercise:**Problem:**

Two lenses of focal lengths of f_1 and f_2 are glued together with transparent material of negligible thickness. Show that the total power of the two lenses simply add.

Exercise:**Problem:**

What will be the angular magnification of a convex lens with the focal length 2.5 cm?

Solution:

11

Exercise:**Problem:**

What will be the formula for the angular magnification of a convex lens of focal length f if the eye is very close to the lens and the near point is located a distance D from the eye?

Additional Problems**Exercise:**

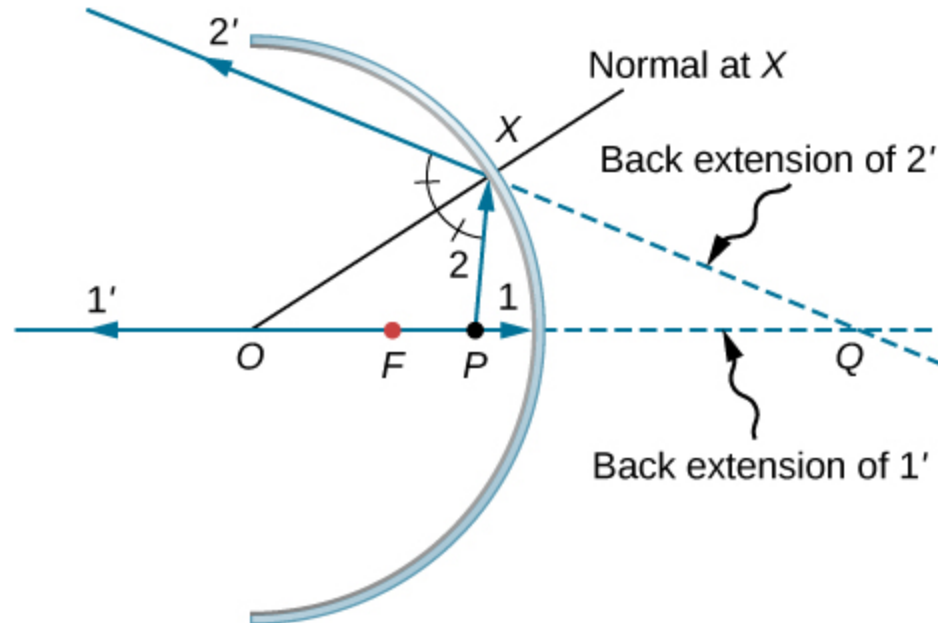
Problem:

Use a ruler and a protractor to draw rays to find images in the following cases.

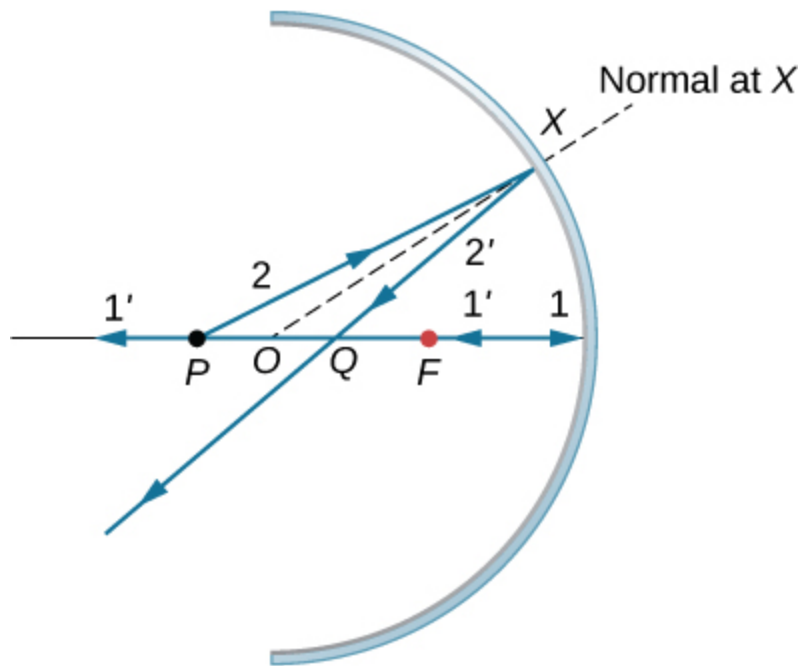
- (a) A point object located on the axis of a concave mirror located at a point within the focal length from the vertex.
- (b) A point object located on the axis of a concave mirror located at a point farther than the focal length from the vertex.
- (c) A point object located on the axis of a convex mirror located at a point within the focal length from the vertex.
- (d) A point object located on the axis of a convex mirror located at a point farther than the focal length from the vertex.
- (e) Repeat (a)–(d) for a point object off the axis.

Solution:

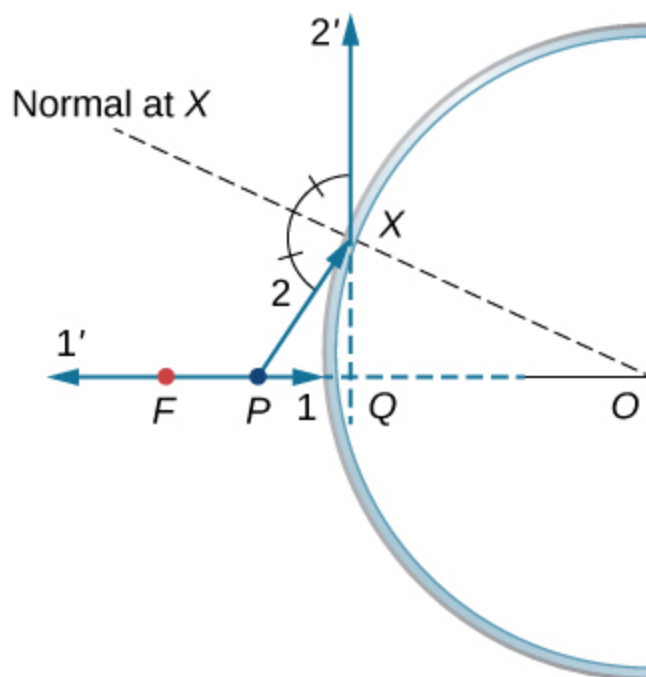
a.



b.

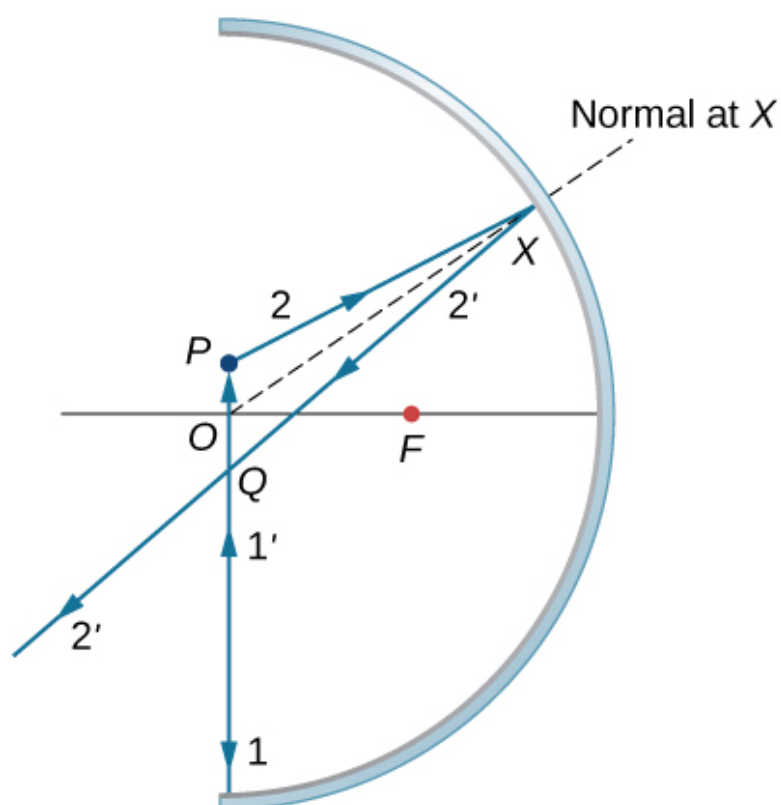
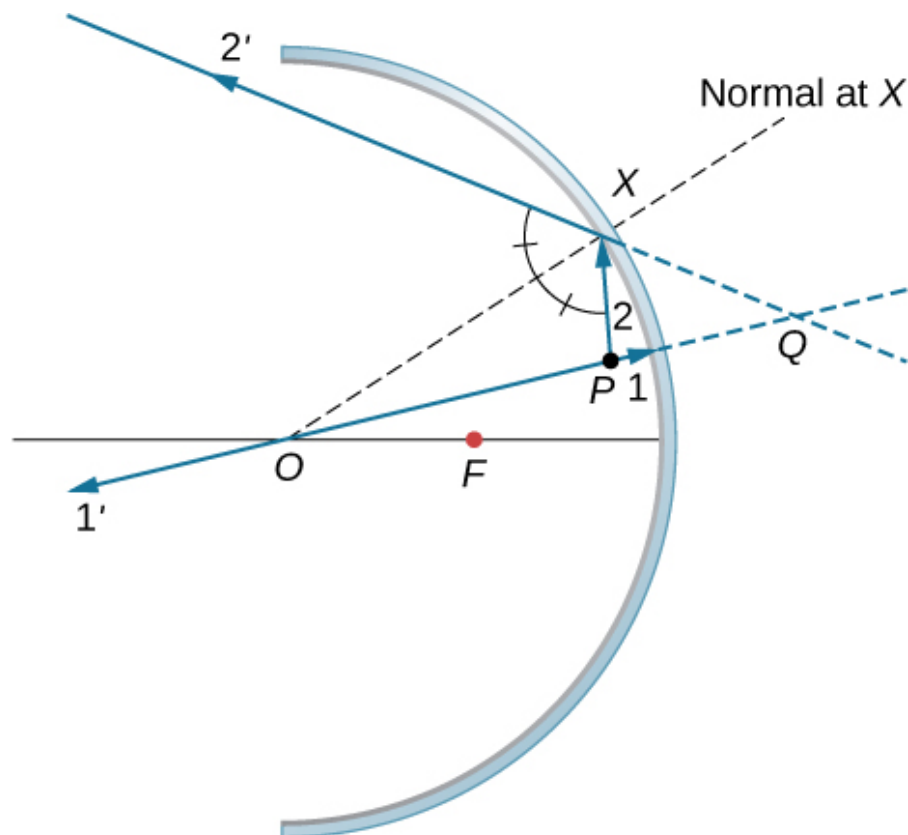


c.



d. similar to the previous picture but with point P outside the focal length; e. Repeat (a)–(d) for a point object off the axis. For a point

object placed off axis in front of a concave mirror corresponding to parts (a) and (b), the case for convex mirror left as exercises.



Exercise:**Problem:**

Where should a 3 cm tall object be placed in front of a concave mirror of radius 20 cm so that its image is real and 2 cm tall?

Exercise:**Problem:**

A 3 cm tall object is placed 5 cm in front of a convex mirror of radius of curvature 20 cm. Where is the image formed? How tall is the image? What is the orientation of the image?

Solution:

$d_i = -10/3$ cm, $h_i = 2$ cm, upright

Exercise:**Problem:**

You are looking for a mirror so that you can see a four-fold magnified virtual image of an object when the object is placed 5 cm from the vertex of the mirror. What kind of mirror you will need? What should be the radius of curvature of the mirror?

Exercise:

Problem: Derive the following equation for a convex mirror:

$$\frac{1}{VO} - \frac{1}{VI} = -\frac{1}{VF},$$

where VO is the distance to the object O from vertex V , VI the distance to the image I from V , and VF is the distance to the focal point F from V . (*Hint: use two sets of similar triangles.*)

Solution:

proof

Exercise:

Problem:

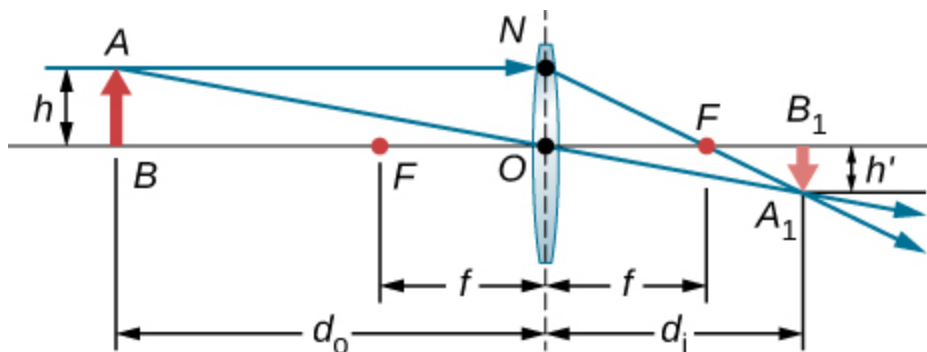
(a) Draw rays to form the image of a vertical object on the optical axis and farther than the focal point from a converging lens. (b) Use plane geometry in your figure and prove that the magnification m is given by $m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}$.

Exercise:

Problem:

Use another ray-tracing diagram for the same situation as given in the previous problem to derive the thin-lens equation, $\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$.

Solution:



Triangles BAO and B_1A_1O are similar triangles. Thus, $\frac{A_1B_1}{AB} = \frac{d_i}{d_o}$.
Triangles NOF and B_1A_1F are similar triangles. Thus, $\frac{NO}{f} = \frac{A_1B_1}{d_i - f}$.
Noting that $NO = AB$ gives $\frac{AB}{f} = \frac{A_1B_1}{d_i - f}$ or $\frac{AB}{A_1B_1} = \frac{f}{d_i - f}$.
Inverting this gives $\frac{A_1B_1}{AB} = \frac{d_i - f}{f}$. Equating the two expressions for the ratio $\frac{A_1B_1}{AB}$ gives $\frac{d_i}{d_o} = \frac{d_i - f}{f}$. Dividing through by d_i gives $\frac{1}{d_o} = \frac{1}{f} - \frac{1}{d_i}$ or $\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$.

Exercise:

Problem:

You photograph a 2.0-m-tall person with a camera that has a 5.0 cm-focal length lens. The image on the film must be no more than 2.0 cm high. (a) What is the closest distance the person can stand to the lens? (b) For this distance, what should be the distance from the lens to the film?

Exercise:**Problem:**

Find the focal length of a thin plano-convex lens. The front surface of this lens is flat, and the rear surface has a radius of curvature of $R_2 = -35$ cm. Assume that the index of refraction of the lens is 1.5.

Solution:

70 cm

Exercise:**Problem:**

Find the focal length of a meniscus lens with $R_1 = 20$ cm and $R_2 = 15$ cm. Assume that the index of refraction of the lens is 1.5.

Exercise:**Problem:**

A nearsighted man cannot see objects clearly beyond 20 cm from his eyes. How close must he stand to a mirror in order to see what he is doing when he shaves?

Solution:

The plane mirror has an infinite focal point, so that $d_i = -d_o$. The total apparent distance of the man in the mirror will be his actual distance, plus the apparent image distance, or $d_o + (-d_i) = 2d_o$. If this distance must be less than 20 cm, he should stand at $d_o = 10$ cm.

Exercise:**Problem:**

A mother sees that her child's contact lens prescription is 0.750 D. What is the child's near point?

Exercise:**Problem:**

Repeat the previous problem for glasses that are 2.20 cm from the eyes.

Solution:

Here we want $d_o = 25 \text{ cm} - 2.20 \text{ cm} = 0.228 \text{ m}$. If $x =$ near point, $d_i = -(x - 0.0220 \text{ m})$. Thus, $P = \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{0.228 \text{ m}} + \frac{1}{x - 0.0220 \text{ m}}$. Using $P = 0.75 \text{ D}$ gives $x = 0.253 \text{ m}$, so the near point is 25.3 cm.

Exercise:**Problem:**

The contact-lens prescription for a nearsighted person is -4.00 D and the person has a far point of 22.5 cm. What is the power of the tear layer between the cornea and the lens if the correction is ideal, taking the tear layer into account?

Exercise:**Problem:**

Unreasonable Results A boy has a near point of 50 cm and a far point of 500 cm. Will a -4.00 D lens correct his far point to infinity?

Solution:

Assuming a lens at 2.00 cm from the boy's eye, the image distance must be $d_i = -(500 \text{ cm} - 2.00 \text{ cm}) = -498 \text{ cm}$. For an infinite-

distance object, the required power is $P = \frac{1}{d_i} = -0.200 \text{ D}$.

Therefore, the -4.00 D lens will correct the nearsightedness.

Exercise:

Problem:

Find the angular magnification of an image by a magnifying glass of $f = 5.0 \text{ cm}$ if the object is placed $d_o = 4.0 \text{ cm}$ from the lens and the lens is close to the eye.

Exercise:

Problem:

Let objective and eyepiece of a compound microscope have focal lengths of 2.5 cm and 10 cm , respectively and be separated by 12 cm . A $70\text{-}\mu\text{m}$ object is placed 6.0 cm from the objective. How large is the virtual image formed by the objective-eyepiece system?

Solution:

$87 \mu\text{m}$

Exercise:

Problem:

Draw rays to scale to locate the image at the retina if the eye lens has a focal length 2.5 cm and the near point is 24 cm . (*Hint:* Place an object at the near point.)

Exercise:

Problem:

The objective and the eyepiece of a microscope have the focal lengths 3 cm and 10 cm respectively. Decide about the distance between the objective and the eyepiece if we need a $10 \times$ magnification from the objective/eyepiece compound system.

Solution:

Use, $M_{\text{net}} = -\frac{d_i^{\text{obj}}(f^{\text{eye}} + 25 \text{ cm})}{f^{\text{obj}} f^{\text{eye}}}$. The image distance for the objective is

$$d_i^{\text{obj}} = -\frac{M_{\text{net}} f^{\text{obj}} f^{\text{eye}}}{f^{\text{eye}} + 25 \text{ cm}}.$$
 Using

$$f^{\text{obj}} = 3.0 \text{ cm}, f^{\text{eye}} = 10 \text{ cm}, \text{ and } M = -10 \text{ gives } d_i^{\text{obj}} = 8.6 \text{ cm}.$$

We want this image to be at the focal point of the eyepiece so that the eyepiece forms an image at infinity for comfortable viewing. Thus, the distance d between the lenses should be

$$d = f^{\text{eye}} + d_i^{\text{obj}} = 10 \text{ cm} + 8.6 \text{ cm} = 19 \text{ cm}.$$

Exercise:

Problem:

A far-sighted person has a near point of 100 cm. How far in front or behind the retina does the image of an object placed 25 cm from the eye form? Use the cornea to retina distance of 2.5 cm.

Exercise:

Problem:

A near-sighted person has a far point of 80 cm. (a) What kind of corrective lens the person will need if the lens is to be placed 1.5 cm from the eye? (b) What would be the power of the contact lens needed? Assume distance to contact lens from the eye to be zero.

Solution:

a. focal length of the corrective lens $f_c = -80 \text{ cm}$; b. -1.25 D

Exercise:

Problem:

In a reflecting telescope the objective is a concave mirror of radius of curvature 2 m and an eyepiece is a convex lens of focal length 5 cm. Find the apparent size of a 25-m tree at a distance of 10 km that you would perceive when looking through the telescope.

Exercise:

Problem:

Two stars that are 10^9 km apart are viewed by a telescope and found to be separated by an angle of 10^{-5} radians. If the eyepiece of the telescope has a focal length of 1.5 cm and the objective has a focal length of 3 meters, how far away are the stars from the observer?

Solution:

$$2 \times 10^{16} \text{ km}$$

Exercise:**Problem:**

What is the angular size of the Moon if viewed from a binocular that has a focal length of 1.2 cm for the eyepiece and a focal length of 8 cm for the objective? Use the radius of the moon 1.74×10^6 m and the distance of the moon from the observer to be 3.8×10^8 m.

Exercise:**Problem:**

An unknown planet at a distance of 10^{12} m from Earth is observed by a telescope that has a focal length of the eyepiece of 1 cm and a focal length of the objective of 1 m. If the far away planet is seen to subtend an angle of 10^{-5} radian at the eyepiece, what is the size of the planet?

Solution:

$$10^5 \text{ m}$$

Glossary**Cassegrain design**

arrangement of an objective and eyepiece such that the light-gathering concave mirror has a hole in the middle, and light then is incident on

an eyepiece lens

compound microscope

microscope constructed from two convex lenses, the first serving as the eyepiece and the second serving as the objective lens

eyepiece

lens or combination of lenses in an optical instrument nearest to the eye of the observer

net magnification

(M_{net}) of the compound microscope is the product of the linear magnification of the objective and the angular magnification of the eyepiece

Newtonian design

arrangement of an objective and eyepiece such that the focused light from the concave mirror was reflected to one side of the tube into an eyepiece

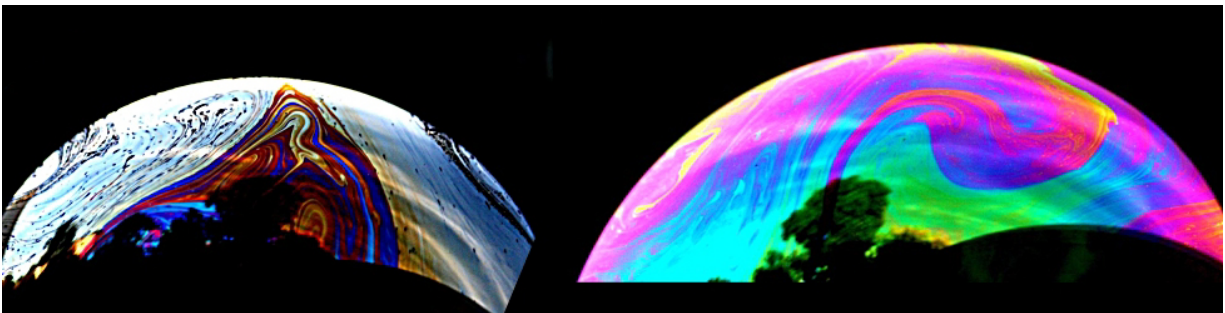
objective

lens nearest to the object being examined.

Introduction

class="introduction"

Soap bubbles are blown from clear fluid into very thin films. The colors we see are not due to any pigmentation but are the result of light interference, which enhances specific wavelengths for a given thickness of the film.



The most certain indication of a wave is interference. This wave characteristic is most prominent when the wave interacts with an object that

is not large compared with the wavelength. Interference is observed for water waves, sound waves, light waves, and, in fact, all types of waves.

If you have ever looked at the reds, blues, and greens in a sunlit soap bubble and wondered how straw-colored soapy water could produce them, you have hit upon one of the many phenomena that can only be explained by the wave character of light (see [\[link\]](#)). The same is true for the colors seen in an oil slick or in the light reflected from a DVD disc. These and other interesting phenomena cannot be explained fully by geometric optics. In these cases, light interacts with objects and exhibits wave characteristics. The branch of optics that considers the behavior of light when it exhibits wave characteristics is called wave optics (sometimes called physical optics). It is the topic of this chapter.

Young's Double-Slit Interference

By the end of this section, you will be able to:

- Explain the phenomenon of interference
- Define constructive and destructive interference for a double slit

The Dutch physicist Christiaan Huygens (1629–1695) thought that light was a wave, but Isaac Newton did not. Newton thought that there were other explanations for color, and for the interference and diffraction effects that were observable at the time. Owing to Newton's tremendous reputation, his view generally prevailed; the fact that Huygens's principle worked was not considered direct evidence proving that light is a wave. The acceptance of the wave character of light came many years later in 1801, when the English physicist and physician Thomas Young (1773–1829) demonstrated optical interference with his now-classic double-slit experiment.

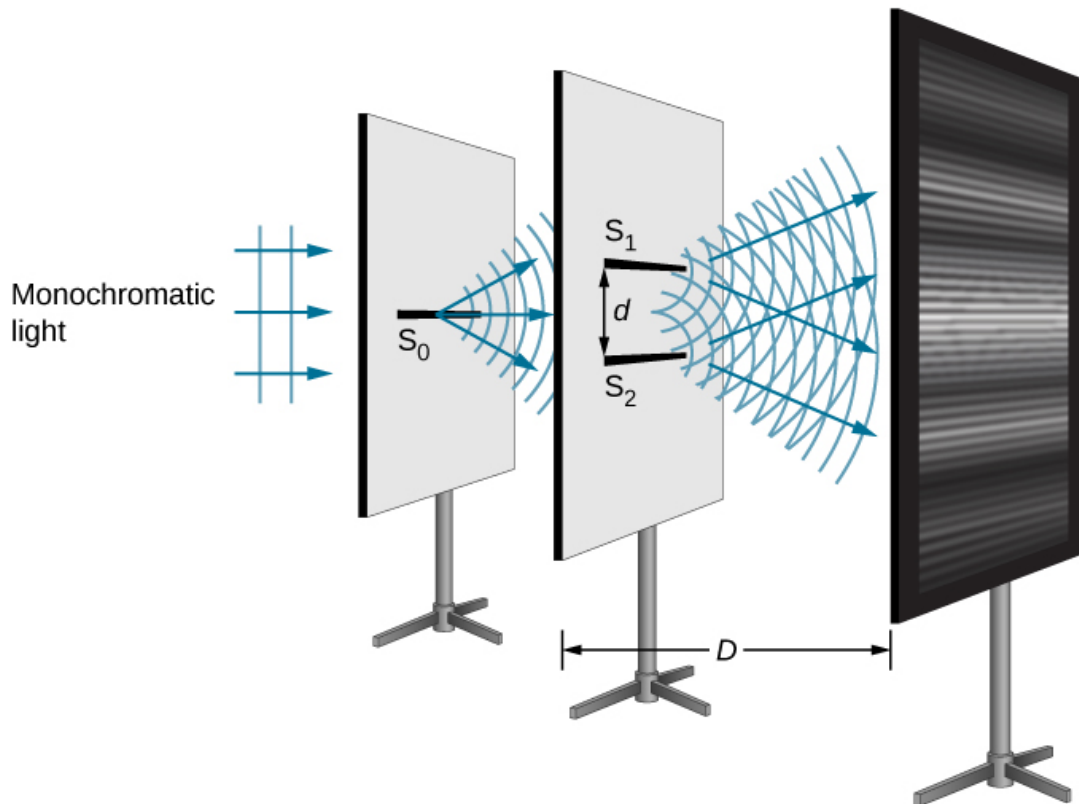
If there were not one but two sources of waves, the waves could be made to interfere, as in the case of waves on water ([\[link\]](#)). If light is an electromagnetic wave, it must therefore exhibit interference effects under appropriate circumstances. In Young's experiment, sunlight was passed through a pinhole on a board. The emerging beam fell on two pinholes on a second board. The light emanating from the two pinholes then fell on a screen where a pattern of bright and dark spots was observed. This pattern, called fringes, can only be explained through interference, a wave phenomenon.



Photograph of an interference pattern produced by circular water waves in a ripple tank. Two thin plungers are vibrated up and down in phase at the surface of the water. Circular water waves are produced by and emanate from each plunger.

We can analyze double-slit interference with the help of [\[link\]](#), which depicts an apparatus analogous to Young's. Light from a monochromatic source falls on a slit S_0 . The light emanating from S_0 is incident on two other slits S_1 and S_2 that are equidistant from S_0 . A pattern of *interference fringes* on the screen is then produced by the light emanating from S_1 and S_2 . All slits are assumed

to be so narrow that they can be considered secondary point sources for Huygens' wavelets ([The Nature of Light](#)). Slits S_1 and S_2 are a distance d apart ($d \leq 1 \text{ mm}$), and the distance between the screen and the slits is D ($\approx 1 \text{ m}$), which is much greater than d .

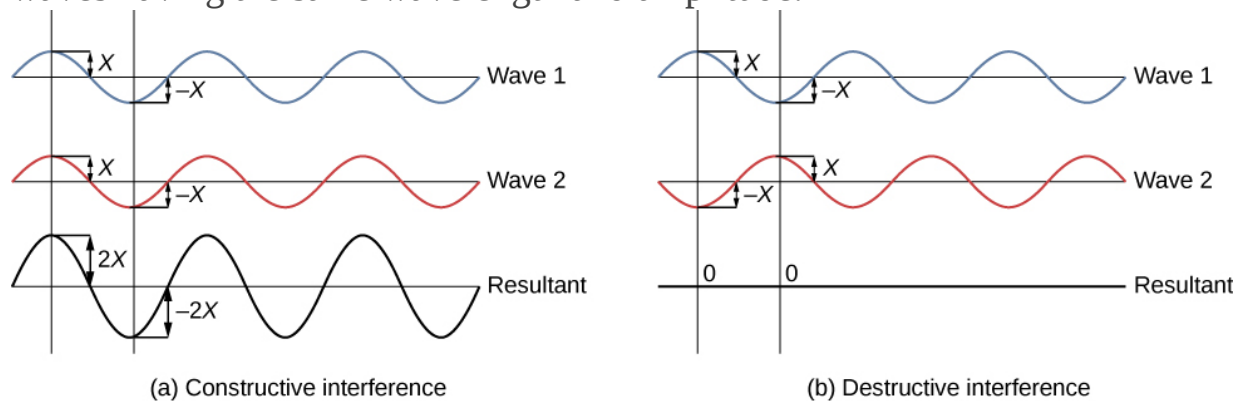


The double-slit interference experiment using monochromatic light and narrow slits. Fringes produced by interfering Huygens wavelets from slits S_1 and S_2 are observed on the screen.

Since S_0 is assumed to be a point source of monochromatic light, the secondary Huygens wavelets leaving S_1 and S_2 always maintain a constant phase difference (zero in this case because S_1 and S_2 are equidistant from S_0) and have the same frequency. The sources S_1 and S_2 are then said to be coherent. By **coherent waves**, we mean the waves are in phase or have a definite phase relationship. The term **incoherent** means the waves have random phase relationships, which would be the case if S_1 and S_2 were

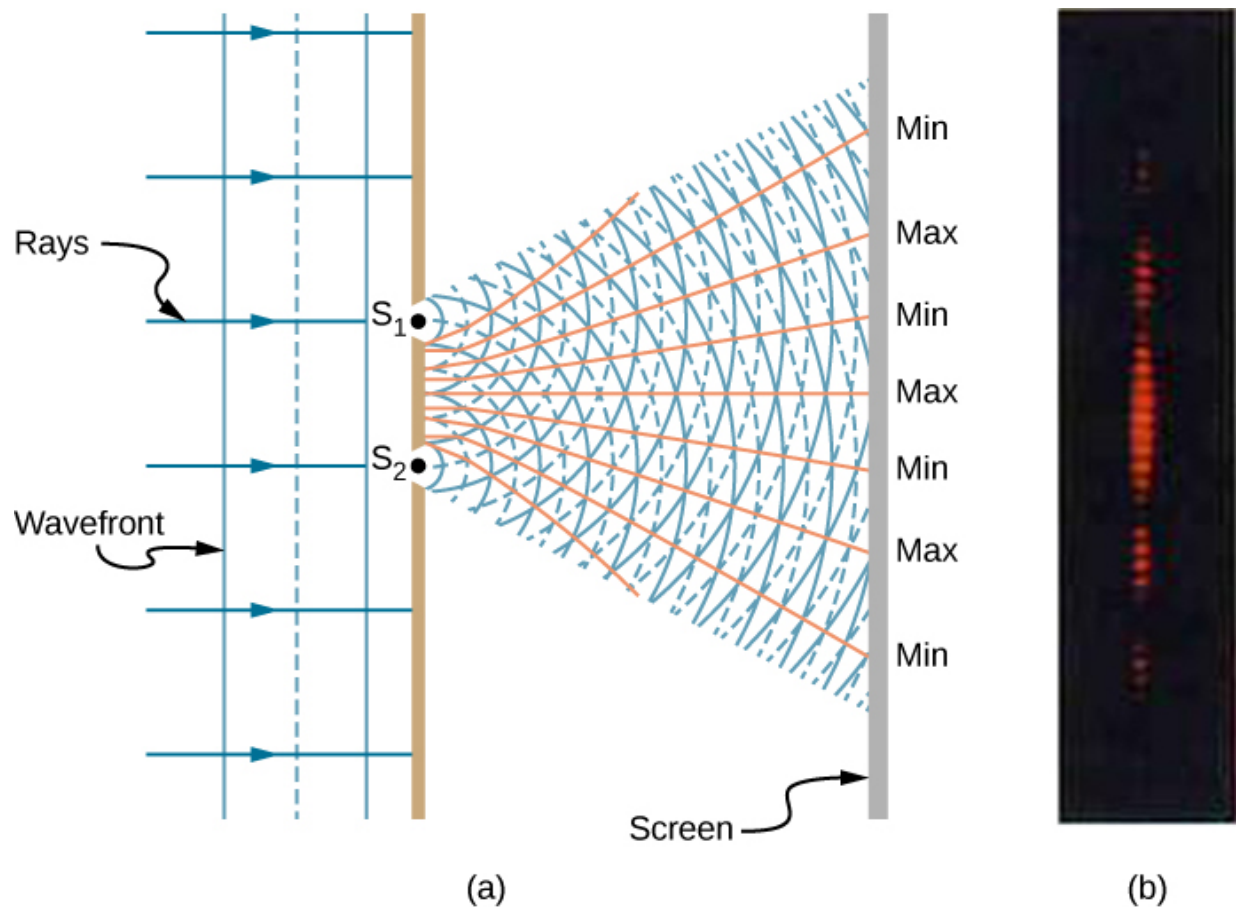
illuminated by two independent light sources, rather than a single source S_0 . Two independent light sources (which may be two separate areas within the same lamp or the Sun) would generally not emit their light in unison, that is, not coherently. Also, because S_1 and S_2 are the same distance from S_0 , the amplitudes of the two Huygens wavelets are equal.

Young used sunlight, where each wavelength forms its own pattern, making the effect more difficult to see. In the following discussion, we illustrate the double-slit experiment with **monochromatic** light (single λ) to clarify the effect. [\[link\]](#) shows the pure constructive and destructive interference of two waves having the same wavelength and amplitude.



The amplitudes of waves add. (a) Pure constructive interference is obtained when identical waves are in phase. (b) Pure destructive interference occurs when identical waves are exactly out of phase, or shifted by half a wavelength.

When light passes through narrow slits, the slits act as sources of coherent waves and light spreads out as semicircular waves, as shown in [\[link\]](#)(a). Pure *constructive interference* occurs where the waves are crest to crest or trough to trough. Pure *destructive interference* occurs where they are crest to trough. The light must fall on a screen and be scattered into our eyes for us to see the pattern. An analogous pattern for water waves is shown in [\[link\]](#). Note that regions of constructive and destructive interference move out from the slits at well-defined angles to the original beam. These angles depend on wavelength and the distance between the slits, as we shall see below.



Double slits produce two coherent sources of waves that interfere. (a) Light spreads out (diffracts) from each slit, because the slits are narrow. These waves overlap and interfere constructively (bright lines) and destructively (dark regions). We can only see this if the light falls onto a screen and is scattered into our eyes. (b) When light that has passed through double slits falls on a screen, we see a pattern such as this.

To understand the double-slit interference pattern, consider how two waves travel from the slits to the screen ([\[link\]](#)). Each slit is a different distance from a given point on the screen. Thus, different numbers of wavelengths fit into each path. Waves start out from the slits in phase (crest to crest), but they may end up out of phase (crest to trough) at the screen if the paths differ in length by half a wavelength, interfering destructively. If the paths differ by a whole wavelength, then the waves arrive in phase (crest to crest) at the screen, interfering constructively. More generally, if the path length difference Δl

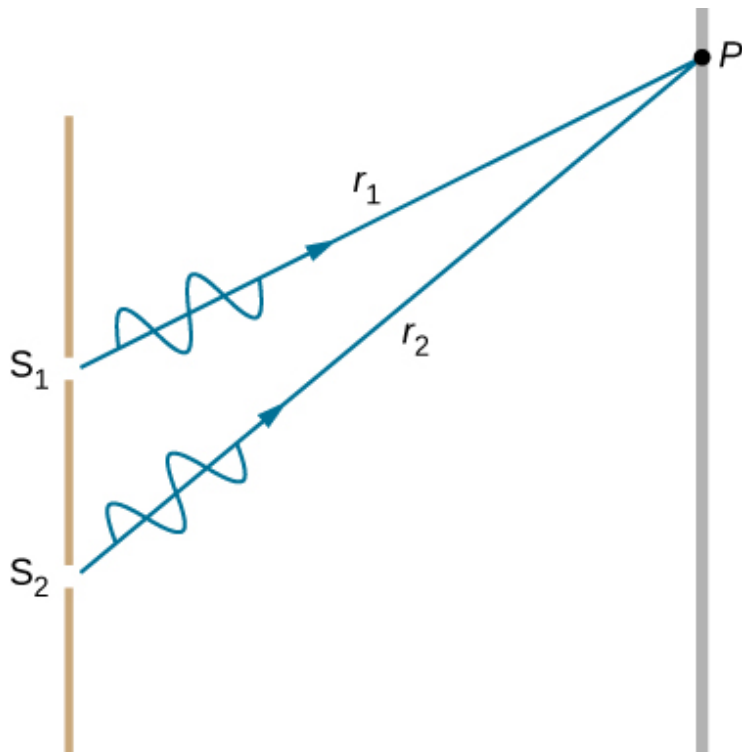
between the two waves is any half-integral number of wavelengths $[(1/2)\lambda, (3/2)\lambda, (5/2)\lambda, \text{etc.}]$, then destructive interference occurs. Similarly, if the path length difference is any integral number of wavelengths $(\lambda, 2\lambda, 3\lambda, \text{etc.})$, then constructive interference occurs. These conditions can be expressed as equations:

Equation:

$$\Delta l = m\lambda, \quad \text{for } m = 0, \pm 1, \pm 2, \pm 3 \dots \text{ (constructive interference)}$$

Equation:

$$\Delta l = (m + \frac{1}{2})\lambda, \quad \text{for } m = 0, \pm 1, \pm 2, \pm 3 \dots \text{ (destructive interference)}$$



Waves follow different paths from the slits to a common point P on a screen. Destructive interference occurs where one path is a half wavelength longer than the other—the waves start in phase but arrive out of phase. Constructive interference occurs where one path is a

whole wavelength longer than the other
—the waves start out and arrive in phase.

Summary

- Young's double-slit experiment gave definitive proof of the wave character of light.
- An interference pattern is obtained by the superposition of light from two slits.

Conceptual Questions

Exercise:

Problem:

Young's double-slit experiment breaks a single light beam into two sources. Would the same pattern be obtained for two independent sources of light, such as the headlights of a distant car? Explain.

Solution:

No. Two independent light sources do not have coherent phase.

Exercise:

Problem:

Is it possible to create an experimental setup in which there is only destructive interference? Explain.

Exercise:

Problem:

Why won't two small sodium lamps, held close together, produce an interference pattern on a distant screen? What if the sodium lamps were replaced by two laser pointers held close together?

Solution:

Because both the sodium lamps are not coherent pairs of light sources. Two lasers operating independently are also not coherent so no interference pattern results.

Glossary

coherent waves

waves are in phase or have a definite phase relationship

incoherent

waves have random phase relationships

monochromatic

light composed of one wavelength only

Mathematics of Interference

By the end of this section, you will be able to:

- Determine the angles for bright and dark fringes for double slit interference
- Calculate the positions of bright fringes on a screen

[\[link\]](#)(a) shows how to determine the path length difference Δl for waves traveling from two slits to a common point on a screen. If the screen is a large distance away compared with the distance between the slits, then the angle θ between the path and a line from the slits to the screen [part (b)] is nearly the same for each path. In other words, r_1 and r_2 are essentially parallel. The lengths of r_1 and r_2 differ by Δl , as indicated by the two dashed lines in the figure. Simple trigonometry shows

Equation:

$$\Delta l = d \sin \theta$$

where d is the distance between the slits. Combining this result with [\[link\]](#), we obtain constructive interference for a double slit when the path length difference is an integral multiple of the wavelength, or

Equation:

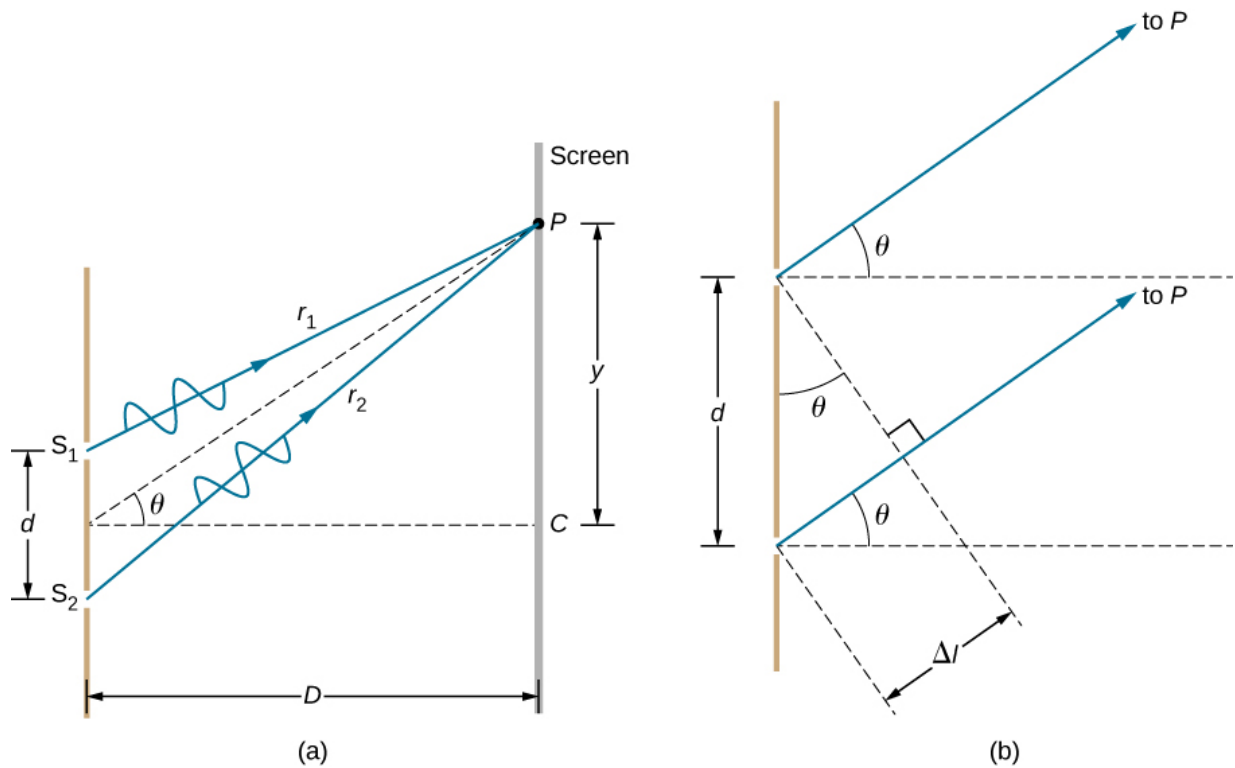
$$d \sin \theta = m\lambda, \text{ for } m = 0, \pm 1, \pm 2, \pm 3, \dots \text{ (constructive interference).}$$

Similarly, to obtain destructive interference for a double slit, the path length difference must be a half-integral multiple of the wavelength, or

Equation:

$$d \sin \theta = (m + \frac{1}{2})\lambda, \text{ for } m = 0, \pm 1, \pm 2, \pm 3, \dots \text{ (destructive interference)}$$

where λ is the wavelength of the light, d is the distance between slits, and θ is the angle from the original direction of the beam as discussed above. We call m the **order** of the interference. For example, $m = 4$ is fourth-order interference.



(a) To reach P , the light waves from S_1 and S_2 must travel different distances. (b) The path difference between the two rays is Δl .

The equations for double-slit interference imply that a series of bright and dark lines are formed. For vertical slits, the light spreads out horizontally on either side of the incident beam into a pattern called interference **fringes** ([\[link\]](#)). The closer the slits are, the more the bright fringes spread apart. We can see this by examining the equation

$d \sin \theta = m\lambda$, for $m = 0, \pm 1, \pm 2, \pm 3, \dots$. For fixed λ and m , the smaller d is, the larger θ must be, since $\sin \theta = m\lambda/d$. This is consistent with our contention that wave effects are most noticeable when the object the wave encounters (here, slits a distance d apart) is small. Small d gives large θ , hence, a large effect.

Referring back to part (a) of the figure, θ is typically small enough that $\sin \theta \approx \tan \theta \approx y_m/D$, where y_m is the distance from the central maximum to the m th bright fringe and D is the distance between the slit and the screen.

[\[link\]](#) may then be written as

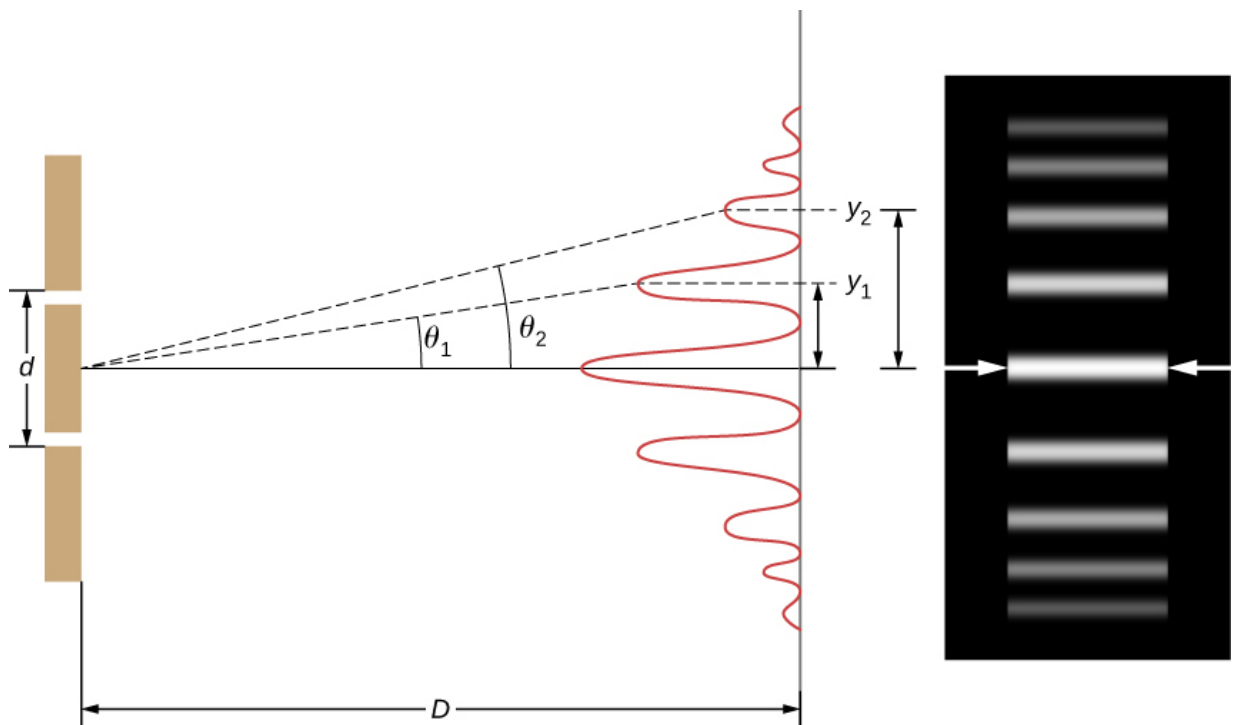
Equation:

$$d \frac{y_m}{D} = m\lambda$$

or

Equation:

$$y_m = \frac{m\lambda D}{d}.$$



The interference pattern for a double slit has an intensity that falls off with angle. The image shows multiple bright and dark lines, or fringes, formed by light passing through a double slit.

Example:

Finding a Wavelength from an Interference Pattern

Suppose you pass light from a He-Ne laser through two slits separated by 0.0100 mm and find that the third bright line on a screen is formed at an angle of 10.95° relative to the incident beam. What is the wavelength of the light?

Strategy

The phenomenon is two-slit interference as illustrated in [\[link\]](#) and the third bright line is due to third-order constructive interference, which means that $m = 3$. We are given $d = 0.0100$ mm and $\theta = 10.95^\circ$. The wavelength can thus be found using the equation $d \sin \theta = m\lambda$ for constructive interference.

Solution

Solving $d \sin \theta = m\lambda$ for the wavelength λ gives

Equation:

$$\lambda = \frac{d \sin \theta}{m}.$$

Substituting known values yields

Equation:

$$\lambda = \frac{(0.0100 \text{ mm})(\sin 10.95^\circ)}{3} = 6.33 \times 10^{-4} \text{ mm} = 633 \text{ nm}.$$

Significance

To three digits, this is the wavelength of light emitted by the common He-Ne laser. Not by coincidence, this red color is similar to that emitted by neon lights. More important, however, is the fact that interference patterns can be used to measure wavelength. Young did this for visible wavelengths. This analytical technique is still widely used to measure electromagnetic spectra. For a given order, the angle for constructive interference increases with λ , so that spectra (measurements of intensity versus wavelength) can be obtained.

Example:

Calculating the Highest Order Possible

Interference patterns do not have an infinite number of lines, since there is a limit to how big m can be. What is the highest-order constructive interference possible with the system described in the preceding example?

Strategy

The equation $d \sin \theta = m\lambda$ (for $m = 0, \pm 1, \pm 2, \pm 3, \dots$) describes constructive interference from two slits. For fixed values of d and λ , the larger m is, the larger $\sin \theta$ is. However, the maximum value that $\sin \theta$ can have is 1, for an angle of 90° . (Larger angles imply that light goes backward and does not reach the screen at all.) Let us find what value of m corresponds to this maximum diffraction angle.

Solution

Solving the equation $d \sin \theta = m\lambda$ for m gives

Equation:

$$m = \frac{d \sin \theta}{\lambda}.$$

Taking $\sin \theta = 1$ and substituting the values of d and λ from the preceding example gives

Equation:

$$m = \frac{(0.0100 \text{ mm})(1)}{633 \text{ nm}} \approx 15.8.$$

Therefore, the largest integer m can be is 15, or $m = 15$.

Significance

The number of fringes depends on the wavelength and slit separation. The number of fringes is very large for large slit separations. However, recall (see [The Propagation of Light](#) and the introduction for this chapter) that wave interference is only prominent when the wave interacts with objects that are not large compared to the wavelength. Therefore, if the slit separation and the sizes of the slits become much greater than the wavelength, the intensity pattern of light on the screen changes, so there are simply two bright lines cast by the slits, as expected, when light behaves like rays. We also note that the fringes get fainter farther away from the center. Consequently, not all 15 fringes may be observable.

Note:

Exercise:

Problem:

Check Your Understanding In the system used in the preceding examples, at what angles are the first and the second bright fringes formed?

Solution:

3.63° and 7.27° , respectively

Summary

- In double-slit diffraction, constructive interference occurs when $d \sin \theta = m\lambda$ (for $m = 0, \pm 1, \pm 2, \pm 3, \dots$), where d is the distance between the slits, θ is the angle relative to the incident direction, and m is the order of the interference.
- Destructive interference occurs when $d \sin \theta = (m + \frac{1}{2})\lambda$ for $m = 0, \pm 1, \pm 2, \pm 3, \dots$

Conceptual Questions

Exercise:**Problem:**

Suppose you use the same double slit to perform Young's double-slit experiment in air and then repeat the experiment in water. Do the angles to the same parts of the interference pattern get larger or smaller? Does the color of the light change? Explain.

Exercise:**Problem:**

Why is monochromatic light used in the double slit experiment? What would happen if white light were used?

Solution:

Monochromatic sources produce fringes at angles according to $d \sin \theta = m\lambda$. With white light, each constituent wavelength will produce fringes at its own set of angles, blending into the fringes of adjacent wavelengths. This results in rainbow patterns.

Problems

Exercise:

Problem:

At what angle is the first-order maximum for 450-nm wavelength blue light falling on double slits separated by 0.0500 mm?

Exercise:

Problem:

Calculate the angle for the third-order maximum of 580-nm wavelength yellow light falling on double slits separated by 0.100 mm.

Solution:

0.997°

Exercise:

Problem:

What is the separation between two slits for which 610-nm orange light has its first maximum at an angle of 30.0° ?

Exercise:

Problem:

Find the distance between two slits that produces the first minimum for 410-nm violet light at an angle of 45.0° .

Solution:

$0.290 \mu\text{m}$

Exercise:**Problem:**

Calculate the wavelength of light that has its third minimum at an angle of 30.0° when falling on double slits separated by $3.00\ \mu\text{m}$. Explicitly show how you follow the steps from the [Problem-Solving Strategy: Wave Optics](#), located at the end of the chapter.

Exercise:**Problem:**

What is the wavelength of light falling on double slits separated by $2.00\ \mu\text{m}$ if the third-order maximum is at an angle of 60.0° ?

Solution:

$$5.77 \times 10^{-7}\ \text{m} = 577\ \text{nm}$$

Exercise:**Problem:**

At what angle is the fourth-order maximum for the situation in the preceding problem?

Exercise:**Problem:**

What is the highest-order maximum for 400-nm light falling on double slits separated by $25.0\ \mu\text{m}$?

Solution:

62.5; since m must be an integer, the highest order is then $m = 62$.

Exercise:**Problem:**

Find the largest wavelength of light falling on double slits separated by $1.20\ \mu\text{m}$ for which there is a first-order maximum. Is this in the visible part of the spectrum?

Exercise:**Problem:**

What is the smallest separation between two slits that will produce a second-order maximum for 720-nm red light?

Solution:

1.44 μm

Exercise:**Problem:**

(a) What is the smallest separation between two slits that will produce a second-order maximum for any visible light? (b) For all visible light?

Exercise:**Problem:**

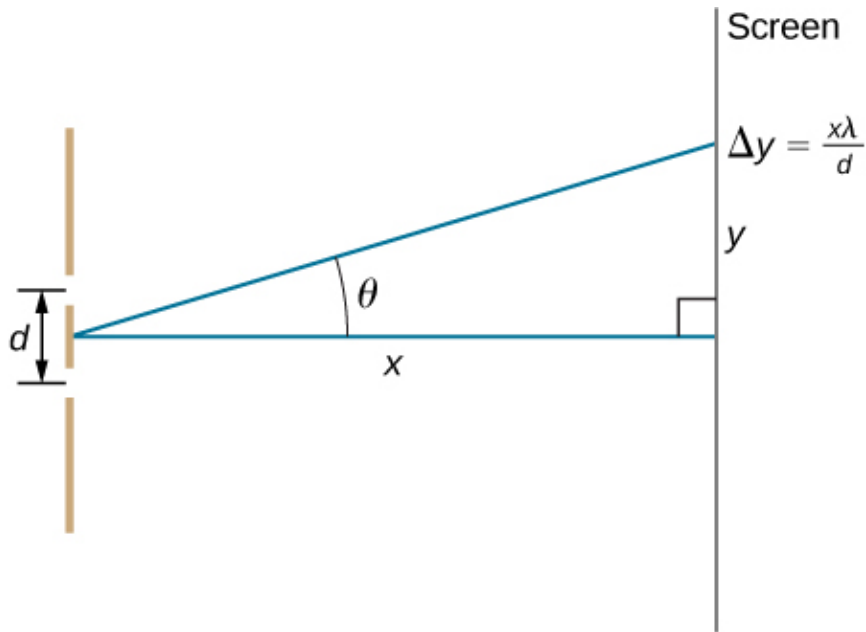
(a) If the first-order maximum for monochromatic light falling on a double slit is at an angle of 10.0° , at what angle is the second-order maximum? (b) What is the angle of the first minimum? (c) What is the highest-order maximum possible here?

Solution:

a. 20.3° ; b. 4.98° ; c. 5.76, the highest order is $m = 5$.

Exercise:**Problem:**

Shown below is a double slit located a distance x from a screen, with the distance from the center of the screen given by y . When the distance d between the slits is relatively large, numerous bright spots appear, called fringes. Show that, for small angles (where $\sin \theta \approx \theta$, with θ in radians), the distance between fringes is given by $\Delta y = x\lambda/d$



Exercise:

Problem:

Using the result of the preceding problem, (a) calculate the distance between fringes for 633-nm light falling on double slits separated by 0.0800 mm, located 3.00 m from a screen. (b) What would be the distance between fringes if the entire apparatus were submersed in water, whose index of refraction is 1.33?

Solution:

a. 2.37 cm; b. 1.78 cm

Exercise:

Problem:

Using the result of the problem two problems prior, find the wavelength of light that produces fringes 7.50 mm apart on a screen 2.00 m from double slits separated by 0.120 mm.

Exercise:

Problem:

In a double-slit experiment, the fifth maximum is 2.8 cm from the central maximum on a screen that is 1.5 m away from the slits. If the slits are 0.15 mm apart, what is the wavelength of the light being used?

Solution:

560 nm

Exercise:**Problem:**

The source in Young's experiment emits at two wavelengths. On the viewing screen, the fourth maximum for one wavelength is located at the same spot as the fifth maximum for the other wavelength. What is the ratio of the two wavelengths?

Exercise:**Problem:**

If 500-nm and 650-nm light illuminates two slits that are separated by 0.50 mm, how far apart are the second-order maxima for these two wavelengths on a screen 2.0 m away?

Solution:

1.2 mm

Exercise:**Problem:**

Red light of wavelength of 700 nm falls on a double slit separated by 400 nm. (a) At what angle is the first-order maximum in the diffraction pattern? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Glossary

fringes

bright and dark patterns of interference

order

integer m used in the equations for constructive and destructive interference for a double slit

Multiple-Slit Interference

By the end of this section, you will be able to:

- Describe the locations and intensities of secondary maxima for multiple-slit interference

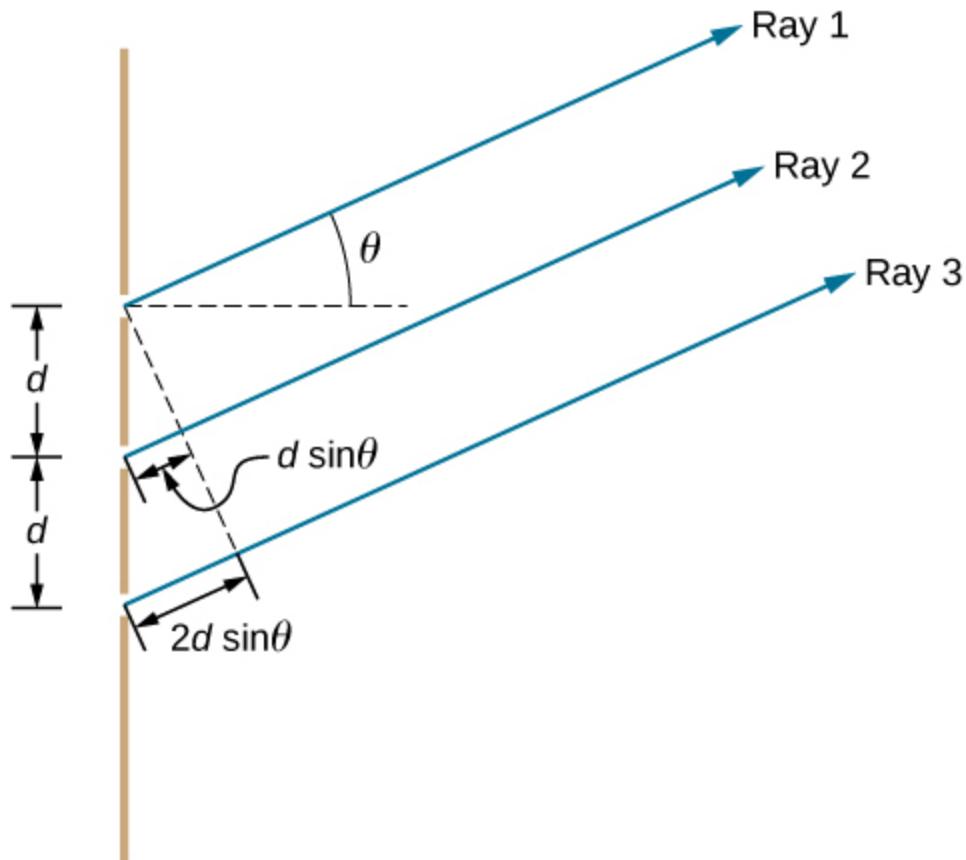
Analyzing the interference of light passing through two slits lays out the theoretical framework of interference and gives us a historical insight into Thomas Young's experiments. However, much of the modern-day application of slit interference uses not just two slits but many, approaching infinity for practical purposes. The key optical element is called a diffraction grating, an important tool in optical analysis, which we discuss in detail in [Diffraction](#). Here, we start the analysis of multiple-slit interference by taking the results from our analysis of the double slit ($N = 2$) and extending it to configurations with three, four, and much larger numbers of slits.

[\[link\]](#) shows the simplest case of multiple-slit interference, with three slits, or $N = 3$. The spacing between slits is d , and the path length difference between adjacent slits is $d \sin \theta$, same as the case for the double slit. What is new is that the path length difference for the first and the third slits is $2d \sin \theta$. The condition for constructive interference is the same as for the double slit, that is

Equation:

$$d \sin \theta = m\lambda.$$

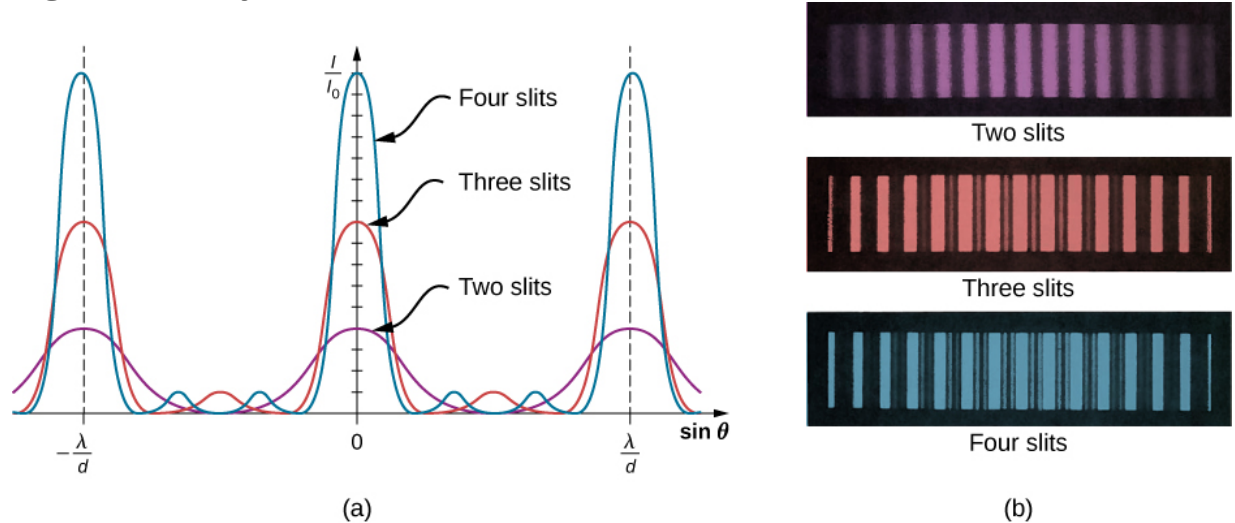
When this condition is met, $2d \sin \theta$ is automatically a multiple of λ , so all three rays combine constructively, and the bright fringes that occur here are called **principal maxima**. But what happens when the path length difference between adjacent slits is only $\lambda/2$? We can think of the first and second rays as interfering destructively, but the third ray remains unaltered. Instead of obtaining a dark fringe, or a minimum, as we did for the double slit, we see a **secondary maximum** with intensity lower than the principal maxima.



Interference with three slits. Different pairs of emerging rays can combine constructively or destructively at the same time, leading to secondary maxima.

In general, for N slits, these secondary maxima occur whenever an unpaired ray is present that does not go away due to destructive interference. This occurs at $(N - 2)$ evenly spaced positions between the principal maxima. The amplitude of the electromagnetic wave is correspondingly diminished to $1/N$ of the wave at the principal maxima, and the light intensity, being proportional to the square of the wave amplitude, is diminished to $1/N^2$ of the intensity compared to the principal maxima. As [\[link\]](#) shows, a dark fringe is located between every maximum (principal or secondary). As N grows larger and the number of bright and dark fringes increase, the widths of the maxima become narrower due to the closely located neighboring dark

fringes. Because the total amount of light energy remains unaltered, narrower maxima require that each maximum reaches a correspondingly higher intensity.



Interference fringe patterns for two, three and four slits. As the number of slits increases, more secondary maxima appear, but the principal maxima become brighter and narrower. (a) Graph and (b) photographs of fringe patterns.

Summary

- Interference from multiple slits ($N > 2$) produces principal as well as secondary maxima.
- As the number of slits is increased, the intensity of the principal maxima increases and the width decreases.

Problems

Exercise:

Problem:

Ten narrow slits are equally spaced 0.25 mm apart and illuminated with yellow light of wavelength 580 nm. (a) What are the angular positions of the third and fourth principal maxima? (b) What is the separation of these maxima on a screen 2.0 m from the slits?

Solution:

a. $0.40^\circ, 0.53^\circ$; b. $4.6 \times 10^{-3} \text{ m}$

Exercise:**Problem:**

The width of bright fringes can be calculated as the separation between the two adjacent dark fringes on either side. Find the angular widths of the third- and fourth-order bright fringes from the preceding problem.

Exercise:**Problem:**

For a three-slit interference pattern, find the ratio of the peak intensities of a secondary maximum to a principal maximum.

Solution:

1:9

Exercise:**Problem:**

What is the angular width of the central fringe of the interference pattern of (a) 20 slits separated by $d = 2.0 \times 10^{-3} \text{ mm}$? (b) 50 slits with the same separation? Assume that $\lambda = 600 \text{ nm}$.

Glossary

principal maximum

brightest interference fringes seen with multiple slits

secondary maximum

bright interference fringes of intensity lower than the principal maxima

Interference in Thin Films

By the end of this section, you will be able to:

- Describe the phase changes that occur upon reflection
- Describe fringes established by reflected rays of a common source
- Explain the appearance of colors in thin films

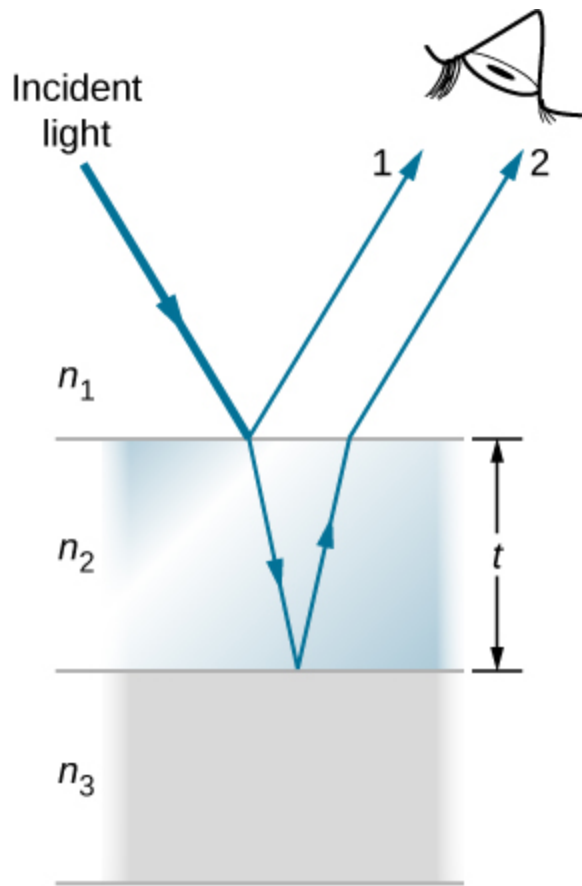
The bright colors seen in an oil slick floating on water or in a sunlit soap bubble are caused by interference. The brightest colors are those that interfere constructively. This interference is between light reflected from different surfaces of a thin film; thus, the effect is known as **thin-film interference**.

As we noted before, interference effects are most prominent when light interacts with something having a size similar to its wavelength. A thin film is one having a thickness t smaller than a few times the wavelength of light, λ . Since color is associated indirectly with λ and because all interference depends in some way on the ratio of λ to the size of the object involved, we should expect to see different colors for different thicknesses of a film, as in [\[link\]](#).



These soap bubbles exhibit brilliant colors when exposed to sunlight. (credit: Scott Robinson)

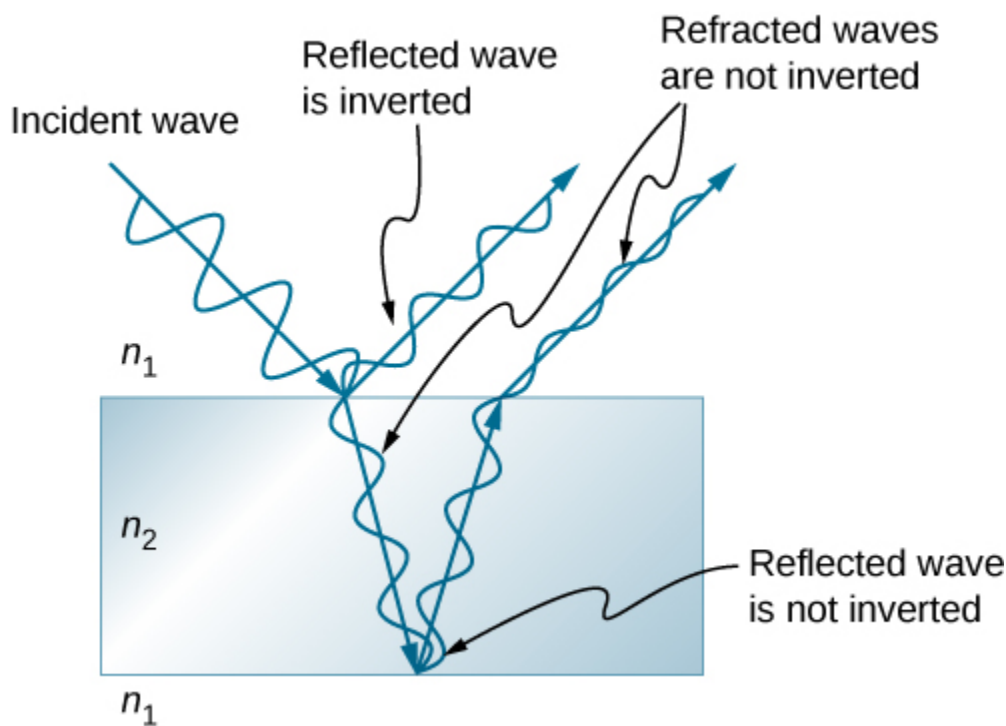
What causes thin-film interference? [\[link\]](#) shows how light reflected from the top and bottom surfaces of a film can interfere. Incident light is only partially reflected from the top surface of the film (ray 1). The remainder enters the film and is itself partially reflected from the bottom surface. Part of the light reflected from the bottom surface can emerge from the top of the film (ray 2) and interfere with light reflected from the top (ray 1). The ray that enters the film travels a greater distance, so it may be in or out of phase with the ray reflected from the top. However, consider for a moment, again, the bubbles in [\[link\]](#). The bubbles are darkest where they are thinnest. Furthermore, if you observe a soap bubble carefully, you will note it gets dark at the point where it breaks. For very thin films, the difference in path lengths of rays 1 and 2 in [\[link\]](#) is negligible, so why should they interfere destructively and not constructively? The answer is that a phase change can occur upon reflection, as discussed next.



Light striking a thin film is partially reflected (ray 1) and partially refracted at the top surface. The refracted ray is partially reflected at the bottom surface and emerges as ray 2. These rays interfere in a way that depends on the thickness of the film and the indices of refraction of the various media.

Changes in Phase due to Reflection

We saw earlier ([Waves](#)) that reflection of mechanical waves can involve a 180° phase change. For example, a traveling wave on a string is inverted (i.e., a 180° phase change) upon reflection at a boundary to which a heavier string is tied. However, if the second string is lighter (or more precisely, of a lower linear density), no inversion occurs. Light waves produce the same effect, but the deciding parameter for light is the index of refraction. Light waves undergo a 180° or π radians phase change upon reflection at an interface beyond which is a medium of higher index of refraction. No phase change takes place when reflecting from a medium of lower refractive index ([link](#)). Because of the periodic nature of waves, this phase change or inversion is equivalent to $\pm\lambda/2$ in distance travelled, or path length. Both the path length and refractive indices are important factors in thin-film interference.



Reflection at an interface for light traveling from a medium with index of refraction n_1 to a medium with index of refraction n_2 , $n_1 < n_2$, causes the phase of the wave to change by π radians.

If the film in [\[link\]](#) is a **soap bubble** (essentially water with air on both sides), then a phase shift of $\lambda/2$ occurs for ray 1 but not for ray 2. Thus, when the film is very thin and the path length difference between the two rays is negligible, they are exactly out of phase, and destructive interference occurs at all wavelengths. Thus, the soap bubble is dark here. The thickness of the film relative to the wavelength of light is the other crucial factor in thin-film interference. Ray 2 in [\[link\]](#) travels a greater distance than ray 1. For light incident perpendicular to the surface, ray 2 travels a distance approximately $2t$ farther than ray 1. When this distance is an integral or half-integral multiple of the wavelength in the medium ($\lambda_n = \lambda/n$, where λ is the wavelength in vacuum and n is the index of refraction), constructive or destructive interference occurs, depending also on whether there is a phase change in either ray.

Example:

Calculating the Thickness of a Nonreflective Lens Coating

Sophisticated cameras use a series of several lenses. Light can reflect from the surfaces of these various lenses and degrade image clarity. To limit these reflections, lenses are coated with a thin layer of magnesium fluoride, which causes destructive thin-film interference. What is the thinnest this film can be, if its index of refraction is 1.38 and it is designed to limit the reflection of 550-nm light, normally the most intense visible wavelength? Assume the index of refraction of the glass is 1.52.

Strategy

Refer to [\[link\]](#) and use $n_1 = 1.00$ for air, $n_2 = 1.38$, and $n_3 = 1.52$. Both ray 1 and ray 2 have a $\lambda/2$ shift upon reflection. Thus, to obtain destructive interference, ray 2 needs to travel a half wavelength farther than ray 1. For rays incident perpendicularly, the path length difference is $2t$.

Solution

To obtain destructive interference here,

Equation:

$$2t = \frac{\lambda_{n2}}{2}$$

where λ_{n2} is the wavelength in the film and is given by $\lambda_{n2} = \lambda/n_2$.
Thus,

Equation:

$$2t = \frac{\lambda/n_2}{2}.$$

Solving for t and entering known values yields

Equation:

$$t = \frac{\lambda/n_2}{4} = \frac{(500 \text{ nm})/1.38}{4} = 90.6 \text{ nm}.$$

Significance

Films such as the one in this example are most effective in producing destructive interference when the thinnest layer is used, since light over a broader range of incident angles is reduced in intensity. These films are called nonreflective coatings; this is only an approximately correct description, though, since other wavelengths are only partially cancelled. Nonreflective coatings are also used in car windows and sunglasses.

Combining Path Length Difference with Phase Change

Thin-film interference is most constructive or most destructive when the path length difference for the two rays is an integral or half-integral wavelength. That is, for rays incident perpendicularly,

Equation:

$$2t = \lambda_n, 2\lambda_n, 3\lambda_n, \dots \text{ or } 2t = \lambda_n/2, 3\lambda_n/2, 5\lambda_n/2, \dots$$

To know whether interference is constructive or destructive, you must also determine if there is a phase change upon reflection. Thin-film interference

thus depends on film thickness, the wavelength of light, and the refractive indices. For white light incident on a film that varies in thickness, you can observe rainbow colors of constructive interference for various wavelengths as the thickness varies.

Example:**Soap Bubbles**

(a) What are the three smallest thicknesses of a soap bubble that produce constructive interference for red light with a wavelength of 650 nm? The index of refraction of soap is taken to be the same as that of water. (b) What three smallest thicknesses give destructive interference?

Strategy

Use [\[link\]](#) to visualize the bubble, which acts as a thin film between two layers of air. Thus $n_1 = n_3 = 1.00$ for air, and $n_2 = 1.333$ for soap (equivalent to water). There is a $\lambda/2$ shift for ray 1 reflected from the top surface of the bubble and no shift for ray 2 reflected from the bottom surface. To get constructive interference, then, the path length difference ($2t$) must be a half-integral multiple of the wavelength—the first three being $\lambda_n/2$, $3\lambda_n/2$, and $5\lambda_n/2$. To get destructive interference, the path length difference must be an integral multiple of the wavelength—the first three being 0, λ_n , and $2\lambda_n$.

Solution

a. Constructive interference occurs here when

Equation:

$$2t_c = \frac{\lambda_n}{2}, \frac{3\lambda_n}{2}, \frac{5\lambda_n}{2}, \dots$$

Thus, the smallest constructive thickness t_c is

Equation:

$$t_c = \frac{\lambda_n}{4} = \frac{\lambda/n}{4} = \frac{(650 \text{ nm})/1.333}{4} = 122 \text{ nm}.$$

The next thickness that gives constructive interference is $t'_c = 3\lambda_n/4$, so that

Equation:

$$t'_c = 366 \text{ nm.}$$

Finally, the third thickness producing constructive interference is $t'_c = 5\lambda_n/4$, so that

Equation:

$$t'_c = 610 \text{ nm.}$$

b. For destructive interference, the path length difference here is an integral multiple of the wavelength. The first occurs for zero thickness, since there is a phase change at the top surface, that is,

Equation:

$$t_d = 0,$$

the very thin (or negligibly thin) case discussed above. The first non-zero thickness producing destructive interference is

Equation:

$$2t'_d = \lambda_n.$$

Substituting known values gives

Equation:

$$t'_d = \frac{\lambda}{2} = \frac{\lambda/n}{2} = \frac{(650 \text{ nm})/1.333}{2} = 244 \text{ nm.}$$

Finally, the third destructive thickness is $2t''_d = 2\lambda_n$, so that

Equation:

$$t''_d = \lambda_n = \frac{\lambda}{n} = \frac{650 \text{ nm}}{1.333} = 488 \text{ nm.}$$

Significance

If the bubble were illuminated with pure red light, we would see bright and dark bands at very uniform increases in thickness. First would be a dark band at 0 thickness, then bright at 122 nm thickness, then dark at 244 nm,

bright at 366 nm, dark at 488 nm, and bright at 610 nm. If the bubble varied smoothly in thickness, like a smooth wedge, then the bands would be evenly spaced.

Note:

Exercise:

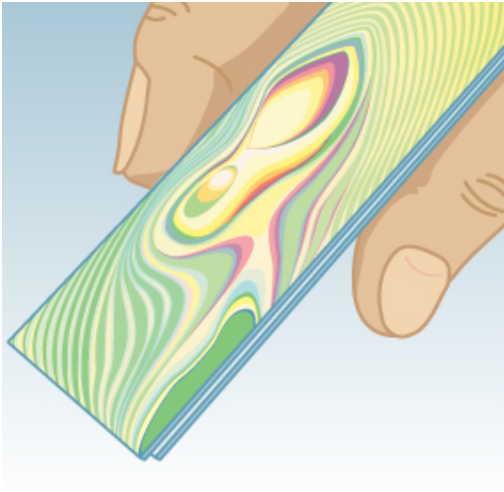
Problem:

Check Your Understanding Going further with [\[link\]](#), what are the next two thicknesses of soap bubble that would lead to (a) constructive interference, and (b) destructive interference?

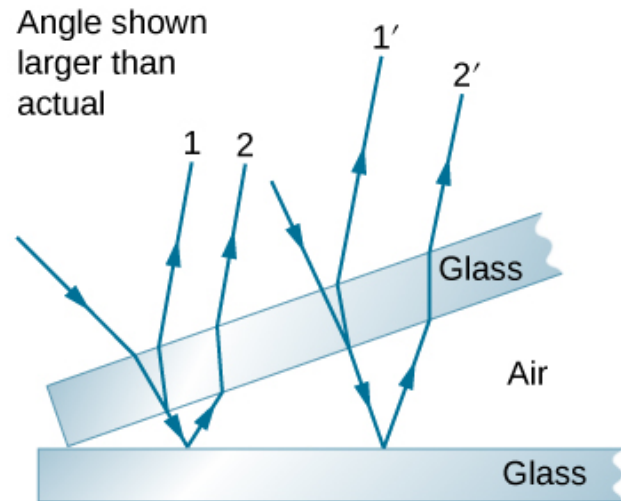
Solution:

a. 853 nm, 1097 nm; b. 731 nm, 975 nm

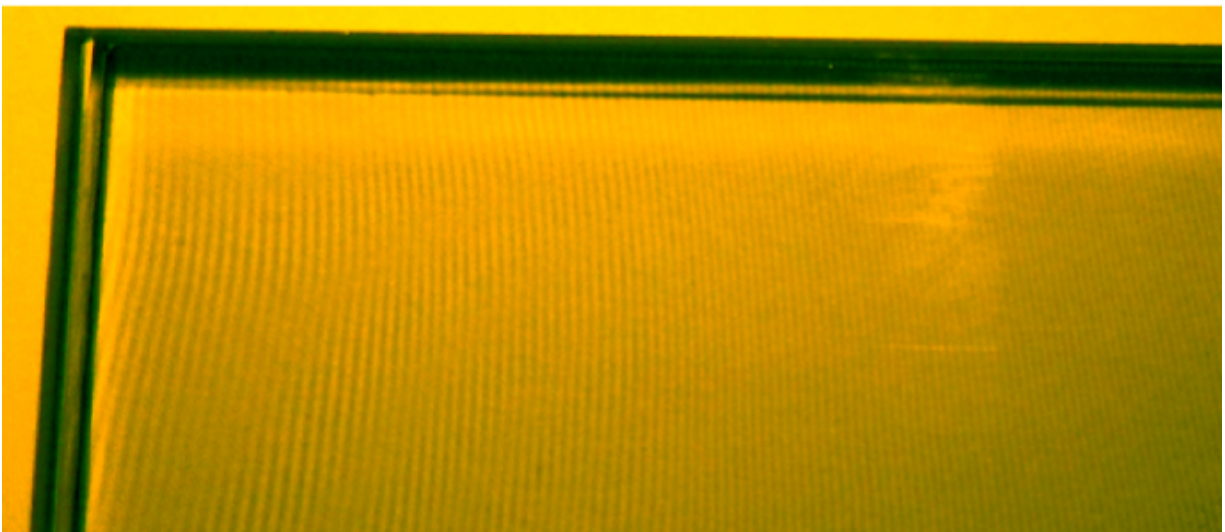
Another example of thin-film interference can be seen when microscope slides are separated (see [\[link\]](#)). The slides are very flat, so that the wedge of air between them increases in thickness very uniformly. A phase change occurs at the second surface but not the first, so a dark band forms where the slides touch. The rainbow colors of constructive interference repeat, going from violet to red again and again as the distance between the slides increases. As the layer of air increases, the bands become more difficult to see, because slight changes in incident angle have greater effects on path length differences. If monochromatic light instead of white light is used, then bright and dark bands are obtained rather than repeating rainbow colors.



(a)



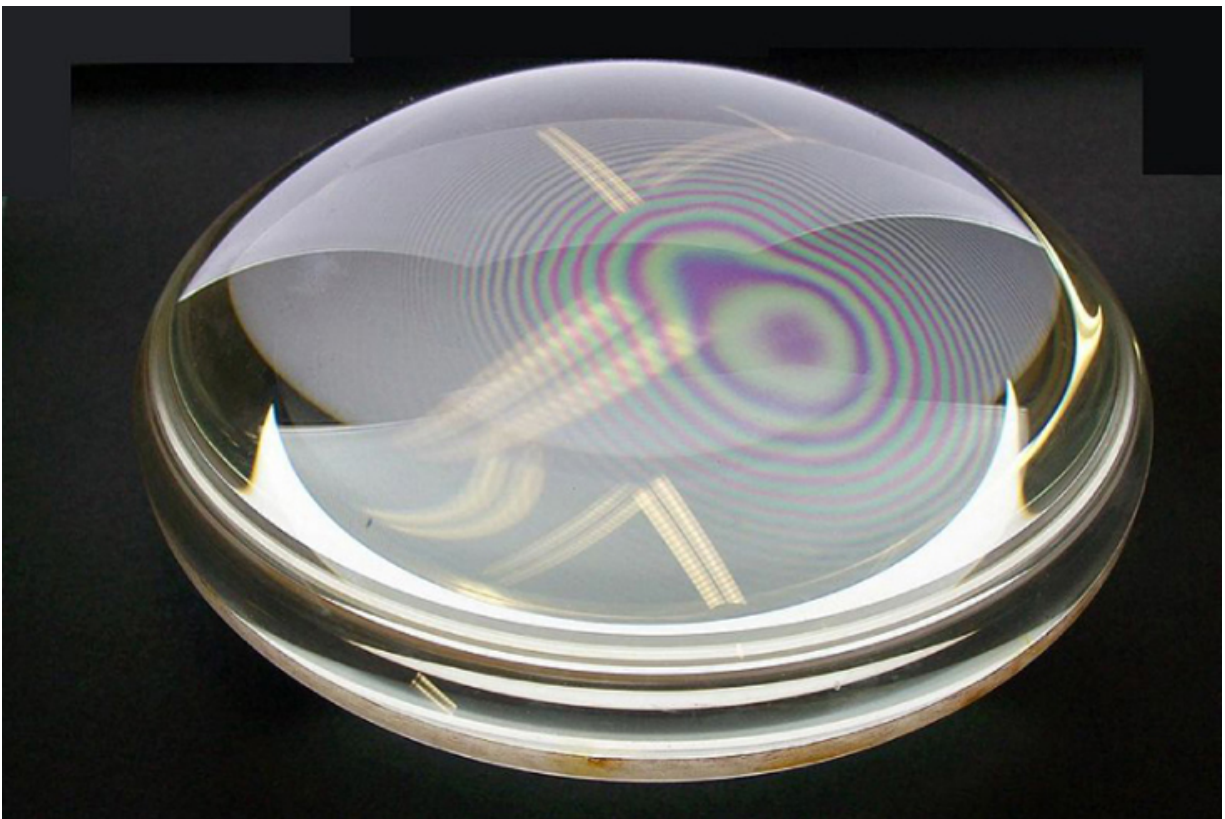
(b)



(c)

(a) The rainbow-color bands are produced by thin-film interference in the air between the two glass slides. (b) Schematic of the paths taken by rays in the wedge of air between the slides. (c) If the air wedge is illuminated with monochromatic light, bright and dark bands are obtained rather than repeating rainbow colors.

An important application of thin-film interference is found in the manufacturing of optical instruments. A lens or mirror can be compared with a master as it is being ground, allowing it to be shaped to an accuracy of less than a wavelength over its entire surface. [\[link\]](#) illustrates the phenomenon called **Newton's rings**, which occurs when the plane surfaces of two lenses are placed together. (The circular bands are called Newton's rings because Isaac Newton described them and their use in detail. Newton did not discover them; Robert Hooke did, and Newton did not believe they were due to the wave character of light.) Each successive ring of a given color indicates an increase of only half a wavelength in the distance between the lens and the blank, so that great precision can be obtained. Once the lens is perfect, no rings appear.



“Newton's rings” interference fringes are produced when two plano-convex lenses are placed together with their plane surfaces in contact. The rings are created by interference between the light reflected off the two surfaces as a result of a slight gap between them, indicating that

these surfaces are not precisely plane but are slightly convex. (credit: Ulf Seifert)

Thin-film interference has many other applications, both in nature and in manufacturing. The wings of certain moths and butterflies have nearly iridescent colors due to thin-film interference. In addition to pigmentation, the wing's color is affected greatly by constructive interference of certain wavelengths reflected from its film-coated surface. Some car manufacturers offer special paint jobs that use thin-film interference to produce colors that change with angle. This expensive option is based on variation of thin-film path length differences with angle. Security features on credit cards, banknotes, driving licenses, and similar items prone to forgery use thin-film interference, diffraction gratings, or holograms. As early as 1998, Australia led the way with dollar bills printed on polymer with a diffraction grating security feature, making the currency difficult to forge. Other countries, such as Canada, New Zealand, and Taiwan, are using similar technologies, while US currency includes a thin-film interference effect.

Summary

- When light reflects from a medium having an index of refraction greater than that of the medium in which it is traveling, a 180° phase change (or a $\lambda/2$ shift) occurs.
- Thin-film interference occurs between the light reflected from the top and bottom surfaces of a film. In addition to the path length difference, there can be a phase change.

Conceptual Questions

Exercise:

Problem:

What effect does increasing the wedge angle have on the spacing of interference fringes? If the wedge angle is too large, fringes are not observed. Why?

Exercise:**Problem:**

How is the difference in paths taken by two originally in-phase light waves related to whether they interfere constructively or destructively? How can this be affected by reflection? By refraction?

Solution:

Differing path lengths result in different phases at destination resulting in constructive or destructive interference accordingly. Reflection can cause a 180° phase change, which also affects how waves interfere. Refraction into another medium changes the wavelength inside that medium such that a wave can emerge from the medium with a different phase compared to another wave that travelled the same distance in a different medium.

Exercise:**Problem:**

Is there a phase change in the light reflected from either surface of a contact lens floating on a person's tear layer? The index of refraction of the lens is about 1.5, and its top surface is dry.

Exercise:**Problem:**

In placing a sample on a microscope slide, a glass cover is placed over a water drop on the glass slide. Light incident from above can reflect from the top and bottom of the glass cover and from the glass slide below the water drop. At which surfaces will there be a phase change in the reflected light?

Solution:

Phase changes occur upon reflection at the top of glass cover and the top of glass slide only.

Exercise:**Problem:**

Answer the above question if the fluid between the two pieces of crown glass is carbon disulfide.

Exercise:**Problem:**

While contemplating the food value of a slice of ham, you notice a rainbow of color reflected from its moist surface. Explain its origin.

Solution:

The surface of the ham being moist means there is a thin layer of fluid, resulting in thin-film interference. Because the exact thickness of the film varies across the piece of ham, which is illuminated by white light, different wavelengths produce bright fringes at different locations, resulting in rainbow colors.

Exercise:**Problem:**

An inventor notices that a soap bubble is dark at its thinnest and realizes that destructive interference is taking place for all wavelengths. How could she use this knowledge to make a nonreflective coating for lenses that is effective at all wavelengths? That is, what limits would there be on the index of refraction and thickness of the coating? How might this be impractical?

Exercise:

Problem:

A nonreflective coating like the one described in [\[link\]](#) works ideally for a single wavelength and for perpendicular incidence. What happens for other wavelengths and other incident directions? Be specific.

Solution:

Other wavelengths will not generally satisfy $t = \frac{\lambda/n}{4}$ for the same value of t so reflections will result in completely destructive interference. For an incidence angle θ , the path length inside the coating will be increased by a factor $1/\cos \theta$ so the new condition for destructive interference becomes $\frac{t}{\cos \theta} = \frac{\lambda/n}{4}$.

Exercise:**Problem:**

Why is it much more difficult to see interference fringes for light reflected from a thick piece of glass than from a thin film? Would it be easier if monochromatic light were used?

Problems**Exercise:****Problem:**

A soap bubble is 100 nm thick and illuminated by white light incident perpendicular to its surface. What wavelength and color of visible light is most constructively reflected, assuming the same index of refraction as water?

Solution:

532 nm (green)

Exercise:

Problem:

An oil slick on water is 120 nm thick and illuminated by white light incident perpendicular to its surface. What color does the oil appear (what is the most constructively reflected wavelength), given its index of refraction is 1.40?

Exercise:**Problem:**

Calculate the minimum thickness of an oil slick on water that appears blue when illuminated by white light perpendicular to its surface. Take the blue wavelength to be 470 nm and the index of refraction of oil to be 1.40.

Solution:

$$8.39 \times 10^{-8} \text{ m} = 83.9 \text{ nm}$$

Exercise:**Problem:**

Find the minimum thickness of a soap bubble that appears red when illuminated by white light perpendicular to its surface. Take the wavelength to be 680 nm, and assume the same index of refraction as water.

Exercise:**Problem:**

A film of soapy water ($n = 1.33$) on top of a plastic cutting board has a thickness of 233 nm. What color is most strongly reflected if it is illuminated perpendicular to its surface?

Solution:

620 nm (orange)

Exercise:**Problem:**

What are the three smallest non-zero thicknesses of soapy water ($n = 1.33$) on Plexiglas if it appears green (constructively reflecting 520-nm light) when illuminated perpendicularly by white light?

Exercise:**Problem:**

Suppose you have a lens system that is to be used primarily for 700-nm red light. What is the second thinnest coating of fluorite (magnesium fluoride) that would be nonreflective for this wavelength?

Solution:

380 nm

Exercise:**Problem:**

(a) As a soap bubble thins it becomes dark, because the path length difference becomes small compared with the wavelength of light and there is a phase shift at the top surface. If it becomes dark when the path length difference is less than one-fourth the wavelength, what is the thickest the bubble can be and appear dark at all visible wavelengths? Assume the same index of refraction as water. (b) Discuss the fragility of the film considering the thickness found.

Exercise:

Problem:

To save money on making military aircraft invisible to radar, an inventor decides to coat them with a nonreflective material having an index of refraction of 1.20, which is between that of air and the surface of the plane. This, he reasons, should be much cheaper than designing Stealth bombers. (a) What thickness should the coating be to inhibit the reflection of 4.00-cm wavelength radar? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. Assuming n for the plane is greater than 1.20, then there are two phase changes: 0.833 cm. b. It is too thick, and the plane would be too heavy. c. It is unreasonable to think the layer of material could be any thickness when used on a real aircraft.

Glossary

Newton's rings

circular interference pattern created by interference between the light reflected off two surfaces as a result of a slight gap between them

thin-film interference

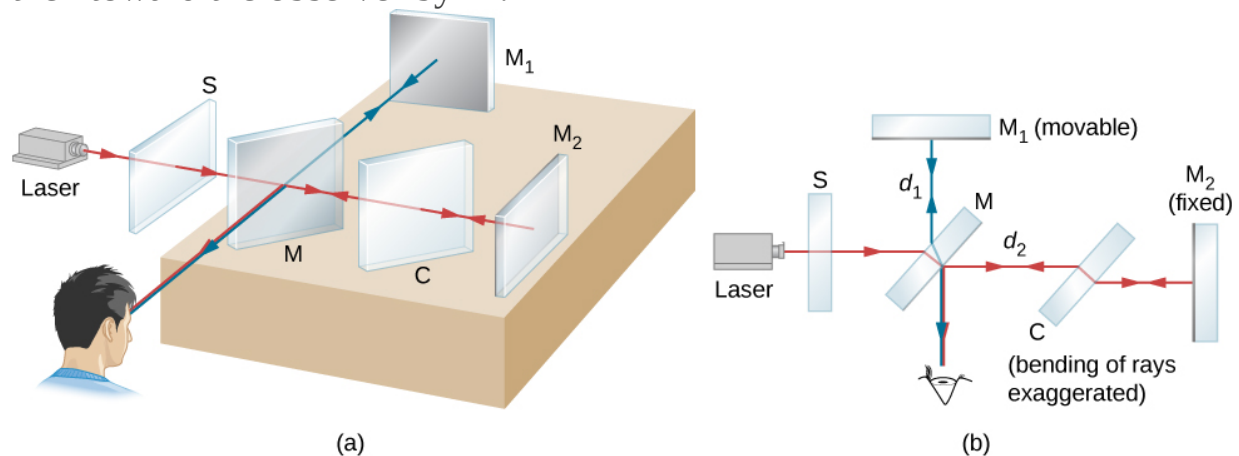
interference between light reflected from different surfaces of a thin film

The Michelson Interferometer

By the end of this section, you will be able to:

- Explain changes in fringes observed with a Michelson interferometer caused by mirror movements
- Explain changes in fringes observed with a Michelson interferometer caused by changes in medium

The Michelson **interferometer** (invented by the American physicist Albert A. Michelson, 1852–1931) is a precision instrument that produces interference fringes by splitting a light beam into two parts and then recombining them after they have traveled different optical paths. [\[link\]](#) depicts the interferometer and the path of a light beam from a single point on the extended source S, which is a ground-glass plate that diffuses the light from a monochromatic lamp of wavelength λ_0 . The beam strikes the half-silvered mirror M, where half of it is reflected to the side and half passes through the mirror. The reflected light travels to the movable plane mirror M_1 , where it is reflected back through M to the observer. The transmitted half of the original beam is reflected back by the stationary mirror M_2 and then toward the observer by M.



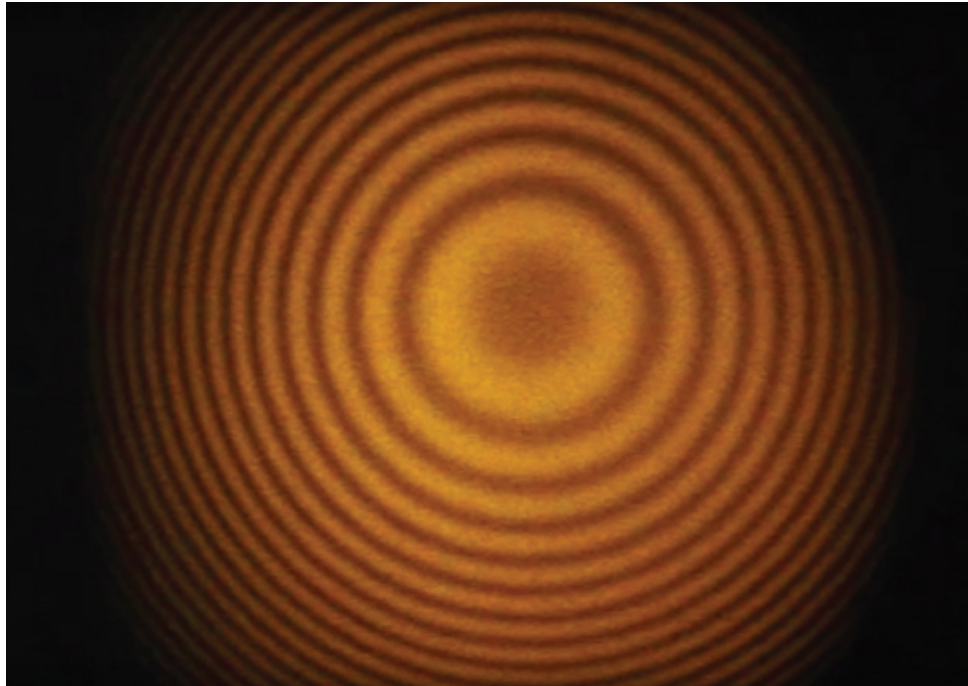
(a) The Michelson interferometer. The extended light source is a ground-glass plate that diffuses the light from a laser. (b) A planar view of the interferometer.

Because both beams originate from the same point on the source, they are coherent and therefore interfere. Notice from the figure that one beam passes through M three times and the other only once. To ensure that both beams traverse the same thickness of glass, a compensator plate C of transparent glass is placed in the arm containing M₂. This plate is a duplicate of M (without the silvering) and is usually cut from the same piece of glass used to produce M. With the compensator in place, any phase difference between the two beams is due solely to the difference in the distances they travel.

The path difference of the two beams when they recombine is $2d_1 - 2d_2$, where d_1 is the distance between M and M₁, and d_2 is the distance between M and M₂. Suppose this path difference is an integer number of wavelengths $m\lambda_0$. Then, constructive interference occurs and a bright image of the point on the source is seen at the observer. Now the light from any other point on the source whose two beams have this same path difference also undergoes constructive interference and produces a bright image. The collection of these point images is a bright fringe corresponding to a path difference of $m\lambda_0$ ([\[link\]](#)). When M₁ is moved a distance $\Delta d = \lambda_0/2$, this path difference changes by λ_0 , and each fringe moves to the position previously occupied by an adjacent fringe. Consequently, by counting the number of fringes m passing a given point as M₁ is moved, an observer can measure minute displacements that are accurate to a fraction of a wavelength, as shown by the relation

Equation:

$$\Delta d = m \frac{\lambda_0}{2}.$$



Fringes produced with a Michelson interferometer.
(credit: "SILLAGESvideos"/YouTube)

Example:**Precise Distance Measurements by Michelson Interferometer**

A red laser light of wavelength 630 nm is used in a Michelson interferometer. While keeping the mirror M_1 fixed, mirror M_2 is moved. The fringes are found to move past a fixed cross-hair in the viewer. Find the distance the mirror M_2 is moved for a single fringe to move past the reference line.

Strategy

Refer to [\[link\]](#) for the geometry. We use the result of the Michelson interferometer interference condition to find the distance moved, Δd .

Solution

For a 630-nm red laser light, and for each fringe crossing ($m = 1$), the distance traveled by M_2 if you keep M_1 fixed is

Equation:

$$\Delta d = m \frac{\lambda_0}{2} = 1 \times \frac{630 \text{ nm}}{2} = 315 \text{ nm} = 0.315 \mu\text{m}.$$

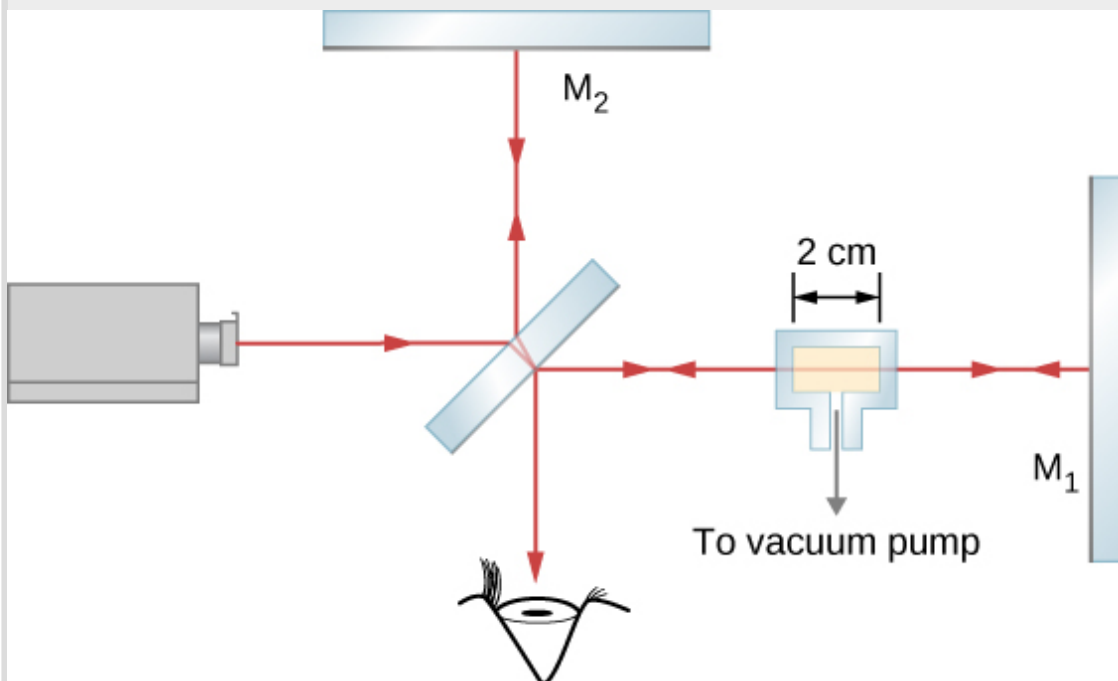
Significance

An important application of this measurement is the definition of the standard meter. As mentioned in [Units and Measurement](#), the length of the standard meter was once defined as the mirror displacement in a Michelson interferometer corresponding to 1,650,763.73 wavelengths of the particular fringe of krypton-86 in a gas discharge tube.

Example:

Measuring the Refractive Index of a Gas

In one arm of a Michelson interferometer, a glass chamber is placed with attachments for evacuating the inside and putting gases in it. The space inside the container is 2 cm wide. Initially, the container is empty. As gas is slowly let into the chamber, you observe that dark fringes move past a reference line in the field of observation. By the time the chamber is filled to the desired pressure, you have counted 122 fringes move past the reference line. The wavelength of the light used is 632.8 nm. What is the refractive index of this gas?



Strategy

The $m = 122$ fringes observed compose the difference between the number of wavelengths that fit within the empty chamber (vacuum) and the number of wavelengths that fit within the same chamber when it is gas-filled. The wavelength in the filled chamber is shorter by a factor of n , the index of refraction.

Solution

The ray travels a distance $t = 2$ cm to the right through the glass chamber and another distance t to the left upon reflection. The total travel is $L = 2t$. When empty, the number of wavelengths that fit in this chamber is

Equation:

$$N_0 = \frac{L}{\lambda_0} = \frac{2t}{\lambda_0}$$

where $\lambda_0 = 632.8$ nm is the wavelength in vacuum of the light used. In any other medium, the wavelength is $\lambda = \lambda_0/n$ and the number of wavelengths that fit in the gas-filled chamber is

Equation:

$$N = \frac{L}{\lambda} = \frac{2t}{\lambda_0/n}.$$

The number of fringes observed in the transition is

Equation:

$$\begin{aligned} m &= N - N_0, \\ &= \frac{2t}{\lambda_0/n} - \frac{2t}{\lambda_0}, \\ &= \frac{2t}{\lambda_0}(n - 1). \end{aligned}$$

Solving for $(n - 1)$ gives

Equation:

$$n - 1 = m \left(\frac{\lambda_0}{2t} \right) = 122 \left(\frac{632.8 \times 10^{-9} \text{ m}}{2(2 \times 10^{-2} \text{ m})} \right) = 0.0019$$

and $n = 1.0019$.

Significance

The indices of refraction for gases are so close to that of vacuum, that we normally consider them equal to 1. The difference between 1 and 1.0019 is so small that measuring it requires a correspondingly sensitive technique such as interferometry. We cannot, for example, hope to measure this value using techniques based simply on Snell's law.

Note:

Exercise:

Problem:

Check Your Understanding Although m , the number of fringes observed, is an integer, which is often regarded as having zero uncertainty, in practical terms, it is all too easy to lose track when counting fringes. In [\[link\]](#), if you estimate that you might have missed as many as five fringes when you reported $m = 122$ fringes, (a) is the value for the index of refraction worked out in [\[link\]](#) too large or too small? (b) By how much?

Solution:

a. too small; b. up to 8×10^{-5}

Note:

Wave Optics

Step 1. *Examine the situation to determine that interference is involved.*

Identify whether slits, thin films, or interferometers are considered in the problem.

Step 2. *If slits are involved*, note that diffraction gratings and double slits produce very similar interference patterns, but that gratings have narrower (sharper) maxima. Single-slit patterns are characterized by a large central maximum and smaller maxima to the sides.

Step 3. *If thin-film interference or an interferometer is involved, take note of the path length difference between the two rays that interfere. Be certain to use the wavelength in the medium involved, since it differs from the wavelength in vacuum. Note also that there is an additional $\lambda/2$ phase shift when light reflects from a medium with a greater index of refraction.*

Step 4. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Draw a diagram of the situation. Labeling the diagram is useful.*

Step 5. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).*

Step 6. *Solve the appropriate equation for the quantity to be determined (the unknown) and enter the knowns. Slits, gratings, and the Rayleigh limit involve equations.*

Step 7. *For thin-film interference, you have constructive interference for a total shift that is an integral number of wavelengths. You have destructive interference for a total shift of a half-integral number of wavelengths.*

Always keep in mind that crest to crest is constructive whereas crest to trough is destructive.

Step 8. *Check to see if the answer is reasonable: Does it make sense? Angles in interference patterns cannot be greater than 90° , for example.*

Summary

- When the mirror in one arm of the interferometer moves a distance of $\lambda/2$ each fringe in the interference pattern moves to the position previously occupied by the adjacent fringe.

Key Equations

Constructive	$\Delta l = m\lambda, \text{ for } m = 0, \pm 1, \pm 2, \pm 3 \dots$
--------------	--

interference	
Destructive interference	$\Delta l = (m + \frac{1}{2})\lambda, \text{ for } m = 0, \pm 1, \pm 2, \pm 3 \dots$
Path length difference for waves from two slits to a common point on a screen	$\Delta l = d \sin \theta$
Constructive interference	$d \sin \theta = m\lambda, \text{ for } m = 0, \pm 1, \pm 2, \pm 3, \dots$
Destructive interference	$d \sin \theta = (m + \frac{1}{2})\lambda, \text{ for } m = 0, \pm 1, \pm 2, \pm 3, \dots$
Distance from central maximum to the m th bright fringe	$y_m = \frac{m\lambda D}{d}$
Displacement measured by a Michelson interferometer	$\Delta d = m \frac{\lambda_0}{2}$

Conceptual Questions

Exercise:

Problem:

Describe how a Michelson interferometer can be used to measure the index of refraction of a gas (including air).

Solution:

In one arm, place a transparent chamber to be filled with the gas. See [\[link\]](#).

Problems**Exercise:****Problem:**

A Michelson interferometer has two equal arms. A mercury light of wavelength 546 nm is used for the interferometer and stable fringes are found. One of the arms is moved by $1.5\mu\text{m}$. How many fringes will cross the observing field?

Exercise:**Problem:**

What is the distance moved by the traveling mirror of a Michelson interferometer that corresponds to 1500 fringes passing by a point of the observation screen? Assume that the interferometer is illuminated with a 606 nm spectral line of krypton-86.

Solution:

$$4.55 \times 10^{-4} \text{ m}$$

Exercise:

Problem:

When the traveling mirror of a Michelson interferometer is moved $2.40 \times 10^{-5} \text{ m}$, 90 fringes pass by a point on the observation screen. What is the wavelength of the light used?

Exercise:**Problem:**

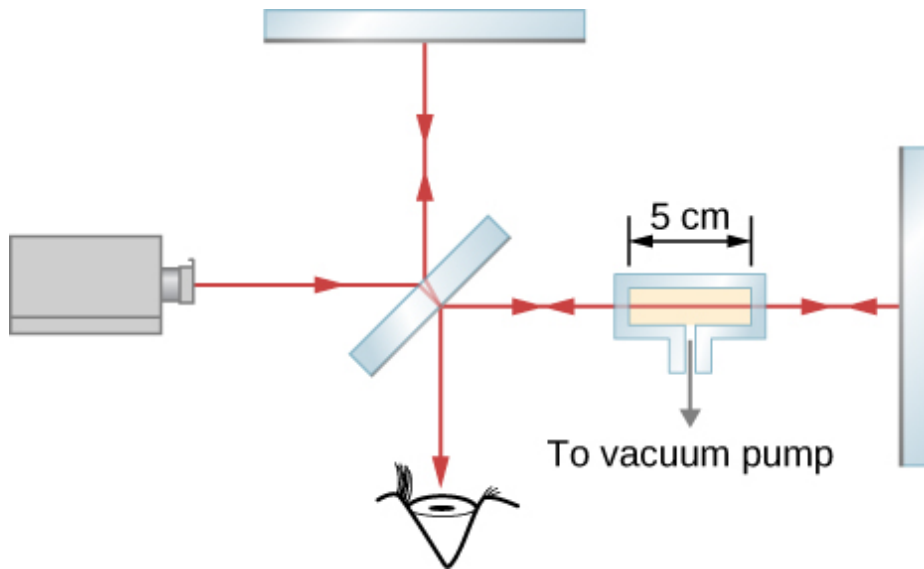
In a Michelson interferometer, light of wavelength 632.8 nm from a He-Ne laser is used. When one of the mirrors is moved by a distance D , 8 fringes move past the field of view. What is the value of the distance D ?

Solution:

$$D = 2.53 \times 10^{-6} \text{ m}$$

Exercise:**Problem:**

A chamber 5.0 cm long with flat, parallel windows at the ends is placed in one arm of a Michelson interferometer (see below). The light used has a wavelength of 500 nm in a vacuum. While all the air is being pumped out of the chamber, 29 fringes pass by a point on the observation screen. What is the refractive index of the air?



Additional Problems

Exercise:

Problem:

For 600-nm wavelength light and a slit separation of 0.12 mm, what are the angular positions of the first and third maxima in the double slit interference pattern?

Solution:

0.29° and 0.86°

Exercise:

Problem:

If the light source in the preceding problem is changed, the angular position of the third maximum is found to be 0.57° . What is the wavelength of light being used now?

Exercise:

Problem:

Red light ($\lambda = 710. \text{ nm}$) illuminates double slits separated by a distance $d = 0.150 \text{ mm}$. The screen and the slits are 3.00 m apart. (a) Find the distance on the screen between the central maximum and the third maximum. (b) What is the distance between the second and the fourth maxima?

Solution:

a. 4.26 cm ; b. 2.84 cm

Exercise:**Problem:**

Two sources as in phase and emit waves with $\lambda = 0.42 \text{ m}$. Determine whether constructive or destructive interference occurs at points whose distances from the two sources are (a) 0.84 and 0.42 m , (b) 0.21 and 0.42 m , (c) 1.26 and 0.42 m , (d) 1.87 and 1.45 m , (e) 0.63 and 0.84 m and (f) 1.47 and 1.26 m .

Exercise:**Problem:**

Two slits $4.0 \times 10^{-6} \text{ m}$ apart are illuminated by light of wavelength 600 nm . What is the highest order fringe in the interference pattern?

Solution:

6

Exercise:**Problem:**

Suppose that the highest order fringe that can be observed is the eighth in a double-slit experiment where 550-nm wavelength light is used. What is the minimum separation of the slits?

Exercise:

Problem:

The interference pattern of a He-Ne laser light ($\lambda = 632.9 \text{ nm}$) passing through two slits 0.031 mm apart is projected on a screen 10.0 m away. Determine the distance between the adjacent bright fringes.

Solution:

0.20 m

Exercise:**Problem:**

Young's double-slit experiment is performed immersed in water ($n = 1.333$). The light source is a He-Ne laser, $\lambda = 632.9 \text{ nm}$ in vacuum. (a) What is the wavelength of this light in water? (b) What is the angle for the third order maximum for two slits separated by 0.100 mm .

Exercise:**Problem:**

A double-slit experiment is to be set up so that the bright fringes appear 1.27 cm apart on a screen 2.13 m away from the two slits. The light source was wavelength 500 nm . What should be the separation between the two slits?

Solution:

0.0839 mm

Exercise:

Problem:

An effect analogous to two-slit interference can occur with sound waves, instead of light. In an open field, two speakers placed 1.30 m apart are powered by a single-function generator producing sine waves at 1200-Hz frequency. A student walks along a line 12.5 m away and parallel to the line between the speakers. She hears an alternating pattern of loud and quiet, due to constructive and destructive interference. What is (a) the wavelength of this sound and (b) the distance between the central maximum and the first maximum (loud) position along this line?

Exercise:**Problem:**

A hydrogen gas discharge lamp emits visible light at four wavelengths, $\lambda = 410, 434, 486, \text{ and } 656 \text{ nm}$. (a) If light from this lamp falls on a N slits separated by 0.025 mm, how far from the central maximum are the third maxima when viewed on a screen 2.0 m from the slits? (b) By what distance are the second and third maxima separated for $\lambda = 486 \text{ nm}$?

Solution:

a. 9.8, 10.4, 11.7, and 15.7 cm; b. 3.9 cm

Exercise:**Problem:**

Monochromatic light of frequency $5.5 \times 10^{14} \text{ Hz}$ falls on 10 slits separated by 0.020 mm. What is the separation between the first and third maxima on a screen that is 2.0 m from the slits?

Exercise:

Problem:

Eight slits equally separated by 0.149 mm is uniformly illuminated by a monochromatic light at $\lambda = 523$ nm. What is the width of the central principal maximum on a screen 2.35 m away?

Solution:

0.0575°

Exercise:**Problem:**

Eight slits equally separated by 0.149 mm is uniformly illuminated by a monochromatic light at $\lambda = 523$ nm. What is the intensity of a secondary maxima compared to that of the principal maxima?

Exercise:**Problem:**

A transparent film of thickness 250 nm and index of refraction of 1.40 is surrounded by air. What wavelength in a beam of white light at near-normal incidence to the film undergoes destructive interference when reflected?

Solution:

700 nm

Exercise:**Problem:**

An intensity minimum is found for 450 nm light transmitted through a transparent film ($n = 1.20$) in air. (a) What is minimum thickness of the film? (b) If this wavelength is the longest for which the intensity minimum occurs, what are the next three lower values of λ for which this happens?

Exercise:

Problem:

A thin film with $n = 1.32$ is surrounded by air. What is the minimum thickness of this film such that the reflection of normally incident light with $\lambda = 500 \text{ nm}$ is minimized?

Solution:

189 nm

Exercise:**Problem:**

Repeat your calculation of the previous problem with the thin film placed on a flat glass ($n = 1.50$) surface.

Exercise:**Problem:**

After a minor oil spill, a thin film of oil ($n = 1.40$) of thickness 450 nm floats on the water surface in a bay. (a) What predominant color is seen by a bird flying overhead? (b) What predominant color is seen by a seal swimming underwater?

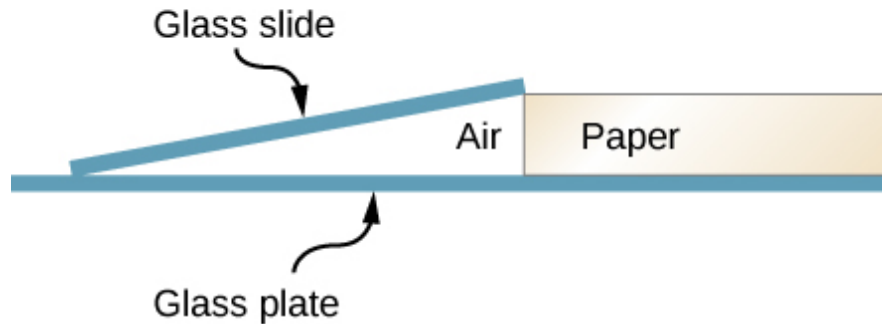
Solution:

a. green (504 nm); b. magenta (white minus green)

Exercise:**Problem:**

A microscope slide 10 cm long is separated from a glass plate at one end by a sheet of paper. As shown below, the other end of the slide is in contact with the plate. The slide is illuminated from above by light from a sodium lamp ($\lambda = 589 \text{ nm}$), and 14 fringes per centimeter are seen along the slide. What is the thickness of the piece of paper?

(Not to scale)



Exercise:

Problem:

Suppose that the setup of the preceding problem is immersed in an unknown liquid. If 18 fringes per centimeter are now seen along the slide, what is the index of refraction of the liquid?

Solution:

1.29

Exercise:

Problem:

A thin wedge filled with air is produced when two flat glass plates are placed on top of one another and a slip of paper is inserted between them at one edge. Interference fringes are observed when monochromatic light falling vertically on the plates are seen in reflection. Is the first fringe near the edge where the plates are in contact a bright fringe or a dark fringe? Explain.

Exercise:

Problem:

Two identical pieces of rectangular plate glass are used to measure the thickness of a hair. The glass plates are in direct contact at one edge and a single hair is placed between them near the opposite edge. When illuminated with a sodium lamp ($\lambda = 589 \text{ nm}$), the hair is seen between the 180th and 181st dark fringes. What are the lower and upper limits on the hair's diameter?

Solution:

$52.7 \mu\text{m}$ and $53.0 \mu\text{m}$

Exercise:**Problem:**

Two microscope slides made of glass are illuminated by monochromatic ($\lambda = 589 \text{ nm}$) light incident perpendicularly. The top slide touches the bottom slide at one end and rests on a thin copper wire at the other end, forming a wedge of air. The diameter of the copper wire is $29.45 \mu\text{m}$. How many bright fringes are seen across these slides?

Exercise:**Problem:**

A good quality camera “lens” is actually a system of lenses, rather than a single lens, but a side effect is that a reflection from the surface of one lens can bounce around many times within the system, creating artifacts in the photograph. To counteract this problem, one of the lenses in such a system is coated with a thin layer of material ($n = 1.28$) on one side. The index of refraction of the lens glass is 1.68. What is the smallest thickness of the coating that reduces the reflection at 640 nm by destructive interference? (In other words, the coating's effect is to be optimized for $\lambda = 640 \text{ nm}$.)

Solution:

125 nm

Exercise:

Problem:

Constructive interference is observed from directly above an oil slick for wavelengths (in air) 440 nm and 616 nm. The index of refraction of this oil is $n = 1.54$. What is the film's minimum possible thickness?

Exercise:

Problem:

A soap bubble is blown outdoors. What colors (indicate by wavelengths) of the reflected sunlight are seen enhanced? The soap bubble has index of refraction 1.36 and thickness 380 nm.

Solution:

413 nm and 689 nm

Exercise:

Problem:

A Michelson interferometer with a He-Ne laser light source ($\lambda = 632.8$ nm) projects its interference pattern on a screen. If the movable mirror is caused to move by $8.54 \mu\text{m}$, how many fringes will be observed shifting through a reference point on a screen?

Exercise:

Problem:

An experimenter detects 251 fringes when the movable mirror in a Michelson interferometer is displaced. The light source used is a sodium lamp, wavelength 589 nm. By what distance did the movable mirror move?

Solution:

$73.9 \mu\text{m}$

Exercise:**Problem:**

A Michelson interferometer is used to measure the wavelength of light put through it. When the movable mirror is moved by exactly 0.100 mm, the number of fringes observed moving through is 316. What is the wavelength of the light?

Exercise:**Problem:**

A 5.08-cm-long rectangular glass chamber is inserted into one arm of a Michelson interferometer using a 633-nm light source. This chamber is initially filled with air ($n = 1.000293$) at standard atmospheric pressure but the air is gradually pumped out using a vacuum pump until a near perfect vacuum is achieved. How many fringes are observed moving by during the transition?

Solution:

47

Exercise:**Problem:**

Into one arm of a Michelson interferometer, a plastic sheet of thickness $75\ \mu\text{m}$ is inserted, which causes a shift in the interference pattern by 86 fringes. The light source has wavelength of 610 nm in air. What is the index of refraction of this plastic?

Exercise:**Problem:**

The thickness of an aluminum foil is measured using a Michelson interferometer that has its movable mirror mounted on a micrometer. There is a difference of 27 fringes in the observed interference pattern when the micrometer clamps down on the foil compared to when the micrometer is empty. Calculate the thickness of the foil?

Solution:

$8.5\ \mu\text{m}$

Exercise:**Problem:**

The movable mirror of a Michelson interferometer is attached to one end of a thin metal rod of length 23.3 mm. The other end of the rod is anchored so it does not move. As the temperature of the rod changes from $15\ ^\circ\text{C}$ to $25\ ^\circ\text{C}$, a change of 14 fringes is observed. The light source is a He Ne laser, $\lambda = 632.8\ \text{nm}$. What is the change in length of the metal bar, and what is its thermal expansion coefficient?

Exercise:**Problem:**

In a thermally stabilized lab, a Michelson interferometer is used to monitor the temperature to ensure it stays constant. The movable mirror is mounted on the end of a 1.00-m-long aluminum rod, held fixed at the other end. The light source is a He Ne laser, $\lambda = 632.8\ \text{nm}$. The resolution of this apparatus corresponds to the temperature difference when a change of just one fringe is observed. What is this temperature difference?

Solution:

$0.013\ ^\circ\text{C}$

Exercise:**Problem:**

A 65-fringe shift results in a Michelson interferometer when a $42.0\text{-}\mu\text{m}$ film made of an unknown material is placed in one arm. The light source has wavelength $632.9\ \text{nm}$. Identify the material using the indices of refraction found in [\[link\]](#).

Challenge Problems

Exercise:

Problem:

Determine what happens to the double-slit interference pattern if one of the slits is covered with a thin, transparent film whose thickness is $\lambda/[2(n - 1)]$, where λ is the wavelength of the incident light and n is the index of refraction of the film.

Solution:

Bright and dark fringes switch places.

Exercise:

Problem:

Fifty-one narrow slits are equally spaced and separated by 0.10 mm. The slits are illuminated by blue light of wavelength 400 nm. What is angular position of the twenty-fifth secondary maximum? What is its peak intensity in comparison with that of the primary maximum?

Exercise:

Problem:

A film of oil on water will appear dark when it is very thin, because the path length difference becomes small compared with the wavelength of light and there is a phase shift at the top surface. If it becomes dark when the path length difference is less than one-fourth the wavelength, what is the thickest the oil can be and appear dark at all visible wavelengths? Oil has an index of refraction of 1.40.

Solution:

The path length must be less than one-fourth of the shortest visible wavelength in oil. The thickness of the oil is half the path length, so it must be less than one-eighth of the shortest visible wavelength in oil. If we take 380 nm to be the shortest visible wavelength in air, 33.9 nm.

Exercise:**Problem:**

[\[link\]](#) shows two glass slides illuminated by monochromatic light incident perpendicularly. The top slide touches the bottom slide at one end and rests on a 0.100-mm-diameter hair at the other end, forming a wedge of air. (a) How far apart are the dark bands, if the slides are 7.50 cm long and 589-nm light is used? (b) Is there any difference if the slides are made from crown or flint glass? Explain.

Exercise:**Problem:**

[\[link\]](#) shows two 7.50-cm-long glass slides illuminated by pure 589-nm wavelength light incident perpendicularly. The top slide touches the bottom slide at one end and rests on some debris at the other end, forming a wedge of air. How thick is the debris, if the dark bands are 1.00 mm apart?

Solution:

$$4.42 \times 10^{-5} \text{ m}$$

Exercise:**Problem:**

A soap bubble is 100 nm thick and illuminated by white light incident at a 45° angle to its surface. What wavelength and color of visible light is most constructively reflected, assuming the same index of refraction as water?

Exercise:**Problem:**

An oil slick on water is 120 nm thick and illuminated by white light incident at a 45° angle to its surface. What color does the oil appear (what is the most constructively reflected wavelength), given its index of refraction is 1.40?

Solution:

for one phase change: 950 nm (infrared); for three phase changes: 317 nm (ultraviolet); Therefore, the oil film will appear black, since the reflected light is not in the visible part of the spectrum.

Glossary

interferometer

instrument that uses interference of waves to make measurements

Introduction

class="introduction"

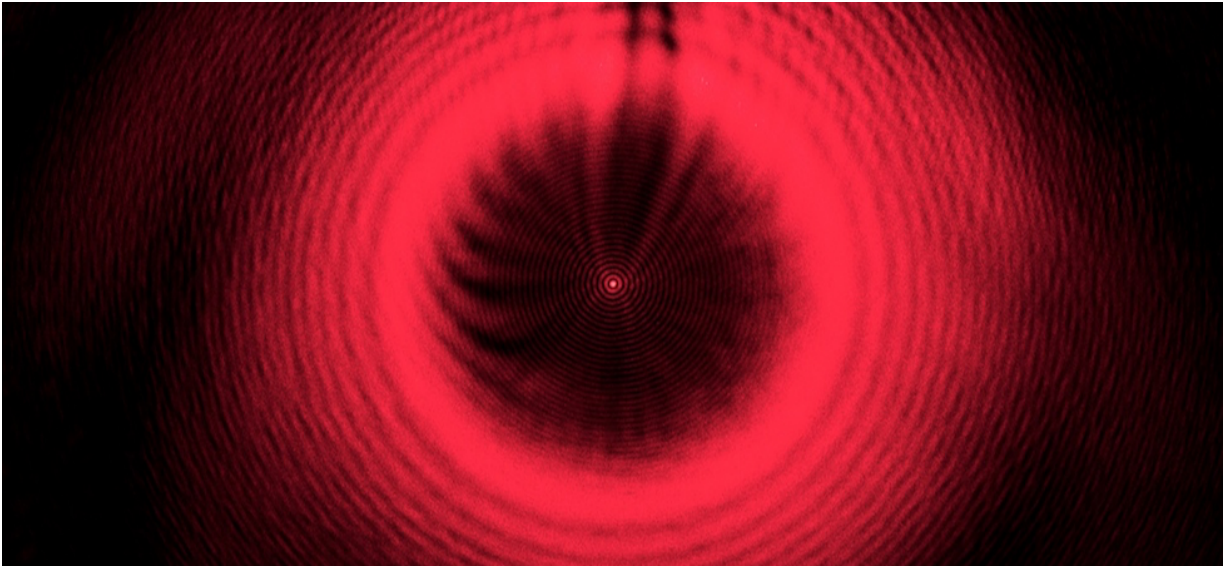
A steel ball
bearing
illuminated by a
laser does not
cast a sharp,
circular shadow.
Instead, a series
of diffraction
fringes and a
central bright
spot are
observed.

Known as
Poisson's spot,
the effect was
first predicted
by Augustin-
Jean Fresnel
(1788–1827) as
a consequence
of diffraction of
light waves.

Based on
principles of ray
optics, Siméon-
Denis Poisson
(1781–1840)
argued against
Fresnel's
prediction.

(credit:
modification of
work by
Harvard Natural

Science Lecture
Demonstrations
)



Imagine passing a monochromatic light beam through a narrow opening—a slit just a little wider than the wavelength of the light. Instead of a simple shadow of the slit on the screen, you will see that an interference pattern appears, even though there is only one slit.

In the chapter on interference, we saw that you need two sources of waves for interference to occur. How can there be an interference pattern when we have only one slit? In [The Nature of Light](#), we learned that, due to Huygens's principle, we can imagine a wave front as equivalent to infinitely many point sources of waves. Thus, a wave from a slit can behave not as one wave but as an infinite number of point sources. These waves can interfere with each other, resulting in an interference pattern without the presence of a second slit. This phenomenon is called *diffraction*.

Another way to view this is to recognize that a slit has a small but finite width. In the preceding chapter, we implicitly regarded slits as objects with positions but no size. The widths of the slits were considered negligible. When the slits have finite widths, each point along the opening can be considered a point source of light—a foundation of Huygens's principle.

Because real-world optical instruments must have finite apertures (otherwise, no light can enter), diffraction plays a major role in the way we interpret the output of these optical instruments. For example, diffraction places limits on our ability to resolve images or objects. This is a problem that we will study later in this chapter.

Single-Slit Diffraction

By the end of this section, you will be able to:

- Explain the phenomenon of diffraction and the conditions under which it is observed
- Describe diffraction through a single slit

After passing through a narrow aperture (opening), a wave propagating in a specific direction tends to spread out. For example, sound waves that enter a room through an open door can be heard even if the listener is in a part of the room where the geometry of ray propagation dictates that there should only be silence. Similarly, ocean waves passing through an opening in a breakwater can spread throughout the bay inside. ([link](#)). The spreading and bending of sound and ocean waves are two examples of **diffraction**, which is the bending of a wave around the edges of an opening or an obstacle—a phenomenon exhibited by all types of waves.

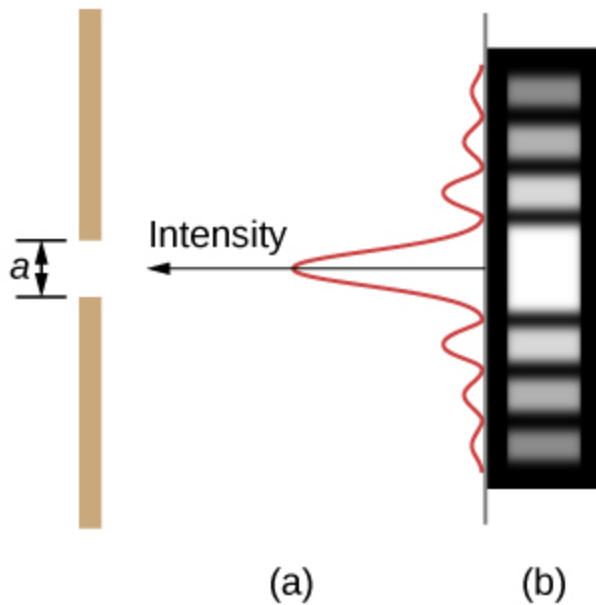


Because of the diffraction of waves, ocean waves entering through an opening in a breakwater can spread throughout the bay. (credit: modification of map data from Google Earth)

The diffraction of sound waves is apparent to us because wavelengths in the audible region are approximately the same size as the objects they encounter, a condition that must be satisfied if diffraction effects are to be observed easily. Since the wavelengths of visible light range from approximately 390 to 770 nm, most objects do not diffract light significantly. However, situations do occur in which apertures are small enough that the diffraction of light is observable. For example, if you place your middle and index fingers close together and look through the opening at a light bulb, you can see a rather clear diffraction pattern, consisting of light and dark lines running parallel to your fingers.

Diffraction through a Single Slit

Light passing through a single slit forms a diffraction pattern somewhat different from those formed by double slits or diffraction gratings, which we discussed in the chapter on interference. [\[link\]](#) shows a single-slit diffraction pattern. Note that the central maximum is larger than maxima on either side and that the intensity decreases rapidly on either side. In contrast, a diffraction grating ([Diffraction Gratings](#)) produces evenly spaced lines that dim slowly on either side of the center.



Single-slit diffraction pattern. (a)

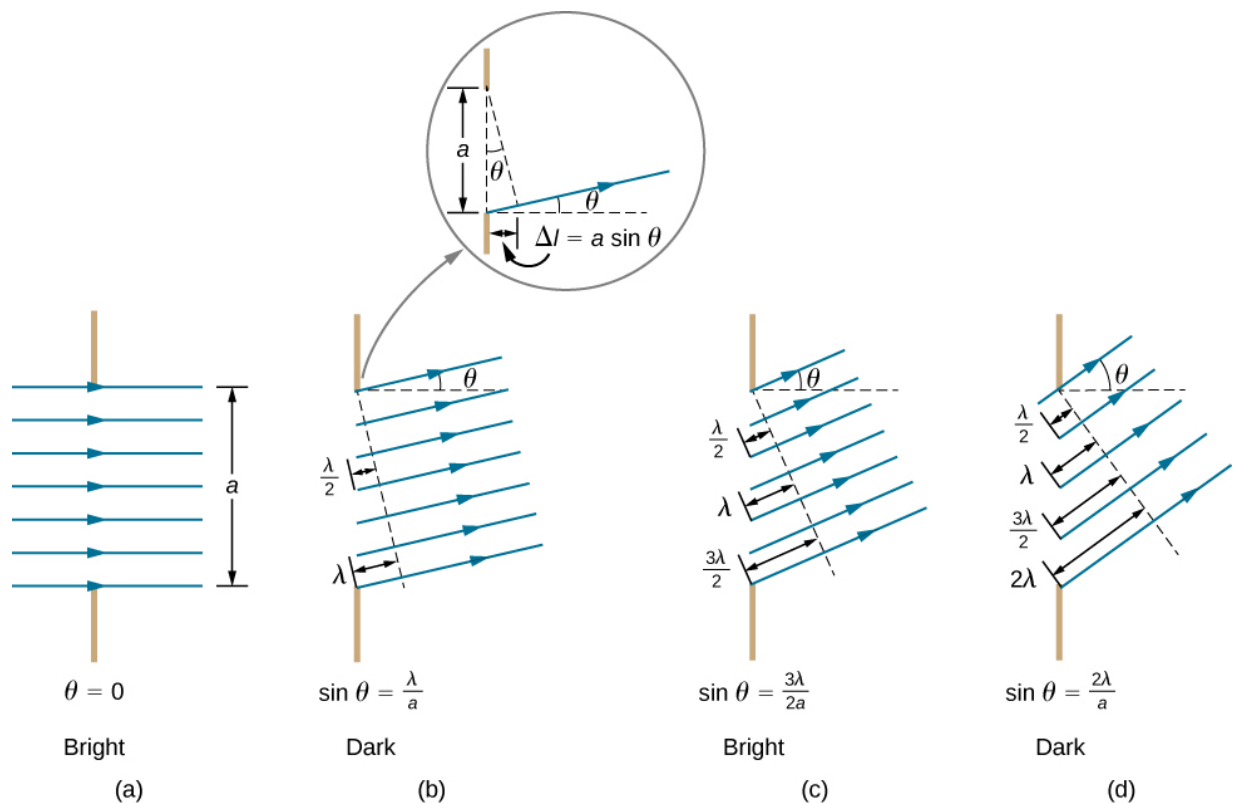
Monochromatic light passing through a single slit has a central maximum and many smaller and dimmer maxima on either side.

The central maximum is six times higher than shown. (b)

The diagram shows the bright central maximum, and the dimmer and thinner maxima on either side.

The analysis of single-slit diffraction is illustrated in [\[link\]](#). Here, the light arrives at the slit, illuminating it uniformly and is in phase across its width. We then consider light propagating onwards from different parts of the *same* slit. According to Huygens's principle, every part of the wave front in the slit emits wavelets, as we discussed in [The Nature of Light](#). These are like rays that start out in phase and head in all directions. (Each ray is perpendicular to the wave front of a wavelet.) Assuming the screen is very far away compared with the size of the slit, rays heading toward a common destination are nearly parallel. When they travel straight ahead, as in part

(a) of the figure, they remain in phase, and we observe a central maximum. However, when rays travel at an angle θ relative to the original direction of the beam, each ray travels a different distance to a common location, and they can arrive in or out of phase. In part (b), the ray from the bottom travels a distance of one wavelength λ farther than the ray from the top. Thus, a ray from the center travels a distance $\lambda/2$ less than the one at the bottom edge of the slit, arrives out of phase, and interferes destructively. A ray from slightly above the center and one from slightly above the bottom also cancel one another. In fact, each ray from the slit interferes destructively with another ray. In other words, a pair-wise cancellation of all rays results in a dark minimum in intensity at this angle. By symmetry, another minimum occurs at the same angle to the right of the incident direction (toward the bottom of the figure) of the light.



Light passing through a single slit is diffracted in all directions and may interfere constructively or destructively, depending on the angle.

The difference in path length for rays from either side of the slit is seen to be $a \sin \theta$.

At the larger angle shown in part (c), the path lengths differ by $3\lambda/2$ for rays from the top and bottom of the slit. One ray travels a distance λ different from the ray from the bottom and arrives in phase, interfering constructively. Two rays, each from slightly above those two, also add constructively. Most rays from the slit have another ray to interfere with constructively, and a maximum in intensity occurs at this angle. However, not all rays interfere constructively for this situation, so the maximum is not as intense as the central maximum. Finally, in part (d), the angle shown is large enough to produce a second minimum. As seen in the figure, the difference in path length for rays from either side of the slit is $a \sin \theta$, and we see that a destructive minimum is obtained when this distance is an integral multiple of the wavelength.

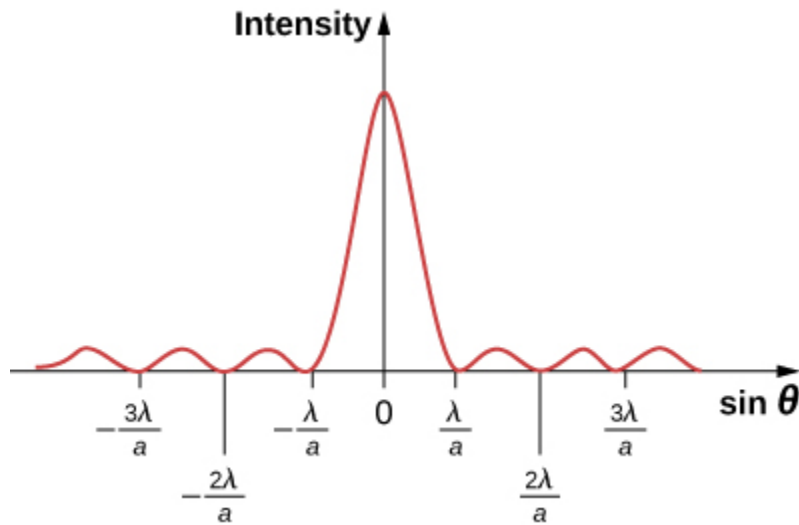
Thus, to obtain **destructive interference for a single slit**,

Note:

Equation:

$$a \sin \theta = m\lambda, \text{ for } m = \pm 1, \pm 2, \pm 3, \dots (\text{destructive}),$$

where a is the slit width, λ is the light's wavelength, θ is the angle relative to the original direction of the light, and m is the order of the minimum. [\[link\]](#) shows a graph of intensity for single-slit interference, and it is apparent that the maxima on either side of the central maximum are much less intense and not as wide. This effect is explored in [Double-Slit Diffraction](#).

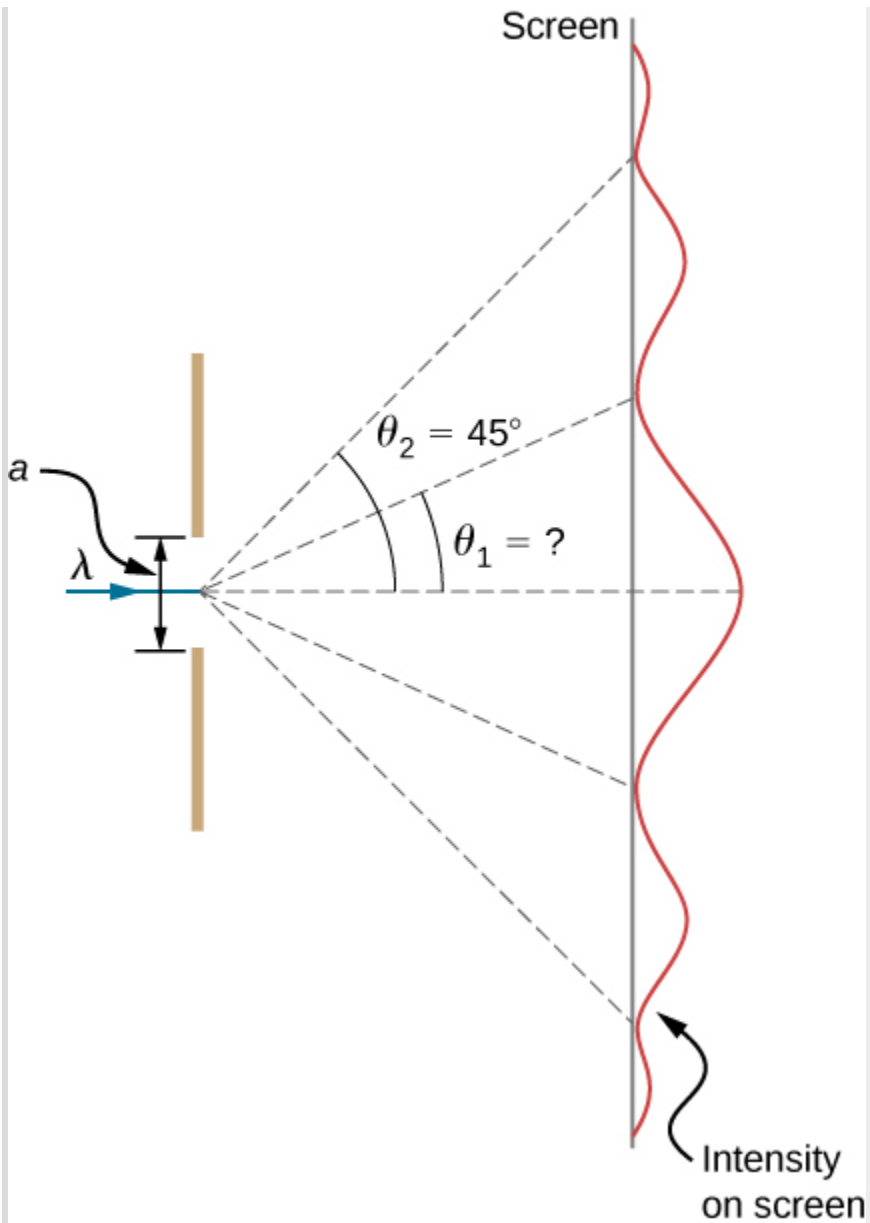


A graph of single-slit diffraction intensity showing the central maximum to be wider and much more intense than those to the sides. In fact, the central maximum is six times higher than shown here.

Example:

Calculating Single-Slit Diffraction

Visible light of wavelength 550 nm falls on a single slit and produces its second diffraction minimum at an angle of 45.0° relative to the incident direction of the light, as in [\[link\]](#). (a) What is the width of the slit? (b) At what angle is the first minimum produced?



In this example, we analyze a graph of the single-slit diffraction pattern.

Strategy

From the given information, and assuming the screen is far away from the slit, we can use the equation $a \sin \theta = m\lambda$ first to find D , and again to find the angle for the first minimum θ_1 .

Solution

- a. We are given that $\lambda = 550 \text{ nm}$, $m = 2$, and $\theta_2 = 45.0^\circ$. Solving the equation $a \sin \theta = m\lambda$ for a and substituting known values gives
Equation:

$$a = \frac{m\lambda}{\sin \theta_2} = \frac{2(550 \text{ nm})}{\sin 45.0^\circ} = \frac{1100 \times 10^{-9} \text{ m}}{0.707} = 1.56 \times 10^{-6} \text{ m}.$$

- b. Solving the equation $a \sin \theta = m\lambda$ for $\sin \theta_1$ and substituting the known values gives

Equation:

$$\sin \theta_1 = \frac{m\lambda}{a} = \frac{1(550 \times 10^{-9} \text{ m})}{1.56 \times 10^{-6} \text{ m}}.$$

Thus the angle θ_1 is

Equation:

$$\theta_1 = \sin^{-1} 0.354 = 20.7^\circ.$$

Significance

We see that the slit is narrow (it is only a few times greater than the wavelength of light). This is consistent with the fact that light must interact with an object comparable in size to its wavelength in order to exhibit significant wave effects such as this single-slit diffraction pattern. We also see that the central maximum extends 20.7° on either side of the original beam, for a width of about 41° . The angle between the first and second minima is only about 24° ($45.0^\circ - 20.7^\circ$). Thus, the second maximum is only about half as wide as the central maximum.

Note:

Exercise:

Problem:

Check Your Understanding Suppose the slit width in [\[link\]](#) is increased to 1.8×10^{-6} m. What are the new angular positions for the first, second, and third minima? Would a fourth minimum exist?

Solution:

$17.8^\circ, 37.7^\circ, 66.4^\circ$; no

Summary

- Diffraction can send a wave around the edges of an opening or other obstacle.
- A single slit produces an interference pattern characterized by a broad central maximum with narrower and dimmer maxima to the sides.

Conceptual Questions

Exercise:**Problem:**

As the width of the slit producing a single-slit diffraction pattern is reduced, how will the diffraction pattern produced change?

Solution:

The diffraction pattern becomes wider.

Exercise:

Problem: Compare interference and diffraction.

Exercise:

Problem:

If you and a friend are on opposite sides of a hill, you can communicate with walkie-talkies but not with flashlights. Explain.

Solution:

Walkie-talkies use radio waves whose wavelengths are comparable to the size of the hill and are thus able to diffract around the hill. Visible wavelengths of the flashlight travel as rays at this size scale.

Exercise:**Problem:**

What happens to the diffraction pattern of a single slit when the entire optical apparatus is immersed in water?

Exercise:**Problem:**

In our study of diffraction by a single slit, we assume that the length of the slit is much larger than the width. What happens to the diffraction pattern if these two dimensions were comparable?

Solution:

The diffraction pattern becomes two-dimensional, with main fringes, which are now spots, running in perpendicular directions and fainter spots in intermediate directions.

Exercise:**Problem:**

A rectangular slit is twice as wide as it is high. Is the central diffraction peak wider in the vertical direction or in the horizontal direction?

Problems

Exercise:**Problem:**

(a) At what angle is the first minimum for 550-nm light falling on a single slit of width $1.00\mu\text{m}$? (b) Will there be a second minimum?

Solution:

a. 33.4° ; b. no

Exercise:**Problem:**

(a) Calculate the angle at which a $2.00\text{-}\mu\text{m}$ -wide slit produces its first minimum for 410-nm violet light. (b) Where is the first minimum for 700-nm red light?

Exercise:**Problem:**

(a) How wide is a single slit that produces its first minimum for 633-nm light at an angle of 28.0° ? (b) At what angle will the second minimum be?

Solution:

a. $1.35 \times 10^{-6} \text{ m}$; b. 69.9°

Exercise:**Problem:**

(a) What is the width of a single slit that produces its first minimum at 60.0° for 600-nm light? (b) Find the wavelength of light that has its first minimum at 62.0° .

Exercise:

Problem:

Find the wavelength of light that has its third minimum at an angle of 48.6° when it falls on a single slit of width $3.00\mu\text{m}$.

Solution:

750 nm

Exercise:**Problem:**

(a) Sodium vapor light averaging 589 nm in wavelength falls on a single slit of width $7.50\mu\text{m}$. At what angle does it produces its second minimum? (b) What is the highest-order minimum produced?

Exercise:**Problem:**

Consider a single-slit diffraction pattern for $\lambda = 589\text{ nm}$, projected on a screen that is 1.00 m from a slit of width 0.25 mm. How far from the center of the pattern are the centers of the first and second dark fringes?

Solution:

2.4 mm, 4.7 mm

Exercise:**Problem:**

(a) Find the angle between the first minima for the two sodium vapor lines, which have wavelengths of 589.1 and 589.6 nm, when they fall upon a single slit of width $2.00\mu\text{m}$. (b) What is the distance between these minima if the diffraction pattern falls on a screen 1.00 m from the slit? (c) Discuss the ease or difficulty of measuring such a distance.

Exercise:

Problem:

(a) What is the minimum width of a single slit (in multiples of λ) that will produce a first minimum for a wavelength λ ? (b) What is its minimum width if it produces 50 minima? (c) 1000 minima?

Solution:

a. 1.00λ ; b. 50.0λ ; c. 1000λ

Exercise:**Problem:**

(a) If a single slit produces a first minimum at 14.5° , at what angle is the second-order minimum? (b) What is the angle of the third-order minimum? (c) Is there a fourth-order minimum? (d) Use your answers to illustrate how the angular width of the central maximum is about twice the angular width of the next maximum (which is the angle between the first and second minima).

Exercise:**Problem:**

If the separation between the first and the second minima of a single-slit diffraction pattern is 6.0 mm, what is the distance between the screen and the slit? The light wavelength is 500 nm and the slit width is 0.16 mm.

Solution:

1.92 m

Exercise:

Problem:

A water break at the entrance to a harbor consists of a rock barrier with a 50.0-m-wide opening. Ocean waves of 20.0-m wavelength approach the opening straight on. At what angles to the incident direction are the boats inside the harbor most protected against wave action?

Exercise:**Problem:**

An aircraft maintenance technician walks past a tall hangar door that acts like a single slit for sound entering the hangar. Outside the door, on a line perpendicular to the opening in the door, a jet engine makes a 600-Hz sound. At what angle with the door will the technician observe the first minimum in sound intensity if the vertical opening is 0.800 m wide and the speed of sound is 340 m/s?

Solution:

45.1°

Glossary

destructive interference for a single slit

occurs when the width of the slit is comparable to the wavelength of light illuminating it

diffraction

bending of a wave around the edges of an opening or an obstacle

Intensity in Single-Slit Diffraction

By the end of this section, you will be able to:

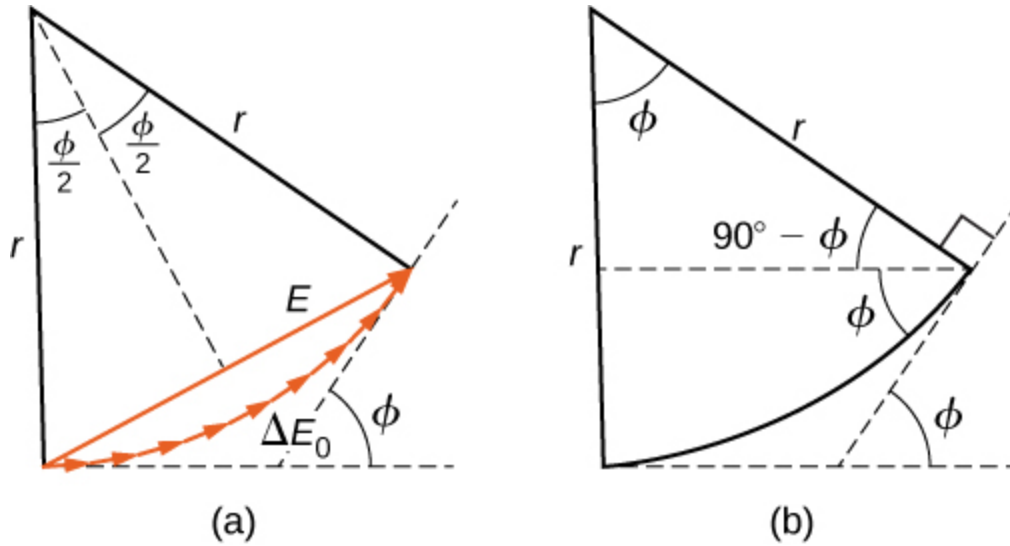
- Calculate the intensity relative to the central maximum of the single-slit diffraction peaks
- Calculate the intensity relative to the central maximum of an arbitrary point on the screen

To calculate the intensity of the diffraction pattern, we follow the phasor method used for calculations with ac circuits in [Alternating-Current Circuits](#). If we consider that there are N Huygens sources across the slit shown in [\[link\]](#), with each source separated by a distance a/N from its adjacent neighbors, the path difference between waves from adjacent sources reaching the arbitrary point P on the screen is $(a/N) \sin \theta$. This distance is equivalent to a phase difference of $(2\pi a/\lambda N) \sin \theta$. The phasor diagram for the waves arriving at the point whose angular position is θ is shown in [\[link\]](#). The amplitude of the phasor for each Huygens wavelet is ΔE_0 , the amplitude of the resultant phasor is E , and the phase difference between the wavelets from the first and the last sources is

Equation:

$$\phi = \left(\frac{2\pi}{\lambda} \right) a \sin \theta.$$

With $N \rightarrow \infty$, the phasor diagram approaches a circular arc of length $N\Delta E_0$ and radius r . Since the length of the arc is $N\Delta E_0$ for any ϕ , the radius r of the arc must decrease as ϕ increases (or equivalently, as the phasors form tighter spirals).



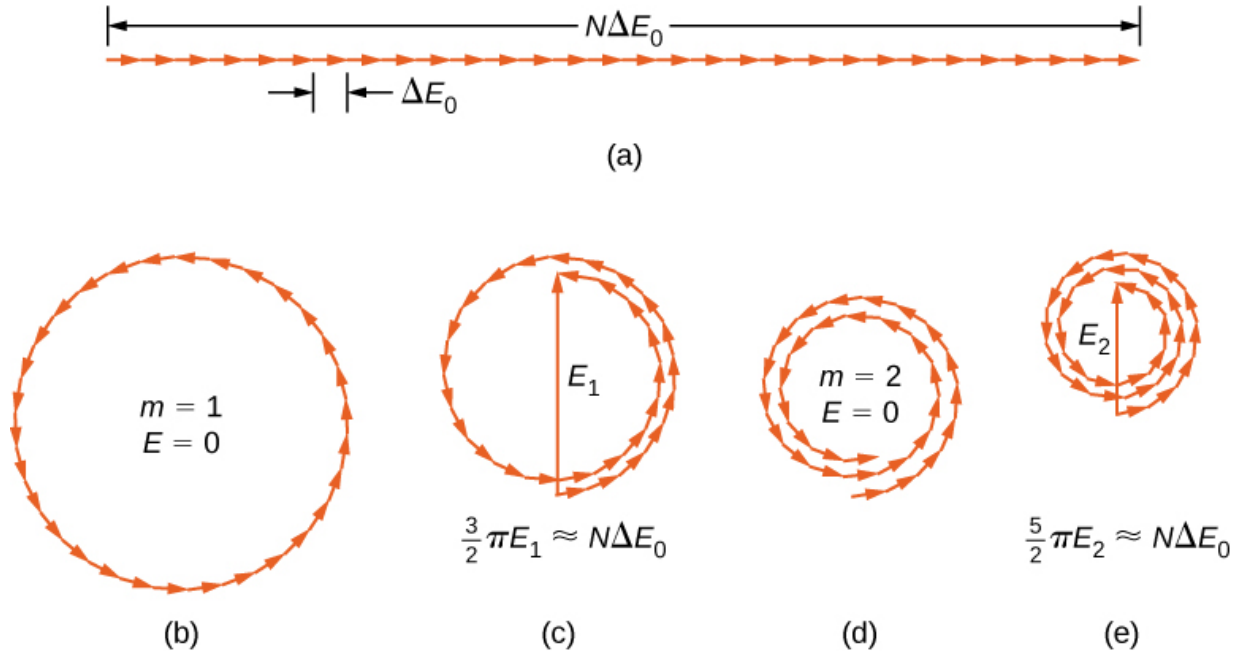
(a) Phasor diagram corresponding to the angular position θ in the single-slit diffraction pattern. The phase difference between the wavelets from the first and last sources is $\phi = (2\pi/\lambda)a \sin \theta$. (b) The geometry of the phasor diagram.

The phasor diagram for $\phi = 0$ (the center of the diffraction pattern) is shown in [\[link\]](#)(a) using $N = 30$. In this case, the phasors are laid end to end in a straight line of length $N\Delta E_0$, the radius r goes to infinity, and the resultant has its maximum value $E = N\Delta E_0$. The intensity of the light can be obtained using the relation $I = \frac{1}{2}c\epsilon_0 E^2$ from [Electromagnetic Waves](#). The intensity of the maximum is then

Equation:

$$I_0 = \frac{1}{2}c\epsilon_0(N\Delta E_0)^2 = \frac{1}{2\mu_0 c}(N\Delta E_0)^2,$$

where $\epsilon_0 = 1/\mu_0 c^2$. The phasor diagrams for the first two zeros of the diffraction pattern are shown in parts (b) and (d) of the figure. In both cases, the phasors add to zero, after rotating through $\phi = 2\pi$ rad for $m = 1$ and 4π rad for $m = 2$.



Phasor diagrams (with 30 phasors) for various points on the single-slit diffraction pattern. Multiple rotations around a given circle have been separated slightly so that the phasors can be seen. (a) Central maximum, (b) first minimum, (c) first maximum beyond central maximum, (d) second minimum, and (e) second maximum beyond central maximum.

The next two maxima beyond the central maxima are represented by the phasor diagrams of parts (c) and (e). In part (c), the phasors have rotated through $\phi = 3\pi$ rad and have formed a resultant phasor of magnitude E_1 . The length of the arc formed by the phasors is $N\Delta E_0$. Since this corresponds to 1.5 rotations around a circle of diameter E_1 , we have

Equation:

$$\frac{3}{2}\pi E_1 \approx N\Delta E_0,$$

so

Equation:

$$E_1 = \frac{2N\Delta E_0}{3\pi}$$

and

Equation:

$$I_1 = \frac{1}{2\mu_0 c} E_1^2 = \frac{4(N\Delta E_0)^2}{(9\pi^2)(2\mu_0 c)} \approx 0.045I_0,$$

where

Equation:

$$I_0 = \frac{(N\Delta E_0)^2}{2\mu_0 c}.$$

In part (e), the phasors have rotated through $\phi = 5\pi$ rad, corresponding to 2.5 rotations around a circle of diameter E_2 and arc length $N\Delta E_0$. This results in $I_2 \approx 0.016I_0$. The proof is left as an exercise for the student ([\[link\]](#)).

These two maxima actually correspond to values of ϕ slightly less than 3π rad and 5π rad. Since the total length of the arc of the phasor diagram is always $N\Delta E_0$, the radius of the arc decreases as ϕ increases. As a result, E_1 and E_2 turn out to be slightly larger for arcs that have not quite curled through 3π rad and 5π rad, respectively. The exact values of ϕ for the maxima are investigated in [\[link\]](#). In solving that problem, you will find that they are less than, but very close to, $\phi = 3\pi, 5\pi, 7\pi, \dots$ rad.

To calculate the intensity at an arbitrary point P on the screen, we return to the phasor diagram of [\[link\]](#). Since the arc subtends an angle ϕ at the center of the circle,

Equation:

$$N\Delta E_0 = r\phi$$

and

Equation:

$$\sin \left(\frac{\phi}{2} \right) = \frac{E}{2r}.$$

where E is the amplitude of the resultant field. Solving the second equation for E and then substituting r from the first equation, we find

Equation:

$$E = 2r \sin \frac{\phi}{2} = 2 \frac{N \Delta E_o}{\phi} \sin \frac{\phi}{2}.$$

Now defining

Note:

Equation:

$$\beta = \frac{\phi}{2} = \frac{\pi a \sin \theta}{\lambda}$$

we obtain

Note:

Equation:

$$E = N \Delta E_0 \frac{\sin \beta}{\beta}$$

This equation relates the amplitude of the resultant field at any point in the diffraction pattern to the amplitude $N\Delta E_0$ at the central maximum. The intensity is proportional to the square of the amplitude, so

Note:

Equation:

$$I = I_0 \left(\frac{\sin \beta}{\beta} \right)^2$$

where $I_0 = (N\Delta E_0)^2 / 2\mu_0 c$ is the intensity at the center of the pattern.

For the central maximum, $\phi = 0$, β is also zero and we see from l'Hôpital's rule that $\lim_{\beta \rightarrow 0} (\sin \beta / \beta) = 1$, so that $\lim_{\phi \rightarrow 0} I = I_0$. For the next maximum, $\phi = 3\pi$ rad, we have $\beta = 3\pi/2$ rad and when substituted into [\[link\]](#), it yields

Equation:

$$I_1 = I_0 \left(\frac{\sin 3\pi/2}{3\pi/2} \right)^2 \approx 0.045 I_0,$$

in agreement with what we found earlier in this section using the diameters and circumferences of phasor diagrams. Substituting $\phi = 5\pi$ rad into [\[link\]](#) yields a similar result for I_2 .

A plot of [\[link\]](#) is shown in [\[link\]](#) and directly below it is a photograph of an actual diffraction pattern. Notice that the central peak is much brighter than the others, and that the zeros of the pattern are located at those points where $\sin \beta = 0$, which occurs when $\beta = m\pi$ rad. This corresponds to

Equation:

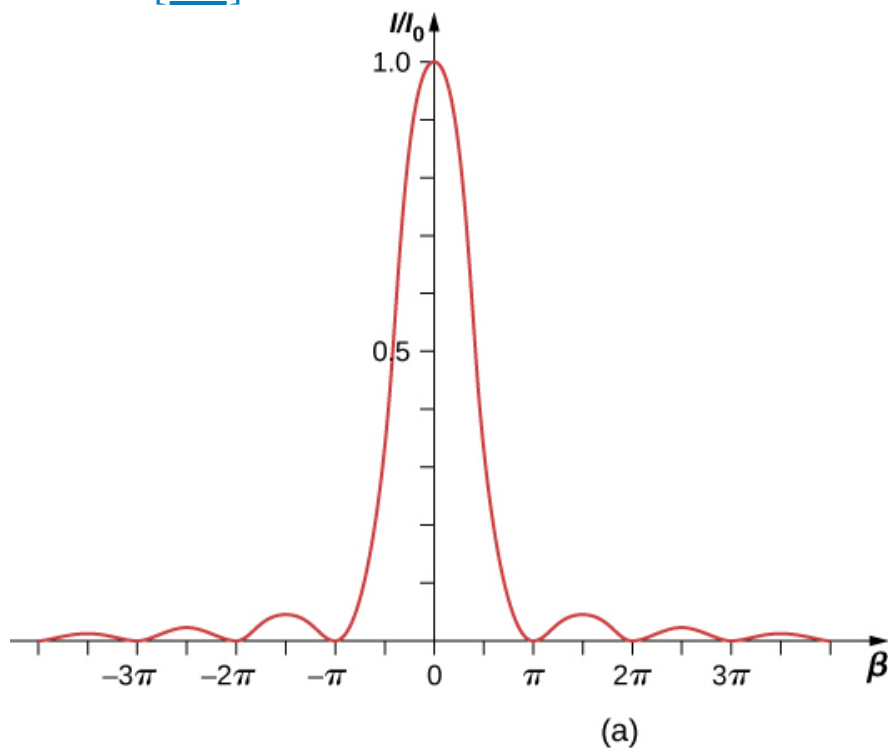
$$\frac{\pi a \sin \theta}{\lambda} = m\pi,$$

or

Equation:

$$a \sin \theta = m\lambda,$$

which is [\[link\]](#).



(b)

(a) The calculated intensity distribution of a single-slit diffraction pattern. (b) The actual diffraction pattern.

Example:**Intensity in Single-Slit Diffraction**

Light of wavelength 550 nm passes through a slit of width 2.00 μm and produces a diffraction pattern similar to that shown in [\[link\]](#). (a) Find the locations of the first two minima in terms of the angle from the central maximum and (b) determine the intensity relative to the central maximum at a point halfway between these two minima.

Strategy

The minima are given by [\[link\]](#), $a \sin \theta = m\lambda$. The first two minima are for $m = 1$ and $m = 2$. [\[link\]](#) and [\[link\]](#) can be used to determine the intensity once the angle has been worked out.

Solution

- a. Solving [\[link\]](#) for θ gives us $\theta_m = \sin^{-1}(m\lambda/a)$, so that

Equation:

$$\theta_1 = \sin^{-1} \left(\frac{(+1) (550 \times 10^{-9} \text{ m})}{2.00 \times 10^{-6} \text{ m}} \right) = +16.0^\circ$$

and

Equation:

$$\theta_2 = \sin^{-1} \left(\frac{(+2) (550 \times 10^{-9} \text{ m})}{2.00 \times 10^{-6} \text{ m}} \right) = +33.4^\circ.$$

- b. The halfway point between θ_1 and θ_2 is

Equation:

$$\theta = (\theta_1 + \theta_2)/2 = (16.0^\circ + 33.4^\circ)/2 = 24.7^\circ.$$

[\[link\]](#) gives

Equation:

$$\beta = \frac{\pi a \sin \theta}{\lambda} = \frac{\pi (2.00 \times 10^{-6} \text{ m}) \sin (24.7^\circ)}{(550 \times 10^{-9} \text{ m})} = 1.52\pi \text{ or } 4.77 \text{ rad.}$$

From [\[link\]](#), we can calculate

Equation:

$$\frac{I}{I_o} = \left(\frac{\sin \beta}{\beta} \right)^2 = \left(\frac{\sin (4.77)}{4.77} \right)^2 = \left(\frac{-0.9985}{4.77} \right)^2 = 0.044.$$

Significance

This position, halfway between two minima, is very close to the location of the maximum, expected near $\beta = 3\pi/2$, or 1.5π .

Note:

Exercise:

Problem:

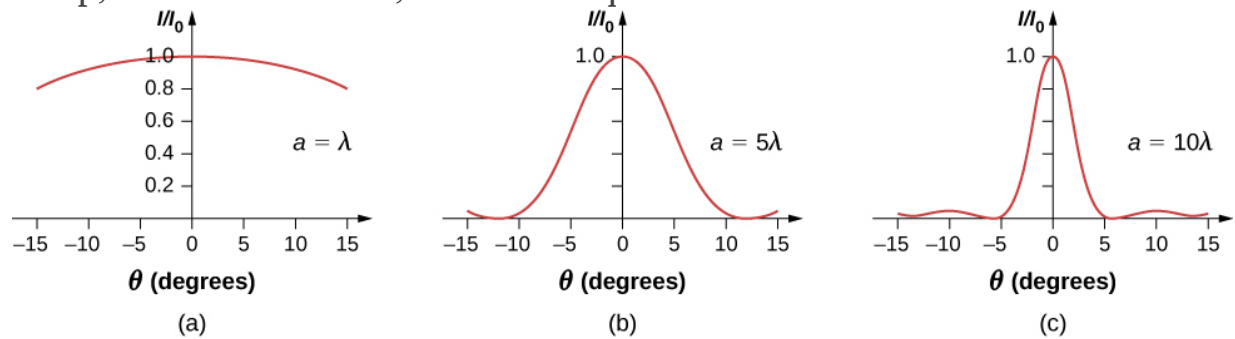
Check Your Understanding For the experiment in [\[link\]](#), at what angle from the center is the third maximum and what is its intensity relative to the central maximum?

Solution:

74.3° , $0.0083I_0$

If the slit width D is varied, the intensity distribution changes, as illustrated in [\[link\]](#). The central peak is distributed over the region from $\sin \theta = -\lambda/a$ to $\sin \theta = +\lambda/a$. For small θ , this corresponds to an angular width $\Delta\theta \approx 2\lambda/a$. Hence, an increase in the slit width results in a decrease in the

width of the central peak. For a slit with $a \gg \lambda$, the central peak is very sharp, whereas if $a \approx \lambda$, it becomes quite broad.



Single-slit diffraction patterns for various slit widths. As the slit width a increases from $a = \lambda$ to 5λ and then to 10λ , the width of the central peak decreases as the angles for the first minima decrease as predicted by [\[link\]](#).

Note:

A diffraction experiment in optics can require a lot of preparation but [this simulation](#) by Andrew Duffy offers not only a quick set up but also the ability to change the slit width instantly. Run the simulation and select “Single slit.” You can adjust the slit width and see the effect on the diffraction pattern on a screen and as a graph.

Summary

- The intensity pattern for diffraction due to a single slit can be calculated using phasors as

Equation:

$$I = I_0 \left(\frac{\sin \beta}{\beta} \right)^2,$$

where $\beta = \frac{\phi}{2} = \frac{\pi a \sin \theta}{\lambda}$, a is the slit width, λ is the wavelength, and θ is the angle from the central peak.

Conceptual Questions

Exercise:

Problem:

In [\[link\]](#), the parameter β looks like an angle but is not an angle that you can measure with a protractor in the physical world. Explain what β represents.

Solution:

The parameter $\beta = \phi/2$ is the arc angle shown in the phasor diagram in [\[link\]](#). The phase difference between the first and last Huygens wavelet across the single slit is 2β and is related to the curvature of the arc that forms the resultant phasor that determines the light intensity.

Problems

Exercise:

Problem:

A single slit of width $3.0 \mu\text{m}$ is illuminated by a sodium yellow light of wavelength 589 nm . Find the intensity at a 15° angle to the axis in terms of the intensity of the central maximum.

Exercise:

Problem:

A single slit of width 0.1 mm is illuminated by a mercury light of wavelength 576 nm . Find the intensity at a 10° angle to the axis in terms of the intensity of the central maximum.

Solution:

$$I/I_0 = 2.2 \times 10^{-5}$$

Exercise:**Problem:**

The width of the central peak in a single-slit diffraction pattern is 5.0 mm. The wavelength of the light is 600 nm, and the screen is 2.0 m from the slit. (a) What is the width of the slit? (b) Determine the ratio of the intensity at 4.5 mm from the center of the pattern to the intensity at the center.

Exercise:**Problem:**

Consider the single-slit diffraction pattern for $\lambda = 600$ nm, $a = 0.025$ mm, and $x = 2.0$ m. Find the intensity in terms of I_0 at $\theta = 0.5^\circ$, 1.0° , 1.5° , 3.0° , and 10.0° .

Solution:

$$0.63I_0, 0.11I_0, 0.0067I_0, 0.0062I_0, 0.00088I_0$$

Glossary

width of the central peak

angle between the minimum for $m = 1$ and $m = -1$

Double-Slit Diffraction

By the end of this section, you will be able to:

- Describe the combined effect of interference and diffraction with two slits, each with finite width
- Determine the relative intensities of interference fringes within a diffraction pattern
- Identify missing orders, if any

When we studied interference in Young's double-slit experiment, we ignored the diffraction effect in each slit. We assumed that the slits were so narrow that on the screen you saw only the interference of light from just two point sources. If the slit is smaller than the wavelength, then [\[link\]](#)(a) shows that there is just a spreading of light and no peaks or troughs on the screen. Therefore, it was reasonable to leave out the diffraction effect in that chapter. However, if you make the slit wider, [\[link\]](#)(b) and (c) show that you cannot ignore diffraction. In this section, we study the complications to the double-slit experiment that arise when you also need to take into account the diffraction effect of each slit.

To calculate the diffraction pattern for two (or any number of) slits, we need to generalize the method we just used for a single slit. That is, across each slit, we place a uniform distribution of point sources that radiate Huygens wavelets, and then we sum the wavelets from all the slits. This gives the intensity at any point on the screen. Although the details of that calculation can be complicated, the final result is quite simple:

Note:

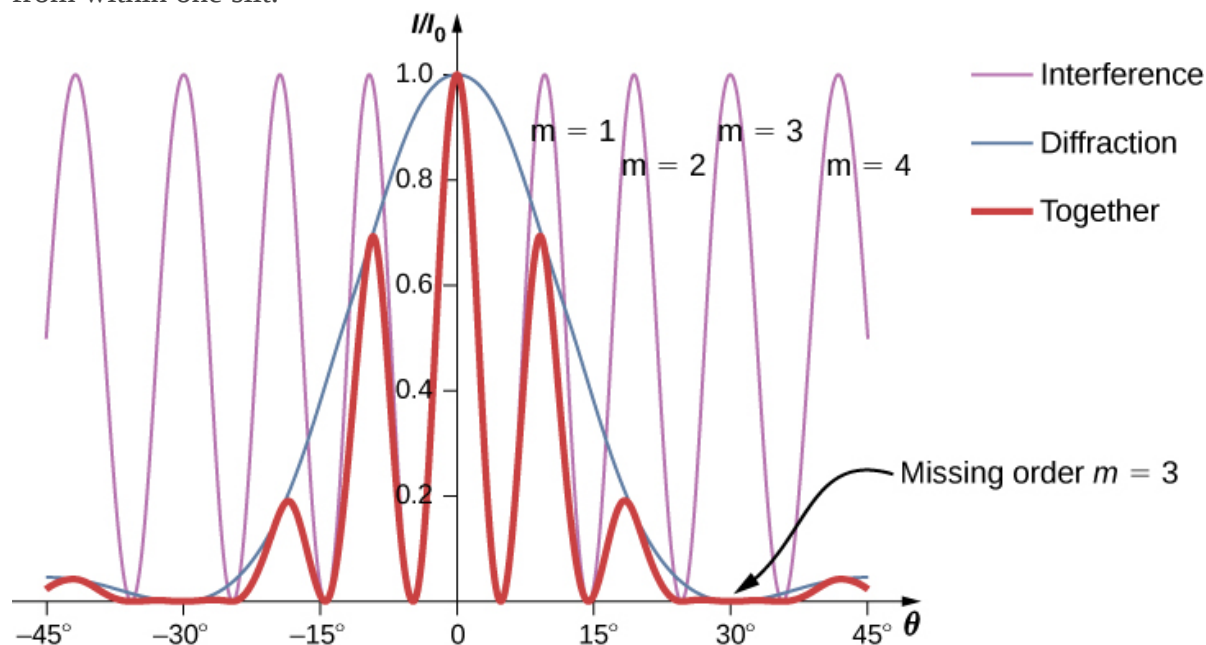
Two-Slit Diffraction Pattern

The diffraction pattern of two slits of width a that are separated by a distance d is the interference pattern of two point sources separated by d multiplied by the diffraction pattern of a slit of width a .

In other words, the *locations* of the interference fringes are given by the equation $d \sin \theta = m\lambda$, the same as when we considered the slits to be point sources, but the *intensities* of the fringes are now reduced by diffraction effects, according to [\[link\]](#). [Note that in the chapter on interference, we wrote $d \sin \theta = m\lambda$ and used the integer m to refer to interference fringes. [\[link\]](#) also uses m , but this time to refer to diffraction minima. If both equations are used simultaneously, it is good practice to use a different variable (such as n) for one of these integers in order to keep them distinct.]

Interference and diffraction effects operate simultaneously and generally produce minima at different angles. This gives rise to a complicated pattern on the screen, in which some of the maxima of interference from the two slits are missing if the maximum of the interference is in the same direction as the minimum of the diffraction. We refer to such a missing peak as a **missing order**. One example of a diffraction pattern on the screen is shown in [\[link\]](#). The

solid line with multiple peaks of various heights is the intensity observed on the screen. It is a product of the interference pattern of waves from separate slits and the diffraction of waves from within one slit.



Diffraction from a double slit. The purple line with peaks of the same height are from the interference of the waves from two slits; the blue line with one big hump in the middle is the diffraction of waves from within one slit; and the thick red line is the product of the two, which is the pattern observed on the screen. The plot shows the expected result for a slit width $a = 2\lambda$ and slit separation $d = 6\lambda$. The maximum of $m = \pm 3$ order for the interference is missing because the minimum of the diffraction occurs in the same direction.

Example:

Intensity of the Fringes

[\[link\]](#) shows that the intensity of the fringe for $m = 3$ is zero, but what about the other fringes? Calculate the intensity for the fringe at $m = 1$ relative to I_0 , the intensity of the central peak.

Strategy

Determine the angle for the double-slit interference fringe, using the equation from [Interference](#), then determine the relative intensity in that direction due to diffraction by using [\[link\]](#).

Solution

From the chapter on interference, we know that the bright interference fringes occur at $d \sin \theta = m\lambda$, or

Equation:

$$\sin \theta = \frac{m\lambda}{d}.$$

From [\[link\]](#),

Equation:

$$I = I_0 \left(\frac{\sin \beta}{\beta} \right)^2, \text{ where } \beta = \frac{\phi}{2} = \frac{\pi a \sin \theta}{\lambda}.$$

Substituting from above,

Equation:

$$\beta = \frac{\pi a \sin \theta}{\lambda} = \frac{\pi a}{\lambda} \cdot \frac{m\lambda}{d} = \frac{m\pi a}{d}.$$

For $a = 2\lambda$, $d = 6\lambda$, and $m = 1$,

Equation:

$$\beta = \frac{(1)\pi(2\lambda)}{(6\lambda)} = \frac{\pi}{3}.$$

Then, the intensity is

Equation:

$$I = I_0 \left(\frac{\sin \beta}{\beta} \right)^2 = I_0 \left(\frac{\sin (\pi/3)}{\pi/3} \right)^2 = 0.684I_0.$$

Significance

Note that this approach is relatively straightforward and gives a result that is almost exactly the same as the more complicated analysis using phasors to work out the intensity values of the double-slit interference (thin line in [\[link\]](#)). The phasor approach accounts for the downward slope in the diffraction intensity (blue line) so that the peak *near* $m = 1$ occurs at a value of θ ever so slightly smaller than we have shown here.

Example:

Two-Slit Diffraction

Suppose that in Young's experiment, slits of width 0.020 mm are separated by 0.20 mm. If the slits are illuminated by monochromatic light of wavelength 500 nm, how many bright fringes are observed in the central peak of the diffraction pattern?

Solution

From [\[link\]](#), the angular position of the first diffraction minimum is

$$\theta \approx \sin \theta = \frac{\lambda}{a} = \frac{5.0 \times 10^{-7} \text{ m}}{2.0 \times 10^{-5} \text{ m}} = 2.5 \times 10^{-2} \text{ rad}.$$

Using $d \sin \theta = m\lambda$ for $\theta = 2.5 \times 10^{-2} \text{ rad}$, we find

Equation:

$$m = \frac{d \sin \theta}{\lambda} = \frac{(0.20 \text{ mm}) (2.5 \times 10^{-2} \text{ rad})}{5.0 \times 10^{-7} \text{ m}} = 10,$$

which is the maximum interference order that fits inside the central peak. We note that $m = \pm 10$ are missing orders as θ matches exactly. Accordingly, we observe bright fringes for

Equation:

$$m = -9, -8, -7, -6, -5, -4, -3, -2, -1, 0, +1, +2, +3, +4, +5, +6, +7, +8, \text{ and } +9$$

for a total of 19 bright fringes.

Note:

Exercise:

Problem:

Check Your Understanding For the experiment in [\[link\]](#), show that $m = 20$ is also a missing order.

Solution:

From $d \sin \theta = m\lambda$, the interference maximum occurs at 2.87° for $m = 20$. From [\[link\]](#), this is also the angle for the second diffraction minimum. (*Note:* Both equations use the index m but they refer to separate phenomena.)

Note:

Explore the effects of double-slit diffraction. In [this simulation](#) written by Fu-Kwun Hwang, select $N = 2$ using the slider and see what happens when you control the slit width, slit separation and the wavelength. Can you make an order go “missing?”

Summary

- With real slits with finite widths, the effects of interference and diffraction operate simultaneously to form a complicated intensity pattern.

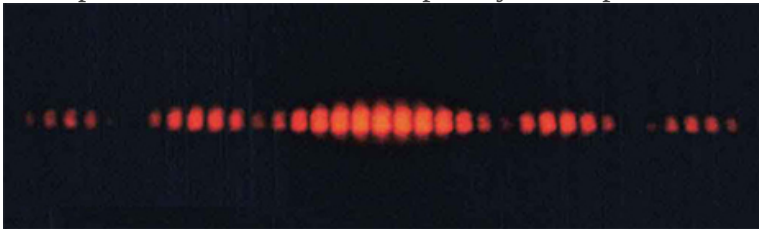
- Relative intensities of interference fringes within a diffraction pattern can be determined.
- Missing orders occur when an interference maximum and a diffraction minimum are located together.

Conceptual Questions

Exercise:

Problem:

Shown below is the central part of the interference pattern for a pure wavelength of red light projected onto a double slit. The pattern is actually a combination of single- and double-slit interference. Note that the bright spots are evenly spaced. Is this a double- or single-slit characteristic? Note that some of the bright spots are dim on either side of the center. Is this a single- or double-slit characteristic? Which is smaller, the slit width or the separation between slits? Explain your responses.



(credit: PASCO)

Problems

Exercise:

Problem:

Two slits of width $2\ \mu\text{m}$, each in an opaque material, are separated by a center-to-center distance of $6\ \mu\text{m}$. A monochromatic light of wavelength $450\ \text{nm}$ is incident on the double-slit. One finds a combined interference and diffraction pattern on the screen.

- How many peaks of the interference will be observed in the central maximum of the diffraction pattern?
- How many peaks of the interference will be observed if the slit width is doubled while keeping the distance between the slits same?
- How many peaks of interference will be observed if the slits are separated by twice the distance, that is, $12\ \mu\text{m}$, while keeping the widths of the slits same?

(d) What will happen in (a) if instead of 450-nm light another light of wavelength 680 nm is used?

(e) What is the value of the ratio of the intensity of the central peak to the intensity of the next bright peak in (a)?

(f) Does this ratio depend on the wavelength of the light?

(g) Does this ratio depend on the width or separation of the slits?

Exercise:

Problem:

A double slit produces a diffraction pattern that is a combination of single- and double-slit interference. Find the ratio of the width of the slits to the separation between them, if the first minimum of the single-slit pattern falls on the fifth maximum of the double-slit pattern. (This will greatly reduce the intensity of the fifth maximum.)

Solution:

0.200

Exercise:

Problem:

For a double-slit configuration where the slit separation is four times the slit width, how many interference fringes lie in the central peak of the diffraction pattern?

Exercise:

Problem:

Light of wavelength 500 nm falls normally on 50 slits that are 2.5×10^{-3} mm wide and spaced 5.0×10^{-3} mm apart. How many interference fringes lie in the central peak of the diffraction pattern?

Solution:

3

Exercise:

Problem:

A monochromatic light of wavelength 589 nm incident on a double slit with slit width $2.5 \mu\text{m}$ and unknown separation results in a diffraction pattern containing nine interference peaks inside the central maximum. Find the separation of the slits.

Exercise:

Problem:

When a monochromatic light of wavelength 430 nm incident on a double slit of slit separation $5\text{ }\mu\text{m}$, there are 11 interference fringes in its central maximum. How many interference fringes will be in the central maximum of a light of the same wavelength and slit widths, but a new slit separation of $4\text{ }\mu\text{m}$?

Solution:

9

Exercise:**Problem:**

Determine the intensities of two interference peaks other than the central peak in the central maximum of the diffraction, if possible, when a light of wavelength 628 nm is incident on a double slit of width 500 nm and separation 1500 nm. Use the intensity of the central spot to be 1 mW/cm^2 .

Glossary

missing order

interference maximum that is not seen because it coincides with a diffraction minimum

two-slit diffraction pattern

diffraction pattern of two slits of width D that are separated by a distance d is the interference pattern of two point sources separated by d multiplied by the diffraction pattern of a slit of width D

Circular Apertures and Resolution

By the end of this section, you will be able to:

- Describe the diffraction limit on resolution
- Describe the diffraction limit on beam propagation

Light diffracts as it moves through space, bending around obstacles, interfering constructively and destructively. This can be used as a spectroscopic tool—a diffraction grating disperses light according to wavelength, for example, and is used to produce spectra—but diffraction also limits the detail we can obtain in images.

[\[link\]](#)(a) shows the effect of passing light through a small circular aperture. Instead of a bright spot with sharp edges, we obtain a spot with a fuzzy edge surrounded by circles of light. This pattern is caused by diffraction, similar to that produced by a single slit. Light from different parts of the circular aperture interferes constructively and destructively. The effect is most noticeable when the aperture is small, but the effect is there for large apertures as well.



(a)



(b)



(c)

(a) Monochromatic light passed through a small circular aperture produces this diffraction pattern. (b) Two point-light sources that are close to one another produce overlapping images because of diffraction. (c) If the sources are closer together, they cannot be distinguished or resolved.

How does diffraction affect the detail that can be observed when light passes through an aperture? [\[link\]](#)(b) shows the diffraction pattern produced by two point-light sources that are close to one another. The pattern is similar to that for a single point source, and it is still possible to tell that there are two light sources rather than one. If they are closer together, as in [\[link\]](#)(c), we cannot distinguish them, thus limiting the detail or **resolution** we can obtain. This limit is an inescapable consequence of the wave nature of light.

Diffraction limits the resolution in many situations. The acuity of our vision is limited because light passes through the pupil, which is the circular aperture of the eye. Be aware that the diffraction-like spreading of light is due to the limited diameter of a light beam, not the interaction with an aperture. Thus, light passing through a lens with a diameter D shows this effect and spreads, blurring the image, just as light passing through an aperture of diameter D does. Thus, diffraction limits the resolution of any system having a lens or mirror. Telescopes are also limited by diffraction, because of the finite diameter D of the primary mirror.

Just what is the limit? To answer that question, consider the diffraction pattern for a circular aperture, which has a central maximum that is wider and brighter than the maxima surrounding it (similar to a slit) ([\[link\]](#)(a)). It can be shown that, for a circular aperture of diameter D , the first minimum in the diffraction pattern occurs at $\theta = 1.22\lambda/D$ (providing the aperture is large compared with the wavelength of light, which is the case for most optical instruments). The accepted criterion for determining the **diffraction limit** to resolution based on this angle is known as the **Rayleigh criterion**, which was developed by Lord Rayleigh in the nineteenth century.

Note:

Rayleigh Criterion

The diffraction limit to resolution states that two images are just resolvable when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other ([\[link\]](#)(b)).

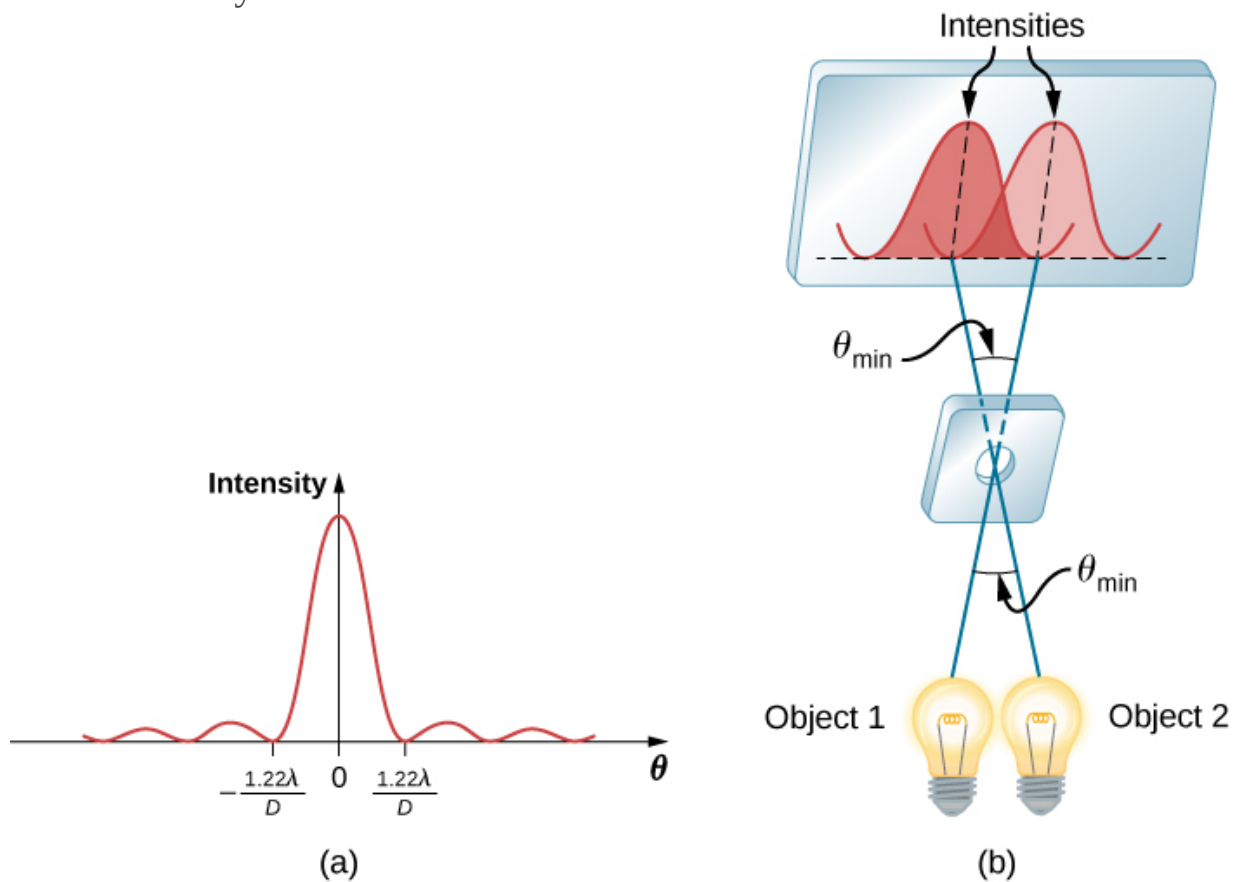
The first minimum is at an angle of $\theta = 1.22\lambda/D$, so that two point objects are just resolvable if they are separated by the angle

Note:

Equation:

$$\theta = 1.22 \frac{\lambda}{D}$$

where λ is the wavelength of light (or other electromagnetic radiation) and D is the diameter of the aperture, lens, mirror, etc., with which the two objects are observed. In this expression, θ has units of radians. This angle is also commonly known as the diffraction limit.



(a) Graph of intensity of the diffraction pattern for a circular aperture. Note that, similar to a single slit, the central maximum is wider and brighter than those to the sides. (b) Two point objects produce overlapping diffraction patterns. Shown here is the Rayleigh criterion for being just resolvable. The central maximum of one pattern lies on the first minimum of the other.

All attempts to observe the size and shape of objects are limited by the wavelength of the probe. Even the small wavelength of light prohibits exact precision. When extremely small wavelength probes are used, as with an electron microscope, the system is disturbed, still limiting our knowledge. Heisenberg's uncertainty principle asserts that this limit is fundamental and inescapable, as we shall see in the chapter on quantum mechanics.

Example:

Calculating Diffraction Limits of the Hubble Space Telescope

The primary mirror of the orbiting Hubble Space Telescope has a diameter of 2.40 m. Being in orbit, this telescope avoids the degrading effects of atmospheric distortion on its resolution. (a) What is the angle between two just-resolvable point light sources (perhaps two stars)? Assume an average light wavelength of 550 nm. (b) If these two stars are at a distance of 2 million light-years, which is the distance of the Andromeda Galaxy, how close together can they be and still be resolved? (A light-year, or ly, is the distance light travels in 1 year.)

Strategy

The Rayleigh criterion stated in [\[link\]](#), $\theta = 1.22\lambda/D$, gives the smallest possible angle θ between point sources, or the best obtainable resolution. Once this angle is known, we can calculate the distance between the stars, since we are given how far away they are.

Solution

a. The Rayleigh criterion for the minimum resolvable angle is

Equation:

$$\theta = 1.22 \frac{\lambda}{D}.$$

Entering known values gives

Equation:

$$\theta = 1.22 \frac{550 \times 10^{-9} \text{ m}}{2.40 \text{ m}} = 2.80 \times 10^{-7} \text{ rad}.$$

- b. The distance s between two objects a distance r away and separated by an angle θ is $s = r\theta$.

Substituting known values gives

Equation:

$$s = (2.0 \times 10^6 \text{ ly}) (2.80 \times 10^{-7} \text{ rad}) = 0.56 \text{ ly}.$$

Significance

The angle found in part (a) is extraordinarily small (less than 1/50,000 of a degree), because the primary mirror is so large compared with the wavelength of light. As noticed, diffraction effects are most noticeable when light interacts with objects having sizes on the order of the wavelength of light. However, the effect is still there, and there is a diffraction limit to what is observable. The actual resolution of the Hubble Telescope is not quite as good as that found here. As with all instruments, there are other effects, such as nonuniformities in mirrors or aberrations in lenses that further limit resolution. However, [\[link\]](#) gives an indication of the extent of the detail observable with the Hubble because of its size and quality, and especially because it is above Earth's atmosphere.



(a)



(b)

These two photographs of the M82 Galaxy give an idea of the observable detail using (a) a ground-based telescope and (b) the Hubble Space Telescope. (credit a: modification of work by “Ricnun”/Wikimedia Commons; credit b: modification of work by NASA, ESA, and The Hubble Heritage Team (STScI/AURA))

The answer in part (b) indicates that two stars separated by about half a light-year can be resolved. The average distance between stars in a galaxy is on the order of five light-years in the outer parts and about one light-year near the galactic center. Therefore, the Hubble can resolve most of the individual stars in Andromeda Galaxy, even though it lies at such a huge distance that its light takes 2 million years to reach us. [\[link\]](#) shows another mirror used to observe radio waves from outer space.



A 305-m-diameter paraboloid at Arecibo in Puerto Rico is lined with reflective material, making it into a radio telescope. It is the largest curved focusing dish in the world. Although D for Arecibo is much larger than for the Hubble Telescope, it detects radiation of a much longer wavelength and its diffraction limit is significantly poorer than Hubble's. The Arecibo telescope is still very useful, because important information is carried by radio waves that is not carried by visible light. (credit: Jeff Hitchcock)

Note:
Exercise:

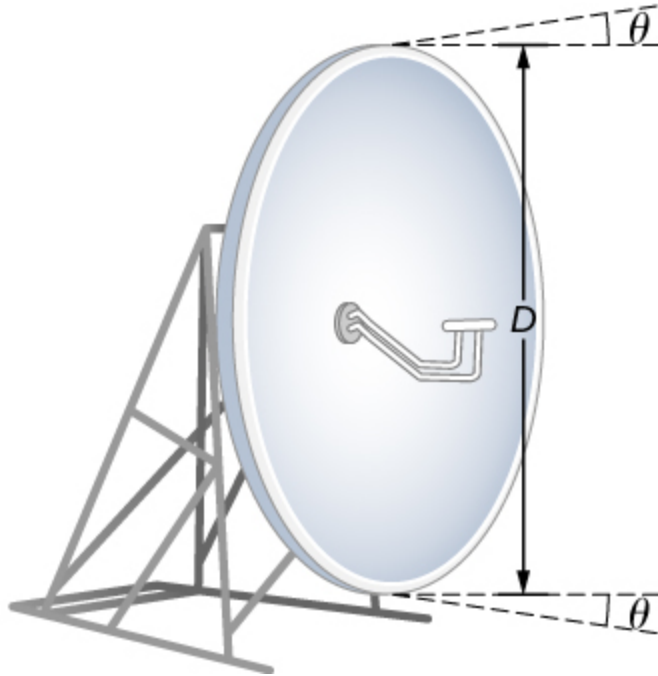
Problem:

Check Your Understanding What is the angular resolution of the Arecibo telescope shown in [\[link\]](#) when operated at 21-cm wavelength? How does it compare to the resolution of the Hubble Telescope?

Solution:

8.4×10^{-4} rad, 3000 times broader than the Hubble Telescope

Diffraction is not only a problem for optical instruments but also for the electromagnetic radiation itself. Any beam of light having a finite diameter D and a wavelength λ exhibits diffraction spreading. The beam spreads out with an angle θ given by [\[link\]](#), $\theta = 1.22\lambda/D$. Take, for example, a laser beam made of rays as parallel as possible (angles between rays as close to $\theta = 0^\circ$ as possible) instead spreads out at an angle $\theta = 1.22\lambda/D$, where D is the diameter of the beam and λ is its wavelength. This spreading is impossible to observe for a flashlight because its beam is not very parallel to start with. However, for long-distance transmission of laser beams or microwave signals, diffraction spreading can be significant ([\[link\]](#)). To avoid this, we can increase D . This is done for laser light sent to the moon to measure its distance from Earth. The laser beam is expanded through a telescope to make D much larger and θ smaller.



The beam produced by this microwave transmission antenna spreads out at a minimum angle $\theta = 1.22\lambda/D$ due to diffraction. It is impossible to produce a near-parallel beam because the beam has a limited diameter.

In most biology laboratories, resolution is an issue when the use of the microscope is introduced. The smaller the distance x by which two objects can be separated and still be seen as distinct, the greater the resolution. The resolving power of a lens is defined as that distance x . An expression for resolving power is obtained from the Rayleigh criterion. [\[link\]](#)(a) shows two point objects separated by a distance x . According to the Rayleigh criterion, resolution is possible when the minimum angular separation is

Equation:

$$\theta = 1.22 \frac{\lambda}{D} = \frac{x}{d},$$

where d is the distance between the specimen and the objective lens, and we have used the small angle approximation (i.e., we have assumed that x is much smaller than d), so that $\tan \theta \approx \sin \theta \approx \theta$. Therefore, the resolving power is

Equation:

$$x = 1.22 \frac{\lambda d}{D}.$$

Another way to look at this is by the concept of numerical aperture (NA), which is a measure of the maximum acceptance angle at which a lens will take light and still contain it within the lens. [\[link\]](#)(b) shows a lens and an object at point P . The NA here is a measure of the ability of the lens to gather light and resolve fine detail. The angle subtended by the lens at its focus is defined to be $\theta = 2\alpha$. From the figure and again using the small angle approximation, we can write

Equation:

$$\sin \alpha = \frac{D/2}{d} = \frac{D}{2d}.$$

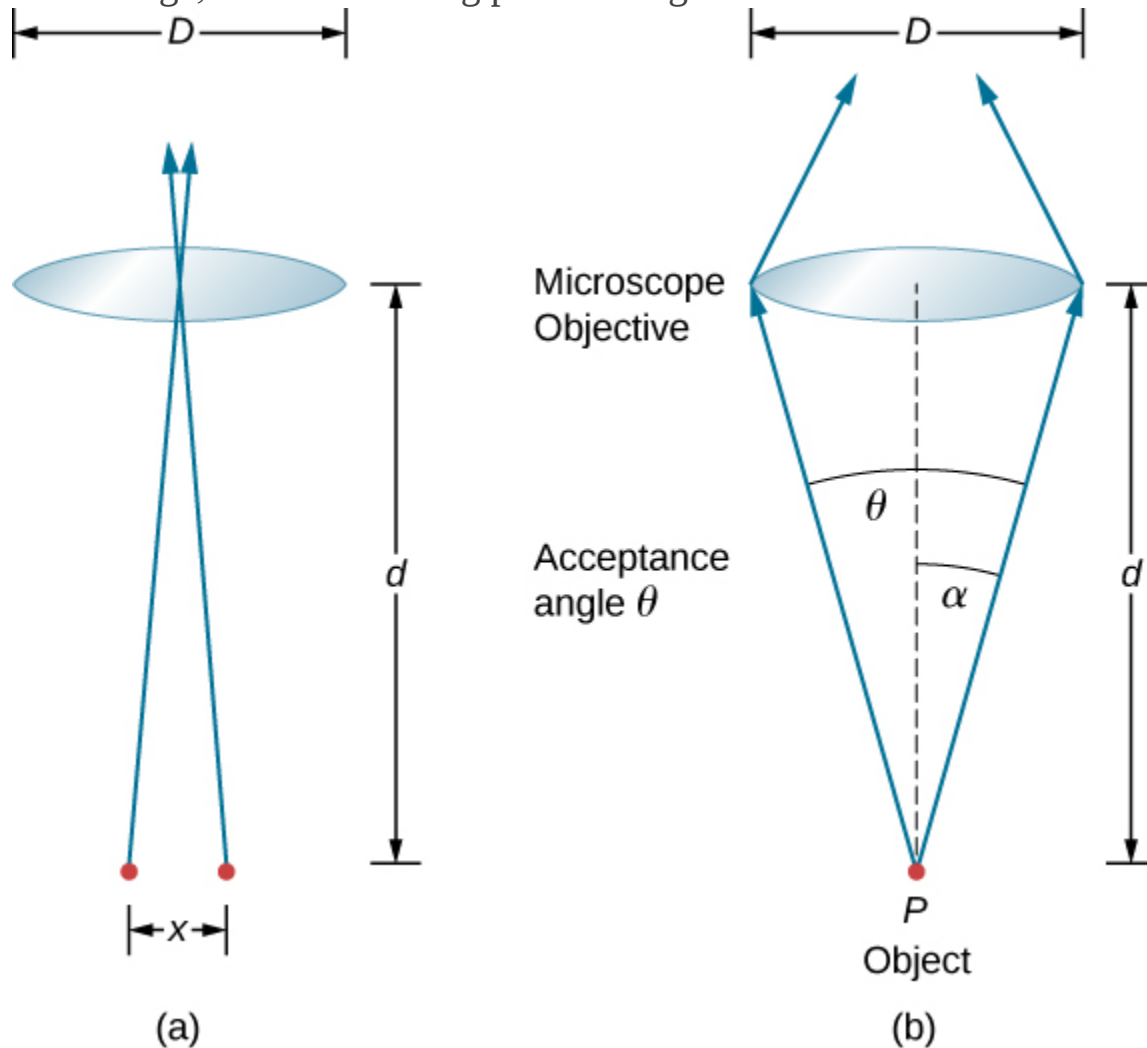
The NA for a lens is $NA = n \sin \alpha$, where n is the index of refraction of the medium between the objective lens and the object at point P . From this definition for NA , we can see that

Equation:

$$x = 1.22 \frac{\lambda d}{D} = 1.22 \frac{\lambda}{2 \sin \alpha} = 0.61 \frac{\lambda n}{NA}.$$

In a microscope, NA is important because it relates to the resolving power of a lens. A lens with a large NA is able to resolve finer details. Lenses with larger NA are also able to collect more light and so give a brighter image. Another way to describe this situation is that the larger the NA , the larger the cone of light that can be brought into the lens, so more of the diffraction

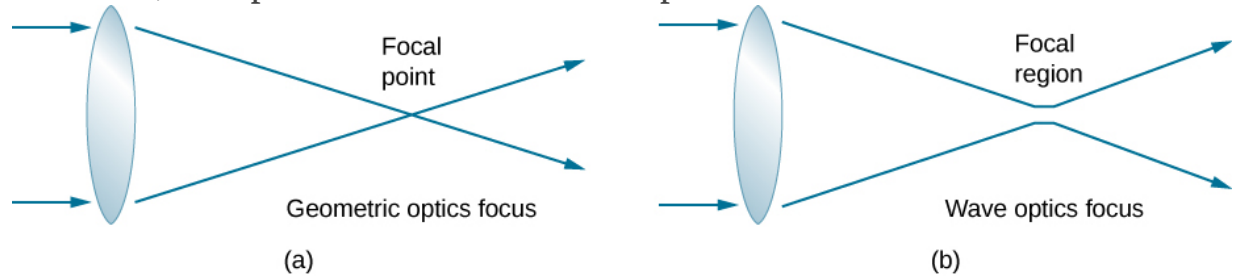
modes are collected. Thus the microscope has more information to form a clear image, and its resolving power is higher.



(a) Two points separated by a distance x and positioned a distance d away from the objective. (b) Terms and symbols used in discussion of resolving power for a lens and an object at point P (credit a: modification of work by “Infopro”/Wikimedia Commons).

One of the consequences of diffraction is that the focal point of a beam has a finite width and intensity distribution. Imagine focusing when only considering geometric optics, as in [\[link\]](#)(a). The focal point is regarded as

an infinitely small point with a huge intensity and the capacity to incinerate most samples, irrespective of the NA of the objective lens—an unphysical oversimplification. For wave optics, due to diffraction, we take into account the phenomenon in which the focal point spreads to become a focal spot ([link](#)(b)) with the size of the spot decreasing with increasing NA . Consequently, the intensity in the focal spot increases with increasing NA . The higher the NA , the greater the chances of photodegrading the specimen. However, the spot never becomes a true point.



(a) In geometric optics, the focus is modelled as a point, but it is not physically possible to produce such a point because it implies infinite intensity. (b) In wave optics, the focus is an extended region.

In a different type of microscope, molecules within a specimen are made to emit light through a mechanism called fluorescence. By controlling the molecules emitting light, it has become possible to construct images with resolution much finer than the Rayleigh criterion, thus circumventing the diffraction limit. The development of super-resolved fluorescence microscopy led to the 2014 Nobel Prize in Chemistry.

Note:

In this Optical Resolution Model, two diffraction patterns for light through two circular apertures are shown side by side in [this simulation](#) by Fu-Kwun Hwang. Watch the patterns merge as you decrease the aperture diameters.

Summary

- Diffraction limits resolution.
- The Rayleigh criterion states that two images are just resolvable when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other.

Conceptual Questions

Exercise:

Problem:

Is higher resolution obtained in a microscope with red or blue light? Explain your answer.

Solution:

blue; The shorter wavelength of blue light results in a smaller angle for diffraction limit.

Exercise:

Problem:

The resolving power of refracting telescope increases with the size of its objective lens. What other advantage is gained with a larger lens?

Exercise:

Problem:

The distance between atoms in a molecule is about 10^{-8} cm. Can visible light be used to “see” molecules?

Solution:

No, these distances are three orders of magnitude smaller than the wavelength of visible light, so visible light makes a poor probe for atoms.

Exercise:**Problem:**

A beam of light always spreads out. Why can a beam not be created with parallel rays to prevent spreading? Why can lenses, mirrors, or apertures not be used to correct the spreading?

Problems**Exercise:****Problem:**

The 305-m-diameter Arecibo radio telescope pictured in [\[link\]](#) detects radio waves with a 4.00-cm average wavelength. (a) What is the angle between two just-resolvable point sources for this telescope? (b) How close together could these point sources be at the 2 million light-year distance of the Andromeda Galaxy?

Exercise:**Problem:**

Assuming the angular resolution found for the Hubble Telescope in [\[link\]](#), what is the smallest detail that could be observed on the moon?

Solution:

107 m

Exercise:**Problem:**

Diffraction spreading for a flashlight is insignificant compared with other limitations in its optics, such as spherical aberrations in its mirror. To show this, calculate the minimum angular spreading of a flashlight beam that is originally 5.00 cm in diameter with an average wavelength of 600 nm.

Exercise:**Problem:**

(a) What is the minimum angular spread of a 633-nm wavelength He-Ne laser beam that is originally 1.00 mm in diameter? (b) If this laser is aimed at a mountain cliff 15.0 km away, how big will the illuminated spot be? (c) How big a spot would be illuminated on the moon, neglecting atmospheric effects? (This might be done to hit a corner reflector to measure the round-trip time and, hence, distance.)

Solution:

a. 7.72×10^{-4} rad; b. 23.2 m; c. 590 km

Exercise:**Problem:**

A telescope can be used to enlarge the diameter of a laser beam and limit diffraction spreading. The laser beam is sent through the telescope in opposite the normal direction and can then be projected onto a satellite or the moon. (a) If this is done with the Mount Wilson telescope, producing a 2.54-m-diameter beam of 633-nm light, what is the minimum angular spread of the beam? (b) Neglecting atmospheric effects, what is the size of the spot this beam would make on the moon, assuming a lunar distance of 3.84×10^8 m ?

Exercise:

Problem:

The limit to the eye's acuity is actually related to diffraction by the pupil. (a) What is the angle between two just-resolvable points of light for a 3.00-mm-diameter pupil, assuming an average wavelength of 550 nm? (b) Take your result to be the practical limit for the eye. What is the greatest possible distance a car can be from you if you can resolve its two headlights, given they are 1.30 m apart? (c) What is the distance between two just-resolvable points held at an arm's length (0.800 m) from your eye? (d) How does your answer to (c) compare to details you normally observe in everyday circumstances?

Solution:

a. 2.24×10^{-4} rad; b. 5.81 km; c. 0.179 mm; d. can resolve details 0.2 mm apart at arm's length

Exercise:**Problem:**

What is the minimum diameter mirror on a telescope that would allow you to see details as small as 5.00 km on the moon some 384,000 km away? Assume an average wavelength of 550 nm for the light received.

Exercise:**Problem:**

Find the radius of a star's image on the retina of an eye if its pupil is open to 0.65 cm and the distance from the pupil to the retina is 2.8 cm. Assume $\lambda = 550$ nm.

Solution:

$2.9 \mu\text{m}$

Exercise:

Problem:

(a) The dwarf planet Pluto and its moon, Charon, are separated by 19,600 km. Neglecting atmospheric effects, should the 5.08-m-diameter Palomar Mountain telescope be able to resolve these bodies when they are 4.50×10^9 km from Earth? Assume an average wavelength of 550 nm. (b) In actuality, it is just barely possible to discern that Pluto and Charon are separate bodies using a ground-based telescope. What are the reasons for this?

Exercise:**Problem:**

A spy satellite orbits Earth at a height of 180 km. What is the minimum diameter of the objective lens in a telescope that must be used to resolve columns of troops marching 2.0 m apart? Assume $\lambda = 550$ nm.

Solution:

6.0 cm

Exercise:**Problem:**

What is the minimum angular separation of two stars that are just-resolvable by the 8.1-m Gemini South telescope, if atmospheric effects do not limit resolution? Use 550 nm for the wavelength of the light from the stars.

Exercise:**Problem:**

The headlights of a car are 1.3 m apart. What is the maximum distance at which the eye can resolve these two headlights? Take the pupil diameter to be 0.40 cm.

Solution:

7.71 km

Exercise:

Problem:

When dots are placed on a page from a laser printer, they must be close enough so that you do not see the individual dots of ink. To do this, the separation of the dots must be less than Rayleigh's criterion. Take the pupil of the eye to be 3.0 mm and the distance from the paper to the eye of 35 cm; find the minimum separation of two dots such that they cannot be resolved. How many dots per inch (dpi) does this correspond to?

Exercise:

Problem:

Suppose you are looking down at a highway from a jetliner flying at an altitude of 6.0 km. How far apart must two cars be if you are able to distinguish them? Assume that $\lambda = 550 \text{ nm}$ and that the diameter of your pupils is 4.0 mm.

Solution:

1.0 m

Exercise:

Problem:

Can an astronaut orbiting Earth in a satellite at a distance of 180 km from the surface distinguish two skyscrapers that are 20 m apart? Assume that the pupils of the astronaut's eyes have a diameter of 5.0 mm and that most of the light is centered around 500 nm.

Exercise:

Problem:

The characters of a stadium scoreboard are formed with closely spaced lightbulbs that radiate primarily yellow light. (Use $\lambda = 600 \text{ nm}$.) How closely must the bulbs be spaced so that an observer 80 m away sees a display of continuous lines rather than the individual bulbs? Assume that the pupil of the observer's eye has a diameter of 5.0 mm.

Solution:

1.2 cm or closer

Exercise:**Problem:**

If a microscope can accept light from objects at angles as large as $\alpha = 70^\circ$, what is the smallest structure that can be resolved when illuminated with light of wavelength 500 nm and (a) the specimen is in air? (b) When the specimen is immersed in oil, with index of refraction of 1.52?

Exercise:**Problem:**

A camera uses a lens with aperture 2.0 cm. What is the angular resolution of a photograph taken at 700 nm wavelength? Can it resolve the millimeter markings of a ruler placed 35 m away?

Solution:

no

Glossary

diffraction limit

fundamental limit to resolution due to diffraction

Rayleigh criterion

two images are just-resolvable when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other

resolution

ability, or limit thereof, to distinguish small details in images

X-Ray Diffraction

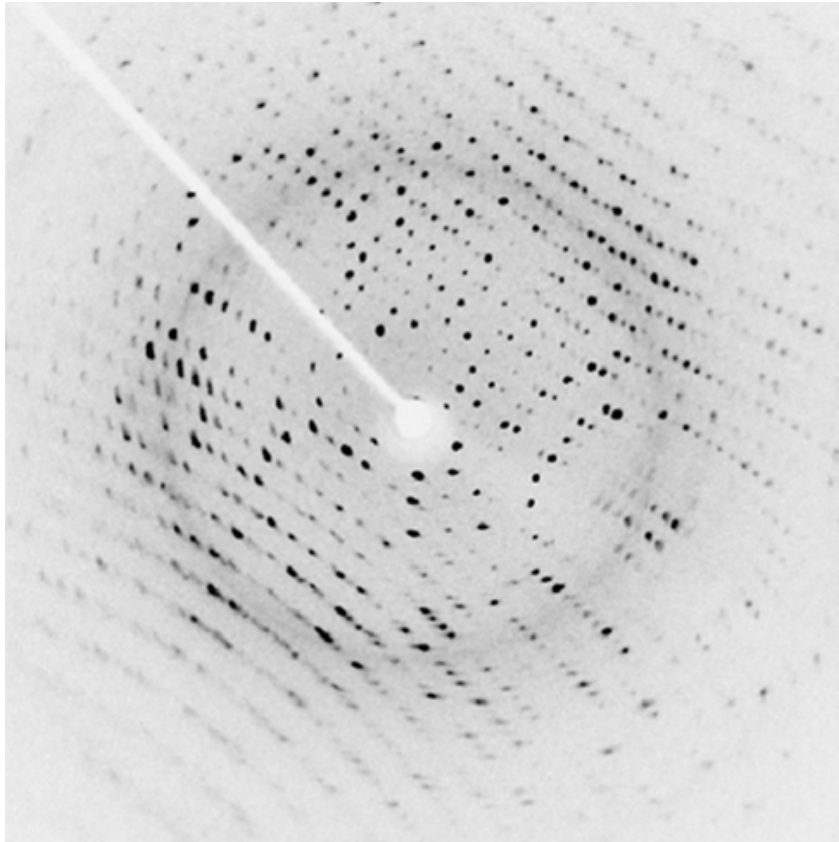
By the end of this section, you will be able to:

- Describe interference and diffraction effects exhibited by X-rays in interaction with atomic-scale structures

Since X-ray photons are very energetic, they have relatively short wavelengths, on the order of 10^{-8} m to 10^{-12} m. Thus, typical X-ray photons act like rays when they encounter macroscopic objects, like teeth, and produce sharp shadows. However, since atoms are on the order of 0.1 nm in size, X-rays can be used to detect the location, shape, and size of atoms and molecules. The process is called **X-ray diffraction**, and it involves the interference of X-rays to produce patterns that can be analyzed for information about the structures that scattered the X-rays.

Perhaps the most famous example of X-ray diffraction is the discovery of the double-helical structure of DNA in 1953 by an international team of scientists working at England's Cavendish Laboratory—American James Watson, Englishman Francis Crick, and New Zealand-born Maurice Wilkins. Using X-ray diffraction data produced by Rosalind Franklin, they were the first to model the double-helix structure of DNA that is so crucial to life. For this work, Watson, Crick, and Wilkins were awarded the 1962 Nobel Prize in Physiology or Medicine. (There is some debate and controversy over the issue that Rosalind Franklin was not included in the prize, although she died in 1958, before the prize was awarded.)

[\[link\]](#) shows a diffraction pattern produced by the scattering of X-rays from a crystal. This process is known as X-ray crystallography because of the information it can yield about crystal structure, and it was the type of data Rosalind Franklin supplied to Watson and Crick for DNA. Not only do X-rays confirm the size and shape of atoms, they give information about the atomic arrangements in materials. For example, more recent research in high-temperature superconductors involves complex materials whose lattice arrangements are crucial to obtaining a superconducting material. These can be studied using X-ray crystallography.



X-ray diffraction from the crystal of a protein (hen egg lysozyme) produced this interference pattern. Analysis of the pattern yields information about the structure of the protein. (credit: "Del45"/Wikimedia Commons)

Historically, the scattering of X-rays from crystals was used to prove that X-rays are energetic electromagnetic (EM) waves. This was suspected from the time of the discovery of X-rays in 1895, but it was not until 1912 that the German Max von Laue (1879–1960) convinced two of his colleagues to scatter X-rays from crystals. If a diffraction pattern is obtained, he reasoned, then the X-rays must be waves, and their wavelength could be determined. (The spacing of atoms in various crystals was reasonably well known at the time, based on good values for Avogadro's number.) The experiments were convincing, and the 1914 Nobel Prize in Physics was given to von Laue for his suggestion leading to the proof that X-rays are EM waves. In 1915, the unique father-and-son team of Sir William Henry Bragg and his son Sir William Lawrence Bragg were awarded a joint Nobel Prize for inventing the X-ray spectrometer and the then-new science of X-ray analysis.

In ways reminiscent of thin-film interference, we consider two plane waves at X-ray wavelengths, each one reflecting off a different plane of atoms within a crystal's lattice, as shown in [\[link\]](#). From the geometry, the difference in path lengths is $2d \sin \theta$. Constructive interference results when this distance is an integer multiple of the wavelength. This condition is captured by the *Bragg equation*,

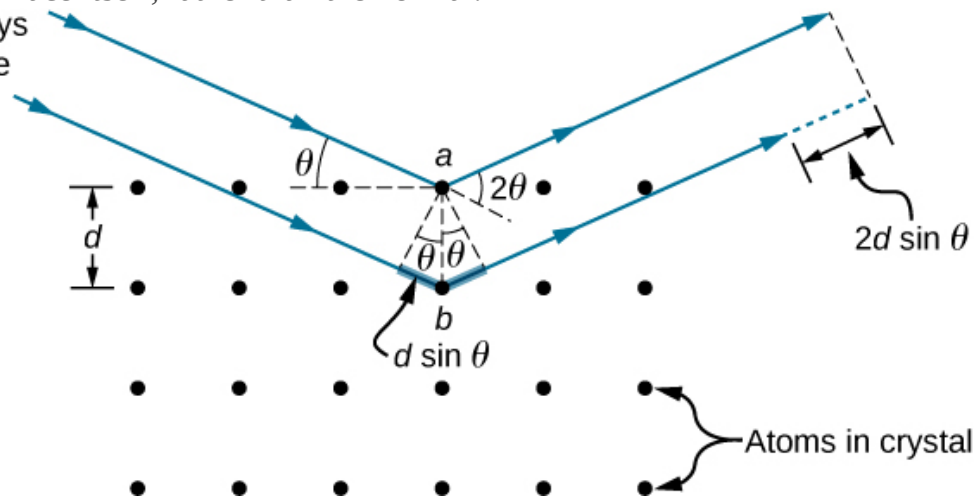
Note:

Equation:

$$m\lambda = 2d \sin \theta, \quad m = 1, 2, 3 \dots$$

where m is a positive integer and d is the spacing between the planes. Following the Law of Reflection, both the incident and reflected waves are described by the same angle, θ , but unlike the general practice in geometric optics, θ is measured with respect to the surface itself, rather than the normal.

Light rays
in phase



X-ray diffraction with a crystal. Two incident waves reflect off two planes of a crystal. The difference in path lengths is indicated by the dashed line.

Example:

X-Ray Diffraction with Salt Crystals

Common table salt is composed mainly of NaCl crystals. In a NaCl crystal, there is a family of planes 0.252 nm apart. If the first-order maximum is observed at an incidence angle of 18.1° , what is the wavelength of the X-ray scattering from this crystal?

Strategy

Use the Bragg equation, [\[link\]](#), $m\lambda = 2d \sin \theta$, to solve for θ .

Solution

For first-order, $m = 1$, and the plane spacing d is known. Solving the Bragg equation for wavelength yields

Equation:

$$\lambda = \frac{2d \sin \theta}{m} = \frac{2 (0.252 \times 10^{-9} \text{ m}) \sin (18.1^\circ)}{1} = 1.57 \times 10^{-10} \text{ m, or } 0.157 \text{ nm.}$$

Significance

The determined wavelength fits within the X-ray region of the electromagnetic spectrum. Once again, the wave nature of light makes itself prominent when the wavelength ($\lambda = 0.157 \text{ nm}$) is comparable to the size of the physical structures ($d = 0.252 \text{ nm}$) it interacts with.

Note:

Exercise:

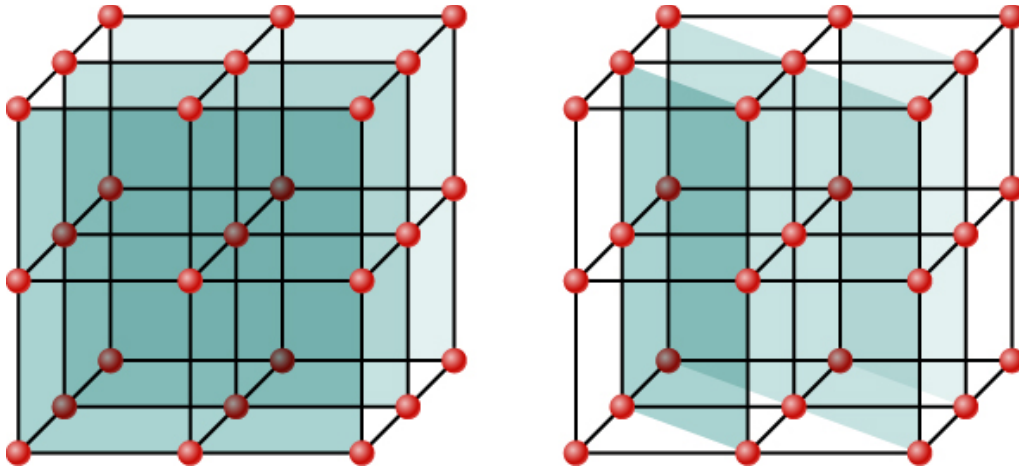
Problem:

Check Your Understanding For the experiment described in [\[link\]](#), what are the two other angles where interference maxima may be observed? What limits the number of maxima?

Solution:

38.4° and 68.8° ; Between $\theta = 0^\circ \rightarrow 90^\circ$, orders 1, 2, and 3, are all that exist.

Although [\[link\]](#) depicts a crystal as a two-dimensional array of scattering centers for simplicity, real crystals are structures in three dimensions. Scattering can occur simultaneously from different families of planes at different orientations and spacing patterns known as called **Bragg planes**, as shown in [\[link\]](#). The resulting interference pattern can be quite complex.



Because of the regularity that makes a crystal structure, one crystal can have many families of planes within its geometry, each one giving rise to X-ray diffraction.

Summary

- X-rays are relatively short-wavelength EM radiation and can exhibit wave characteristics such as interference when interacting with correspondingly small objects.

Conceptual Questions

Exercise:

Problem: Crystal lattices can be examined with X-rays but not UV. Why?

Solution:

UV wavelengths are much larger than lattice spacings in crystals such that there is no diffraction. The Bragg equation implies a value for $\sin\theta$ greater than unity, which has no solution.

Problems

Exercise:

Problem:

X-rays of wavelength 0.103 nm reflect off a crystal and a second-order maximum is recorded at a Bragg angle of 25.5° . What is the spacing between the scattering planes in this crystal?

Exercise:**Problem:**

A first-order Bragg reflection maximum is observed when a monochromatic X-ray falls on a crystal at a 32.3° angle to a reflecting plane. What is the wavelength of this X-ray?

Solution:

0.120 nm

Exercise:**Problem:**

An X-ray scattering experiment is performed on a crystal whose atoms form planes separated by 0.440 nm. Using an X-ray source of wavelength 0.548 nm, what is the angle (with respect to the planes in question) at which the experimenter needs to illuminate the crystal in order to observe a first-order maximum?

Exercise:**Problem:**

The structure of the NaCl crystal forms reflecting planes 0.541 nm apart. What is the smallest angle, measured from these planes, at which X-ray diffraction can be observed, if X-rays of wavelength 0.085 nm are used?

Solution:

4.51°

Exercise:**Problem:**

On a certain crystal, a first-order X-ray diffraction maximum is observed at an angle of 27.1° relative to its surface, using an X-ray source of unknown wavelength. Additionally, when illuminated with a different, this time of known wavelength 0.137 nm, a second-order maximum is detected at 37.3° . Determine (a) the spacing between the reflecting planes, and (b) the unknown wavelength.

Exercise:**Problem:**

Calcite crystals contain scattering planes separated by 0.30 nm. What is the angular separation between first and second-order diffraction maxima when X-rays of 0.130 nm wavelength are used?

Solution:

13.2°

Exercise:**Problem:**

The first-order Bragg angle for a certain crystal is 12.1°. What is the second-order angle?

Glossary

Bragg planes

families of planes within crystals that can give rise to X-ray diffraction

X-ray diffraction

technique that provides the detailed information about crystallographic structure of natural and manufactured materials

Holography

By the end of this section, you will be able to:

- Describe how a three-dimensional image is recorded as a hologram
- Describe how a three-dimensional image is formed from a hologram

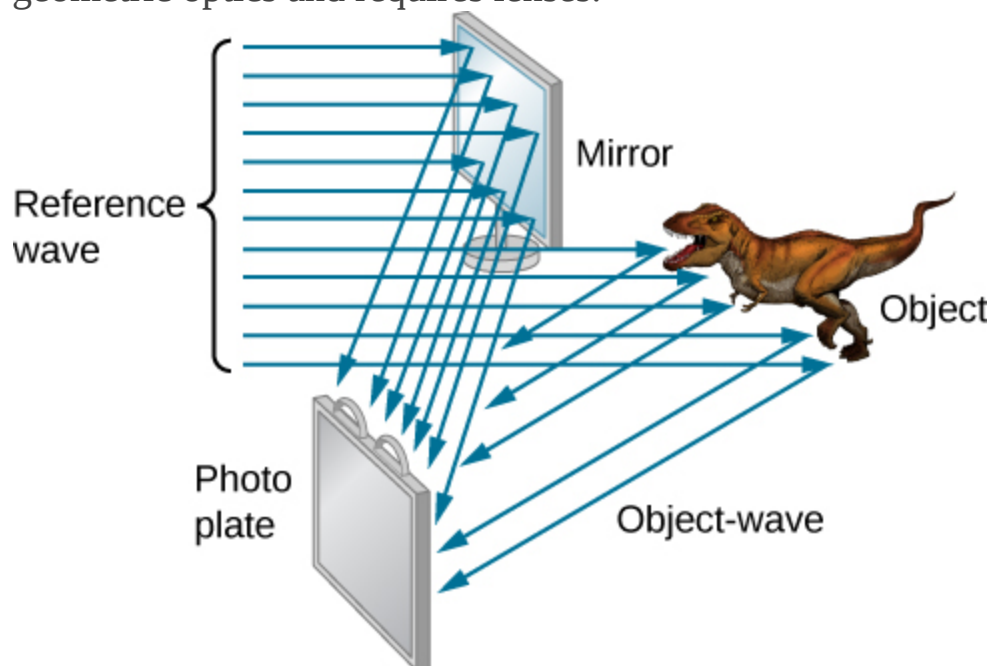
A **hologram**, such as the one in [\[link\]](#), is a true three-dimensional image recorded on film by lasers. Holograms are used for amusement; decoration on novelty items and magazine covers; security on credit cards and driver's licenses (a laser and other equipment are needed to reproduce them); and for serious three-dimensional information storage. You can see that a hologram is a true three-dimensional image because objects change relative position in the image when viewed from different angles.



Credit cards commonly have holograms for logos, making them difficult to reproduce. (credit: Dominic Alves)

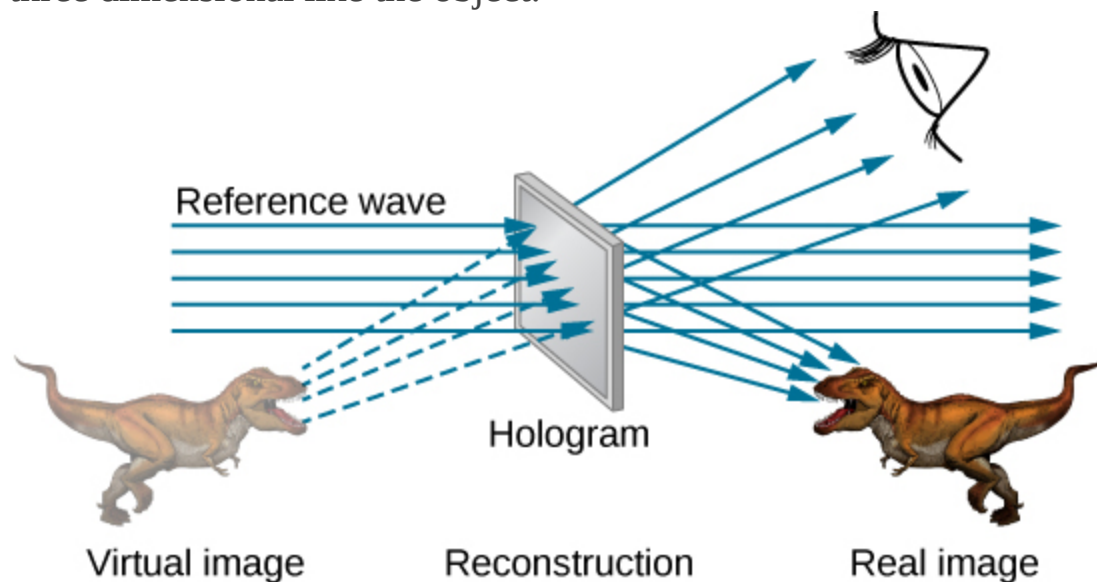
The name hologram means “entire picture” (from the Greek *holo*, as in holistic) because the image is three-dimensional. **Holography** is the process

of producing holograms and, although they are recorded on photographic film, the process is quite different from normal photography. Holography uses light interference or wave optics, whereas normal photography uses geometric optics. [\[link\]](#) shows one method of producing a hologram. Coherent light from a laser is split by a mirror, with part of the light illuminating the object. The remainder, called the reference beam, shines directly on a piece of film. Light scattered from the object interferes with the reference beam, producing constructive and destructive interference. As a result, the exposed film looks foggy, but close examination reveals a complicated interference pattern stored on it. Where the interference was constructive, the film (a negative actually) is darkened. Holography is sometimes called lens-less photography, because it uses the wave characteristics of light, as contrasted to normal photography, which uses geometric optics and requires lenses.



Production of a hologram. Single-wavelength coherent light from a laser produces a well-defined interference pattern on a piece of film. The laser beam is split by a partially silvered mirror, with part of the light illuminating the object and the remainder shining directly on the film. (credit: modification of work by Mariana Ruiz Villarreal)

Light falling on a hologram can form a three-dimensional image of the original object. The process is complicated in detail, but the basics can be understood, as shown in [\[link\]](#), in which a laser of the same type that exposed the film is now used to illuminate it. The myriad tiny exposed regions of the film are dark and block the light, whereas less exposed regions allow light to pass. The film thus acts much like a collection of diffraction gratings with various spacing patterns. Light passing through the hologram is diffracted in various directions, producing both real and virtual images of the object used to expose the film. The interference pattern is the same as that produced by the object. Moving your eye to various places in the interference pattern gives you different perspectives, just as looking directly at the object would. The image thus looks like the object and is three dimensional like the object.



A transmission hologram is one that produces real and virtual images when a laser of the same type as that which exposed the hologram is passed through it. Diffraction from various parts of the film produces the same interference pattern that was produced by the object that was used to expose it. (credit: modification of work by Mariana Ruiz Villarreal)

The hologram illustrated in [\[link\]](#) is a transmission hologram. Holograms that are viewed with reflected light, such as the white light holograms on credit cards, are reflection holograms and are more common. White light holograms often appear a little blurry with rainbow edges, because the diffraction patterns of various colors of light are at slightly different locations due to their different wavelengths. Further uses of holography include all types of three-dimensional information storage, such as of statues in museums, engineering studies of structures, and images of human organs.

Invented in the late 1940s by Dennis Gabor (1900–1970), who won the 1971 Nobel Prize in Physics for his work, holography became far more practical with the development of the laser. Since lasers produce coherent single-wavelength light, their interference patterns are more pronounced. The precision is so great that it is even possible to record numerous holograms on a single piece of film by just changing the angle of the film for each successive image. This is how the holograms that move as you walk by them are produced—a kind of lens-less movie.

In a similar way, in the medical field, holograms have allowed complete three-dimensional holographic displays of objects from a stack of images. Storing these images for future use is relatively easy. With the use of an endoscope, high-resolution, three-dimensional holographic images of internal organs and tissues can be made.

Summary

- Holography is a technique based on wave interference to record and form three-dimensional images.
- Lasers offer a practical way to produce sharp holographic images because of their monochromatic and coherent light for pronounced interference patterns.

Key Equations

Destructive interference for a single slit	$a \sin \theta = m\lambda$ for $m = \pm 1, \pm 2, \pm 3, \dots$
Half phase angle	$\beta = \frac{\phi}{2} = \frac{\pi a \sin \theta}{\lambda}$
Field amplitude in the diffraction pattern	$E = N \Delta E_0 \frac{\sin \beta}{\beta}$
Intensity in the diffraction pattern	$I = I_0 \left(\frac{\sin \beta}{\beta} \right)^2$
Rayleigh criterion for circular apertures	$\theta = 1.22 \frac{\lambda}{D}$
Bragg equation	$m\lambda = 2d \sin \theta, m = 1, 2, 3\dots$

Conceptual Questions

Exercise:

Problem:

How can you tell that a hologram is a true three-dimensional image and that those in three-dimensional movies are not?

Exercise:

Problem:

If a hologram is recorded using monochromatic light at one wavelength but its image is viewed at another wavelength, say 10% shorter, what will you see? What if it is viewed using light of exactly half the original wavelength?

Solution:

Image will appear at slightly different location and/or size when viewed using 10 % shorter wavelength but at exactly half the wavelength, a higher-order interference reconstructs the original image, different color.

Exercise:

Problem:

What image will one see if a hologram is recorded using monochromatic light but its image is viewed in white light? Explain.

Additional Problems

Exercise:

Problem:

White light falls on two narrow slits separated by 0.40 mm. The interference pattern is observed on a screen 3.0 m away. (a) What is the separation between the first maxima for red light ($\lambda = 700 \text{ nm}$) and violet light ($\lambda = 400 \text{ nm}$)? (b) At what point nearest the central maximum will a maximum for yellow light ($\lambda = 600 \text{ nm}$) coincide with a maximum for violet light? Identify the order for each maximum.

Solution:

a. 2.2 mm; b. 0.172° , second-order yellow and third-order violet coincide

Exercise:

Problem:

Microwaves of wavelength 10.0 mm fall normally on a metal plate that contains a slit 25 mm wide. (a) Where are the first minima of the diffraction pattern? (b) Would there be minima if the wavelength were 30.0 mm?

Exercise:**Problem:**

Quasars, or *quasi-stellar radio sources*, are astronomical objects discovered in 1960. They are distant but strong emitters of radio waves with angular size so small, they were originally unresolved, the same as stars. The quasar 3C405 is actually two discrete radio sources that subtend an angle of 82 arcsec. If this object is studied using radio emissions at a frequency of 410 MHz, what is the minimum diameter of a radio telescope that can resolve the two sources?

Solution:

2.2 km

Exercise:**Problem:**

Two slits each of width 1800 nm and separated by the center-to-center distance of 1200 nm are illuminated by plane waves from a krypton ion laser-emitting at wavelength 461.9 nm. Find the number of interference peaks in the central diffraction peak.

Exercise:**Problem:**

A microwave of an unknown wavelength is incident on a single slit of width 6 cm. The angular width of the central peak is found to be 25° . Find the wavelength.

Solution:

1.3 cm

Exercise:

Problem:

Red light (wavelength 632.8 nm in air) from a Helium-Neon laser is incident on a single slit of width 0.05 mm. The entire apparatus is immersed in water of refractive index 1.333. Determine the angular width of the central peak.

Exercise:**Problem:**

A light ray of wavelength 461.9 nm emerges from a 2-mm circular aperture of a krypton ion laser. Due to diffraction, the beam expands as it moves out. How large is the central bright spot at (a) 1 m, (b) 1 km, (c) 1000 km, and (d) at the surface of the moon at a distance of 400,000 km from Earth.

Solution:

a. 0.28 mm; b. 0.28 m; c. 280 m; d. 113 km

Exercise:**Problem:**

How far apart must two objects be on the moon to be distinguishable by eye if only the diffraction effects of the eye's pupil limit the resolution? Assume 550 nm for the wavelength of light, the pupil diameter 5.0 mm, and 400,000 km for the distance to the moon.

Exercise:**Problem:**

How far apart must two objects be on the moon to be resolvable by the 8.1-m-diameter Gemini North telescope at Mauna Kea, Hawaii, if only the diffraction effects of the telescope aperture limit the resolution? Assume 550 nm for the wavelength of light and 400,000 km for the distance to the moon.

Solution:

33 m

Exercise:

Problem:

A spy satellite is reputed to be able to resolve objects 10. cm apart while operating 197 km above the surface of Earth. What is the diameter of the aperture of the telescope if the resolution is only limited by the diffraction effects? Use 550 nm for light.

Exercise:

Problem:

Monochromatic light of wavelength 530 nm passes through a horizontal single slit of width $1.5 \mu\text{m}$ in an opaque plate. A screen of dimensions $2.0 \text{ m} \times 2.0 \text{ m}$ is 1.2 m away from the slit. (a) Which way is the diffraction pattern spread out on the screen? (b) What are the angles of the minima with respect to the center? (c) What are the angles of the maxima? (d) How wide is the central bright fringe on the screen? (e) How wide is the next bright fringe on the screen?

Solution:

a. vertically; b. $\pm 20^\circ$, $\pm 44^\circ$; c. 0, $\pm 31^\circ$, $\pm 60^\circ$; d. 89 cm; e. 71 cm

Exercise:

Problem:

A monochromatic light of unknown wavelength is incident on a slit of width $20 \mu\text{m}$. A diffraction pattern is seen at a screen 2.5 m away where the central maximum is spread over a distance of 10.0 cm. Find the wavelength.

Exercise:

Problem:

A source of light having two wavelengths 550 nm and 600 nm of equal intensity is incident on a slit of width $1.8\ \mu\text{m}$. Find the separation of the $m = 1$ bright spots of the two wavelengths on a screen 30.0 cm away.

Solution:

0.98 cm

Exercise:**Problem:**

A single slit of width 2100 nm is illuminated normally by a wave of wavelength 632.8 nm. Find the phase difference between waves from the top and one third from the bottom of the slit to a point on a screen at a horizontal distance of 2.0 m and vertical distance of 10.0 cm from the center.

Exercise:**Problem:**

A single slit of width $3.0\ \mu\text{m}$ is illuminated by a sodium yellow light of wavelength 589 nm. Find the intensity at a 15° angle to the axis in terms of the intensity of the central maximum.

Solution:

$$I/I_0 = 0.041$$

Exercise:**Problem:**

A single slit of width 0.10 mm is illuminated by a mercury lamp of wavelength 576 nm. Find the intensity at a 10° angle to the axis in terms of the intensity of the central maximum.

Exercise:**Problem:**

A diffraction grating produces a second maximum that is 89.7 cm from the central maximum on a screen 2.0 m away. If the grating has 600 lines per centimeter, what is the wavelength of the light that produces the diffraction pattern?

Solution:

340 nm

Exercise:**Problem:**

A grating with 4000 lines per centimeter is used to diffract light that contains all wavelengths between 400 and 650 nm. How wide is the first-order spectrum on a screen 3.0 m from the grating?

Exercise:**Problem:**

A diffraction grating with 2000 lines per centimeter is used to measure the wavelengths emitted by a hydrogen gas discharge tube. (a) At what angles will you find the maxima of the two first-order blue lines of wavelengths 410 and 434 nm? (b) The maxima of two other first-order lines are found at $\theta_1 = 0.097$ rad and $\theta_2 = 0.132$ rad. What are the wavelengths of these lines?

Solution:

a. 0.082 rad and 0.087 rad; b. 480 nm and 660 nm

Exercise:

Problem:

For white light ($400\text{ nm} < \lambda < 700\text{ nm}$) falling normally on a diffraction grating, show that the second and third-order spectra overlap no matter what the grating constant d is.

Exercise:**Problem:**

How many complete orders of the visible spectrum ($400\text{ nm} < \lambda < 700\text{ nm}$) can be produced with a diffraction grating that contains 5000 lines per centimeter?

Solution:

two orders

Exercise:**Problem:**

Two lamps producing light of wavelength 589 nm are fixed 1.0 m apart on a wooden plank. What is the maximum distance an observer can be and still resolve the lamps as two separate sources of light, if the resolution is affected solely by the diffraction of light entering the eye? Assume light enters the eye through a pupil of diameter 4.5 mm .

Exercise:**Problem:**

On a bright clear day, you are at the top of a mountain and looking at a city 12 km away. There are two tall towers 20.0 m apart in the city. Can your eye resolve the two towers if the diameter of the pupil is 4.0 mm ? If not, what should be the minimum magnification power of the telescope needed to resolve the two towers? In your calculations use 550 nm for the wavelength of the light.

Solution:

yes and N/A

Exercise:

Problem:

Radio telescopes are telescopes used for the detection of radio emission from space. Because radio waves have much longer wavelengths than visible light, the diameter of a radio telescope must be very large to provide good resolution. For example, the radio telescope in Penticton, BC in Canada, has a diameter of 26 m and can be operated at frequencies as high as 6.6 GHz. (a) What is the wavelength corresponding to this frequency? (b) What is the angular separation of two radio sources that can be resolved by this telescope? (c) Compare the telescope's resolution with the angular size of the moon.



(credit: modification of work by Jason Nishiyama)

Exercise:

Problem:

Calculate the wavelength of light that produces its first minimum at an angle of 36.9° when falling on a single slit of width $1.00\ \mu\text{m}$.

Solution:

600 nm

Exercise:**Problem:**

(a) Find the angle of the third diffraction minimum for 633-nm light falling on a slit of width $20.0\ \mu\text{m}$. (b) What slit width would place this minimum at 85.0° ?

Exercise:**Problem:**

As an example of diffraction by apertures of everyday dimensions, consider a doorway of width 1.0 m. (a) What is the angular position of the first minimum in the diffraction pattern of 600-nm light? (b) Repeat this calculation for a musical note of frequency 440 Hz (A above middle C). Take the speed of sound to be 343 m/s.

Solution:

a. $3.4 \times 10^{-5}^\circ$; b. 51°

Exercise:**Problem:**

What are the angular positions of the first and second minima in a diffraction pattern produced by a slit of width 0.20 mm that is illuminated by 400 nm light? What is the angular width of the central peak?

Exercise:**Problem:**

How far would you place a screen from the slit of the previous problem so that the second minimum is a distance of 2.5 mm from the center of the diffraction pattern?

Solution:

0.63 m

Exercise:

Problem:

How narrow is a slit that produces a diffraction pattern on a screen 1.8 m away whose central peak is 1.0 m wide? Assume $\lambda = 589 \text{ nm}$.

Exercise:

Problem:

Suppose that the central peak of a single-slit diffraction pattern is so wide that the first minima can be assumed to occur at angular positions of $\pm 90^\circ$. For this case, what is the ratio of the slit width to the wavelength of the light?

Solution:

1

Exercise:

Problem:

The central diffraction peak of the double-slit interference pattern contains exactly nine fringes. What is the ratio of the slit separation to the slit width?

Exercise:

Problem:

Determine the intensities of three interference peaks other than the central peak in the central maximum of the diffraction, if possible, when a light of wavelength 500 nm is incident normally on a double slit of width 1000 nm and separation 1500 nm. Use the intensity of the central spot to be 1 mW/cm^2 .

Solution:

0.17 mW/cm^2 for $m = 1$ only, no higher orders

Exercise:

Problem:

The yellow light from a sodium vapor lamp *seems* to be of pure wavelength, but it produces two first-order maxima at 36.093° and 36.129° when projected on a 10,000 line per centimeter diffraction grating. What are the two wavelengths to an accuracy of 0.1 nm?

Exercise:

Problem:

Structures on a bird feather act like a reflection grating having 8000 lines per centimeter. What is the angle of the first-order maximum for 600-nm light?

Solution:

28.7°

Exercise:

Problem:

If a diffraction grating produces a first-order maximum for the shortest wavelength of visible light at 30.0° , at what angle will the first-order maximum be for the largest wavelength of visible light?

Exercise:

Problem:

(a) What visible wavelength has its fourth-order maximum at an angle of 25.0° when projected on a 25,000-line per centimeter diffraction grating? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. 42.3 nm; b. This wavelength is not in the visible spectrum. c. The number of slits in this diffraction grating is too large. Etching in integrated circuits can be done to a resolution of 50 nm, so slit separations of 400 nm are at the limit of what we can do today. This line spacing is too small to produce diffraction of light.

Exercise:

Problem:

Consider a spectrometer based on a diffraction grating. Construct a problem in which you calculate the distance between two wavelengths of electromagnetic radiation in your spectrometer. Among the things to be considered are the wavelengths you wish to be able to distinguish, the number of lines per meter on the diffraction grating, and the distance from the grating to the screen or detector. Discuss the practicality of the device in terms of being able to discern between wavelengths of interest.

Exercise:

Problem:

An amateur astronomer wants to build a telescope with a diffraction limit that will allow him to see if there are people on the moons of Jupiter. (a) What diameter mirror is needed to be able to see 1.00-m detail on a Jovian moon at a distance of 7.50×10^8 km from Earth? The wavelength of light averages 600 nm. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. 549 km; b. This is an unreasonably large telescope. c. Unreasonable to assume diffraction limit for optical telescopes unless in space due to atmospheric effects.

Challenge Problems

Exercise:**Problem:**

Blue light of wavelength 450 nm falls on a slit of width 0.25 mm. A converging lens of focal length 20 cm is placed behind the slit and focuses the diffraction pattern on a screen. (a) How far is the screen from the lens? (b) What is the distance between the first and the third minima of the diffraction pattern?

Exercise:**Problem:**

(a) Assume that the maxima are halfway between the minima of a single-slit diffraction pattern. Use the diameter and circumference of the phasor diagram, as described in [Intensity in Single-Slit Diffraction](#), to determine the intensities of the third and fourth maxima in terms of the intensity of the central maximum. (b) Do the same calculation, using [\[link\]](#).

Solution:

a. $I = 0.00500 I_0$, $0.00335 I_0$; b. $I = 0.00500 I_0$, $0.00335 I_0$

Exercise:**Problem:**

(a) By differentiating [\[link\]](#), show that the higher-order maxima of the single-slit diffraction pattern occur at values of β that satisfy $\tan \beta = \beta$. (b) Plot $y = \tan \beta$ and $y = \beta$ versus β and find the intersections of these two curves. What information do they give you about the locations of the maxima? (c) Convince yourself that these points do not appear exactly at $\beta = (n + \frac{1}{2})\pi$, where $n = 0, 1, 2, \dots$, but are quite close to these values.

Exercise:

Problem:

What is the maximum number of lines per centimeter a diffraction grating can have and produce a complete first-order spectrum for visible light?

Solution:

12,800

Exercise:**Problem:**

Show that a diffraction grating cannot produce a second-order maximum for a given wavelength of light unless the first-order maximum is at an angle less than 30.0° .

Exercise:**Problem:**

A He-Ne laser beam is reflected from the surface of a CD onto a wall. The brightest spot is the reflected beam at an angle equal to the angle of incidence. However, fringes are also observed. If the wall is 1.50 m from the CD, and the first fringe is 0.600 m from the central maximum, what is the spacing of grooves on the CD?

Solution:

$1.58 \times 10^{-6} \text{ m}$

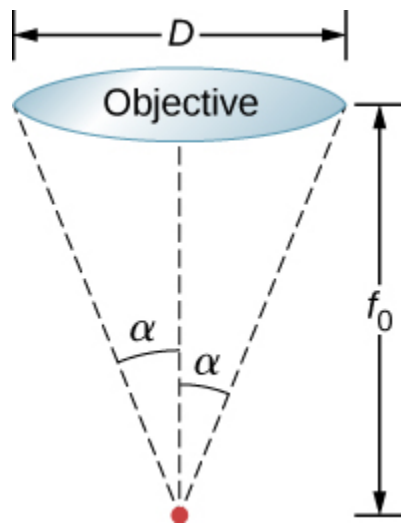
Exercise:**Problem:**

Objects viewed through a microscope are placed very close to the focal point of the objective lens. Show that the minimum separation x of two objects resolvable through the microscope is given by

Equation:

$$x = \frac{1.22\lambda f_0}{D},$$

where f_0 is the focal length and D is the diameter of the objective lens as shown below.



Glossary

hologram

three-dimensional image recorded on film by lasers; the word hologram means *entire picture* (from the Greek word *holo*, as in holistic)

holography

process of producing holograms with the use of lasers

Introduction

class="introduction"

Special
relativity
explains
how time
passes
slightly
differently
on Earth
and within
the rapidly
moving
global
positioning
satellite
(GPS). GPS
units in
vehicles
could not
find their
correct
location on
Earth
without
taking this
correction
into
account.
(credit:
modificatio
n of work
by U.S. Air
Force)



The special theory of relativity was proposed in 1905 by Albert Einstein (1879–1955). It describes how time, space, and physical phenomena appear in different frames of reference that are moving at constant velocity with respect to each other. This differs from Einstein's later work on general relativity, which deals with any frame of reference, including accelerated frames.

The theory of relativity led to a profound change in the way we perceive space and time. The “common sense” rules that we use to relate space and time measurements in the Newtonian worldview differ seriously from the correct rules at speeds near the speed of light. For example, the special theory of relativity tells us that measurements of length and time intervals are not the same in reference frames moving relative to one another. A particle might be observed to have a lifetime of 1.0×10^{-8} s in one reference frame, but a lifetime of 2.0×10^{-8} s in another; and an object might be measured to be 2.0 m long in one frame and 3.0 m long in another frame. These effects are usually significant only at speeds comparable to the speed of light, but even at the much lower speeds of the global positioning satellite, which requires extremely accurate time measurements to function, the different lengths of the same distance in different frames of reference are significant enough that they need to be taken into account.

Unlike Newtonian mechanics, which describes the motion of particles, or Maxwell's equations, which specify how the electromagnetic field behaves,

special relativity is not restricted to a particular type of phenomenon. Instead, its rules on space and time affect all fundamental physical theories.

The modifications of Newtonian mechanics in special relativity do not invalidate classical Newtonian mechanics or require its replacement. Instead, the equations of relativistic mechanics differ meaningfully from those of classical Newtonian mechanics only for objects moving at relativistic speeds (i.e., speeds less than, but comparable to, the speed of light). In the macroscopic world that you encounter in your daily life, the relativistic equations reduce to classical equations, and the predictions of classical Newtonian mechanics agree closely enough with experimental results to disregard relativistic corrections.

Invariance of Physical Laws

By the end of this section, you will be able to:

- Describe the theoretical and experimental issues that Einstein's theory of special relativity addressed.
- State the two postulates of the special theory of relativity.

Suppose you calculate the hypotenuse of a right triangle given the base angles and adjacent sides. Whether you calculate the hypotenuse from one of the sides and the cosine of the base angle, or from the Pythagorean theorem, the results should agree. Predictions based on different principles of physics must also agree, whether we consider them principles of mechanics or principles of electromagnetism.

Albert Einstein pondered a disagreement between predictions based on electromagnetism and on assumptions made in classical mechanics. Specifically, suppose an observer measures the velocity of a light pulse in the observer's own **rest frame**; that is, in the frame of reference in which the observer is at rest. According to the assumptions long considered obvious in classical mechanics, if an observer measures a velocity \vec{v} in one frame of reference, and that frame of reference is moving with velocity \vec{u} past a second reference frame, an observer in the second frame measures the original velocity as $\vec{v}' = \vec{v} + \vec{u}$. This sum of velocities is often referred to as **Galilean relativity**. If this principle is correct, the pulse of light that the observer measures as traveling with speed c travels at speed $c + u$ measured in the frame of the second observer. If we reasonably assume that the laws of electrodynamics are the same in both frames of reference, then the predicted speed of light (in vacuum) in both frames should be $c = 1/\sqrt{\epsilon_0\mu_0}$. Each observer should measure the same speed of the light pulse with respect to that observer's own rest frame. To reconcile difficulties of this kind, Einstein constructed his **special theory of relativity**, which introduced radical new ideas about time and space that have since been confirmed experimentally.

Inertial Frames

All velocities are measured relative to some frame of reference. For example, a car's motion is measured relative to its starting position on the road it travels on; a projectile's motion is measured relative to the surface from which it is launched; and a planet's orbital motion is measured relative to the star it orbits. The frames of reference in which mechanics takes the simplest form are those that are not accelerating. Newton's first law, the law of inertia, holds exactly in such a frame.

Note:

Inertial Reference Frame

An **inertial frame of reference** is a reference frame in which a body at rest remains at rest and a body in motion moves at a constant speed in a straight line unless acted upon by an outside force.

For example, to a passenger inside a plane flying at constant speed and constant altitude, physics seems to work exactly the same as when the passenger is standing on the surface of Earth. When the plane is taking off, however, matters are somewhat more complicated. In this case, the passenger at rest inside the plane concludes that a net force F on an object is not equal to the product of mass and acceleration, ma . Instead, F is equal to ma plus a fictitious force. This situation is not as simple as in an inertial frame. Special relativity handles accelerating frames as a constant and velocities as relative to the observer. General relativity treats both velocity and acceleration as relative to the observer, thus making the use of curved space-time.

Einstein's First Postulate

Not only are the principles of classical mechanics simplest in inertial frames, but they are the same in all inertial frames. Einstein based the **first postulate** of his theory on the idea that this is true for all the laws of physics, not merely those in mechanics.

Note:**First Postulate of Special Relativity**

The laws of physics are the same in all inertial frames of reference.

This postulate denies the existence of a special or preferred inertial frame. The laws of nature do not give us a way to endow any one inertial frame with special properties. For example, we cannot identify any inertial frame as being in a state of “absolute rest.” We can only determine the relative motion of one frame with respect to another.

There is, however, more to this postulate than meets the eye. The laws of physics include only those that satisfy this postulate. We will see that the definitions of energy and momentum must be altered to fit this postulate. Another outcome of this postulate is the famous equation $E = mc^2$, which relates energy to mass.

Einstein's Second Postulate

The second postulate upon which Einstein based his theory of special relativity deals with the speed of light. Late in the nineteenth century, the major tenets of classical physics were well established. Two of the most important were the laws of electromagnetism and Newton's laws. Investigations such as Young's double-slit experiment in the early 1800s had convincingly demonstrated that light is a wave. Maxwell's equations of electromagnetism implied that electromagnetic waves travel at $c = 3.00 \times 10^8$ m/s in a vacuum, but they do not specify the frame of reference in which light has this speed. Many types of waves were known, and all travelled in some medium. Scientists therefore assumed that some medium carried the light, even in a vacuum, and that light travels at a speed c relative to that medium (often called “the aether”).

Starting in the mid-1880s, the American physicist A.A. Michelson, later aided by E.W. Morley, made a series of direct measurements of the speed of light. They intended to deduce from their data the speed v at which Earth was moving through the mysterious medium for light waves. The speed of

light measured on Earth should have been $c + v$ when Earth's motion was opposite to the medium's flow at speed u past the Earth, and $c - v$ when Earth was moving in the same direction as the medium. The results of their measurements were startling.

Note:

Michelson-Morley Experiment

The **Michelson-Morley experiment** demonstrated that the speed of light in a vacuum is independent of the motion of Earth about the Sun.

The eventual conclusion derived from this result is that light, unlike mechanical waves such as sound, does not need a medium to carry it. Furthermore, the Michelson-Morley results implied that the speed of light c is independent of the motion of the source relative to the observer. That is, everyone observes light to move at speed c regardless of how they move relative to the light source or to one another. For several years, many scientists tried unsuccessfully to explain these results within the framework of Newton's laws.

In addition, there was a contradiction between the principles of electromagnetism and the assumption made in Newton's laws about relative velocity. Classically, the velocity of an object in one frame of reference and the velocity of that object in a second frame of reference relative to the first should combine like simple vectors to give the velocity seen in the second frame. If that were correct, then two observers moving at different speeds would see light traveling at different speeds. Imagine what a light wave would look like to a person traveling along with it (in vacuum) at a speed c . If such a motion were possible, then the wave would be stationary relative to the observer. It would have electric and magnetic fields whose strengths varied with position but were constant in time. This is not allowed by Maxwell's equations. So either Maxwell's equations are different in different inertial frames, or an object with mass cannot travel at speed c . Einstein concluded that the latter is true: An object with mass cannot travel

at speed c . Maxwell's equations are correct, but Newton's addition of velocities is not correct for light.

Not until 1905, when Einstein published his first paper on special relativity, was the currently accepted conclusion reached. Based mostly on his analysis that the laws of electricity and magnetism would not allow another speed for light, and only slightly aware of the Michelson-Morley experiment, Einstein detailed his **second postulate of special relativity**.

Note:

Second Postulate of Special Relativity

Light travels in a vacuum with the same speed c in any direction in all inertial frames.

In other words, the speed of light has the same definite speed for any observer, regardless of the relative motion of the source. This deceptively simple and counterintuitive postulate, along with the first postulate, leave all else open for change. Among the changes are the loss of agreement on the time between events, the variation of distance with speed, and the realization that matter and energy can be converted into one another. We describe these concepts in the following sections.

Note:

Exercise:

Problem:

Check Your Understanding Explain how special relativity differs from general relativity.

Solution:

Special relativity applies only to objects moving at constant velocity, whereas general relativity applies to objects that undergo acceleration.

Summary

- Relativity is the study of how observers in different reference frames measure the same event.
- Modern relativity is divided into two parts. Special relativity deals with observers in uniform (unaccelerated) motion, whereas general relativity includes accelerated relative motion and gravity. Modern relativity is consistent with all empirical evidence thus far and, in the limit of low velocity and weak gravitation, gives close agreement with the predictions of classical (Galilean) relativity.
- An inertial frame of reference is a reference frame in which a body at rest remains at rest and a body in motion moves at a constant speed in a straight line unless acted upon by an outside force.
- Modern relativity is based on Einstein's two postulates. The first postulate of special relativity is that the laws of physics are the same in all inertial frames of reference. The second postulate of special relativity is that the speed of light c is the same in all inertial frames of reference, independent of the relative motion of the observer and the light source.
- The Michelson-Morley experiment demonstrated that the speed of light in a vacuum is independent of the motion of Earth about the sun.

Conceptual Questions

Exercise:

Problem:

Which of Einstein's postulates of special relativity includes a concept that does not fit with the ideas of classical physics? Explain.

Solution:

the second postulate, involving the speed of light; classical physics already included the idea that the laws of mechanics, at least, were the same in all inertial frames, but the velocity of a light pulse was different in different frames moving with respect to each other

Exercise:

Problem:

Is Earth an inertial frame of reference? Is the sun? Justify your response.

Exercise:

Problem:

When you are flying in a commercial jet, it may appear to you that the airplane is stationary and Earth is moving beneath you. Is this point of view valid? Discuss briefly.

Solution:

yes, provided the plane is flying at constant velocity relative to the Earth; in that case, an object with no force acting on it within the plane has no change in velocity relative to the plane and no change in velocity relative to the Earth; both the plane and the ground are inertial frames for describing the motion of the object

Glossary

first postulate of special relativity

laws of physics are the same in all inertial frames of reference

Galilean relativity

if an observer measures a velocity in one frame of reference, and that frame of reference is moving with a velocity past a second reference frame, an observer in the second frame measures the original velocity as the vector sum of these velocities

inertial frame of reference

reference frame in which a body at rest remains at rest and a body in motion moves at a constant speed in a straight line unless acted on by an outside force

Michelson-Morley experiment

investigation performed in 1887 that showed that the speed of light in a vacuum is the same in all frames of reference from which it is viewed

rest frame

frame of reference in which the observer is at rest

second postulate of special relativity

light travels in a vacuum with the same speed c in any direction in all inertial frames

special theory of relativity

theory that Albert Einstein proposed in 1905 that assumes all the laws of physics have the same form in every inertial frame of reference, and that the speed of light is the same within all inertial frames

Relativity of Simultaneity

By the end of this section, you will be able to:

- Show from Einstein's postulates that two events measured as simultaneous in one inertial frame are not necessarily simultaneous in all inertial frames.
- Describe how simultaneity is a relative concept for observers in different inertial frames in relative motion.

Do time intervals depend on who observes them? Intuitively, it seems that the time for a process, such as the elapsed time for a foot race ([link](#)), should be the same for all observers. In everyday experiences, disagreements over elapsed time have to do with the accuracy of measuring time. No one would be likely to argue that the actual time interval was different for the moving runner and for the stationary clock displayed. Carefully considering just how time is measured, however, shows that elapsed time does depend on the relative motion of an observer with respect to the process being measured.



Elapsed time for a foot race is the same for all observers, but at relativistic speeds, elapsed time depends on the motion of the observer relative to the location where the process being timed occurs. (credit: "Jason Edward Scott Bain"/Flickr)

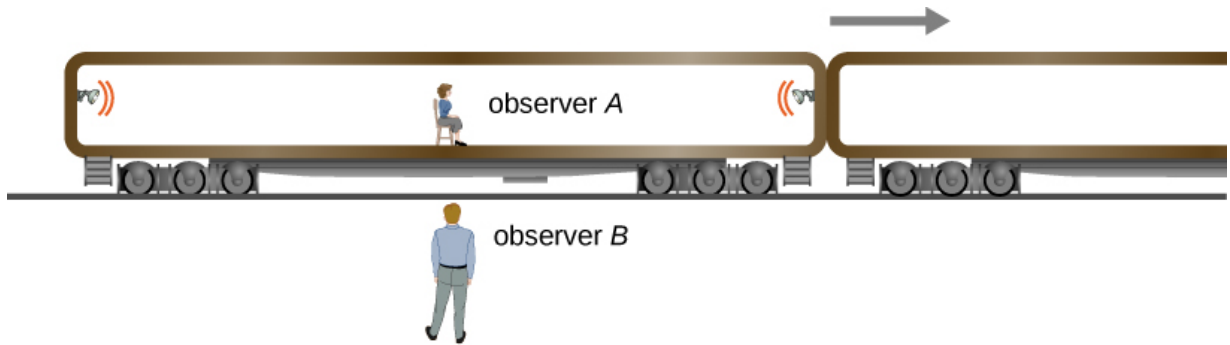
Consider how we measure elapsed time. If we use a stopwatch, for example, how do we know when to start and stop the watch? One method is to use the arrival of light from the event. For example, if you're in a moving car and observe the light arriving from a traffic signal change from green to red, you know it's time to step on the brake pedal. The timing is more accurate if some sort of electronic detection is used, avoiding human reaction times and other complications.

Now suppose two observers use this method to measure the time interval between two flashes of light from flash lamps that are a distance apart ([link](#)). An observer *A* is seated midway on a rail car with two flash lamps at opposite sides equidistant from her. A pulse of light is emitted from each flash lamp and moves toward observer *A*, shown in frame (a) of the figure. The rail car is moving rapidly in the direction indicated by the velocity vector in the diagram. An observer *B* standing on the platform is facing the rail car as it passes and observes both flashes of light reach him simultaneously, as shown in frame (c). He measures the distances from where he saw the pulses originate, finds them equal, and concludes that the pulses were emitted simultaneously.

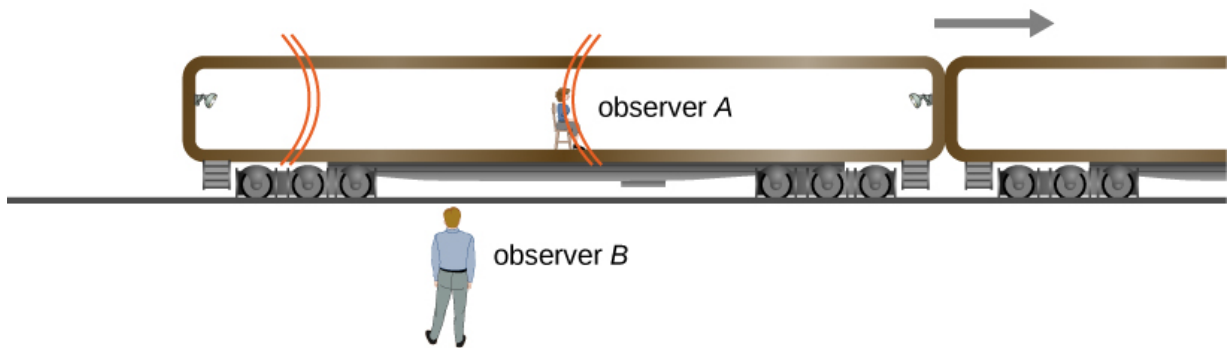
However, because of Observer *A*'s motion, the pulse from the right of the railcar, from the direction the car is moving, reaches her before the pulse from the left, as shown in frame (b). She also measures the distances from within her frame of reference, finds them equal, and concludes that the pulses were not emitted simultaneously.

The two observers reach conflicting conclusions about whether the two events at well-separated locations were simultaneous. Both frames of reference are valid, and both conclusions are valid. Whether two events at

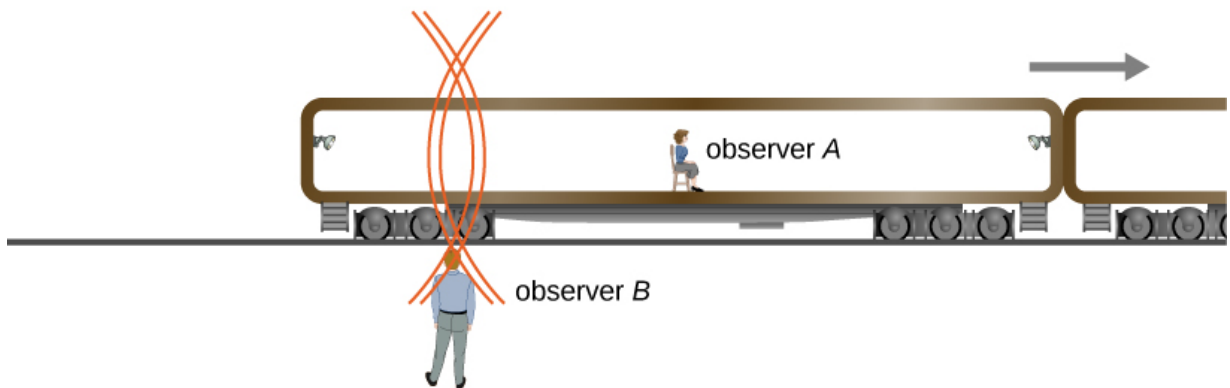
separate locations are simultaneous depends on the motion of the observer relative to the locations of the events.



(a)



(b)



(c)

- (a) Two pulses of light are emitted simultaneously relative to observer B. (c) The pulses reach observer B's position simultaneously. (b) Because of A's motion, she sees the pulse from the right first and

concludes the bulbs did not flash simultaneously. Both conclusions are correct.

Here, the relative velocity between observers affects whether two events a distance apart are observed to be simultaneous. *Simultaneity is not absolute.* We might have guessed (incorrectly) that if light is emitted simultaneously, then two observers halfway between the sources would see the flashes simultaneously. But careful analysis shows this cannot be the case if the speed of light is the same in all inertial frames.

This type of *thought experiment* (in German, “Gedankenexperiment”) shows that seemingly obvious conclusions must be changed to agree with the postulates of relativity. The validity of thought experiments can only be determined by actual observation, and careful experiments have repeatedly confirmed Einstein’s theory of relativity.

Summary

- Two events are defined to be simultaneous if an observer measures them as occurring at the same time (such as by receiving light from the events).
- Two events at locations a distance apart that are simultaneous for an observer at rest in one frame of reference are not necessarily simultaneous for an observer at rest in a different frame of reference.

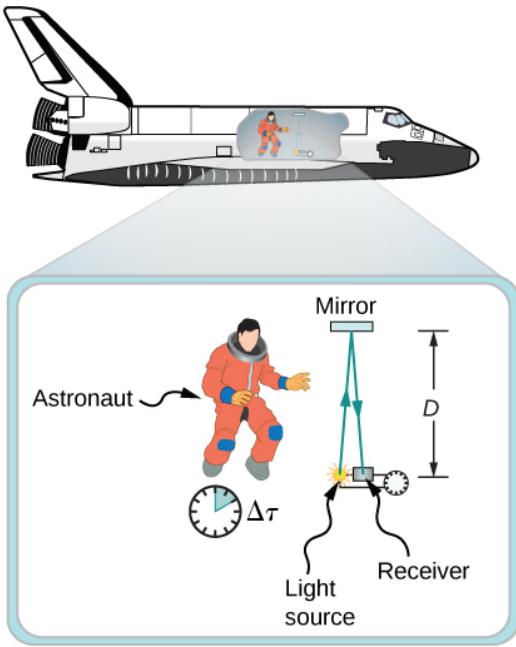
Time Dilation

By the end of this section, you will be able to:

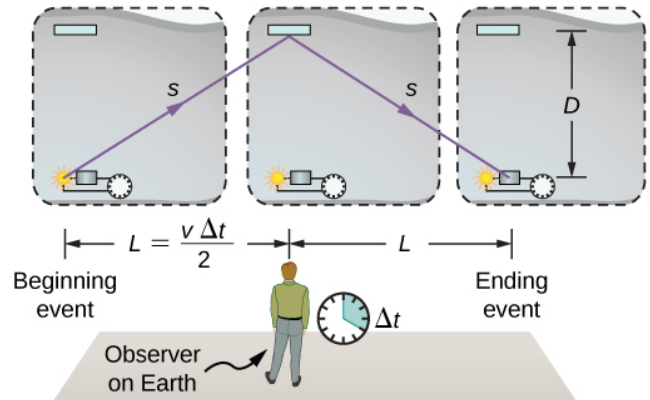
- Explain how time intervals can be measured differently in different reference frames.
- Describe how to distinguish a proper time interval from a dilated time interval.
- Describe the significance of the muon experiment.
- Explain why the twin paradox is not a contradiction.
- Calculate time dilation given the speed of an object in a given frame.

The analysis of simultaneity shows that Einstein's postulates imply an important effect: Time intervals have different values when measured in different inertial frames. Suppose, for example, an astronaut measures the time it takes for a pulse of light to travel a distance perpendicular to the direction of his ship's motion (relative to an earthbound observer), bounce off a mirror, and return ([\[link\]](#)). How does the elapsed time that the astronaut measures in the spacecraft compare with the elapsed time that an earthbound observer measures by observing what is happening in the spacecraft?

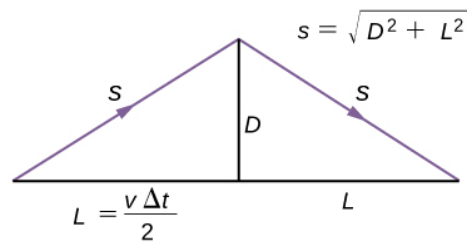
Examining this question leads to a profound result. The elapsed time for a process depends on which observer is measuring it. In this case, the time measured by the astronaut (within the spaceship where the astronaut is at rest) is smaller than the time measured by the earthbound observer (to whom the astronaut is moving). The time elapsed for the same process is different for the observers, because the distance the light pulse travels in the astronaut's frame is smaller than in the earthbound frame, as seen in [\[link\]](#). Light travels at the same speed in each frame, so it takes more time to travel the greater distance in the earthbound frame.



(a)



(b)



(c)

(a) An astronaut measures the time $\Delta\tau$ for light to travel distance $2D$ in the astronaut's frame. (b) A NASA scientist on Earth sees the light follow the longer path $2s$ and take a longer time Δt . (c) These triangles are used to find the relationship between the two distances D and s .

Note:
Time Dilation

Time dilation is the lengthening of the time interval between two events for an observer in an inertial frame that is moving with respect to the rest frame of the events (in which the events occur at the same location).

To quantitatively compare the time measurements in the two inertial frames, we can relate the distances in [\[link\]](#) to each other, then express each distance in terms of the time of travel (respectively either Δt or $\Delta \tau$) of the pulse in the corresponding reference frame. The resulting equation can then be solved for Δt in terms of $\Delta \tau$.

The lengths D and L in [\[link\]](#) are the sides of a right triangle with hypotenuse s . From the Pythagorean theorem,

Equation:

$$s^2 = D^2 + L^2.$$

The lengths $2s$ and $2L$ are, respectively, the distances that the pulse of light and the spacecraft travel in time Δt in the earthbound observer's frame. The length D is the distance that the light pulse travels in time $\Delta \tau$ in the astronaut's frame. This gives us three equations:

Equation:

$$2s = c\Delta t; 2L = v\Delta t; 2D = c\Delta \tau.$$

Note that we used Einstein's second postulate by taking the speed of light to be c in both inertial frames. We substitute these results into the previous expression from the Pythagorean theorem:

Equation:

$$\begin{aligned} s^2 &= D^2 + L^2 \\ \left(c \frac{\Delta t}{2}\right)^2 &= \left(c \frac{\Delta \tau}{2}\right)^2 + \left(v \frac{\Delta t}{2}\right)^2. \end{aligned}$$

Then we rearrange to obtain

Equation:

$$(c\Delta t)^2 - (v\Delta t)^2 = (c\Delta\tau)^2.$$

Finally, solving for Δt in terms of $\Delta\tau$ gives us

Note:

Equation:

$$\Delta t = \frac{\Delta\tau}{\sqrt{1 - (v/c)^2}}.$$

This is equivalent to

Equation:

$$\Delta t = \gamma\Delta\tau,$$

where γ is the relativistic factor (often called the Lorentz factor) given by

Note:

Equation:

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

and v and c are the speeds of the moving observer and light, respectively.

Note the asymmetry between the two measurements. Only one of them is a measurement of the time interval between two events—the emission and arrival of the light pulse—at the same position. It is a measurement of the time interval in the rest frame of a single clock. The measurement in the earthbound frame involves comparing the time interval between two events that occur at different locations. The time interval between events that occur at a single location has a separate name to distinguish it from the time measured by the earthbound observer, and we use the separate symbol $\Delta\tau$ to refer to it throughout this chapter.

Note:

Proper Time

The **proper time** interval $\Delta\tau$ between two events is the time interval measured by an observer for whom both events occur at the same location.

The equation relating Δt and $\Delta\tau$ is truly remarkable. First, as stated earlier, elapsed time is not the same for different observers moving relative to one another, even though both are in inertial frames. A proper time interval $\Delta\tau$ for an observer who, like the astronaut, is moving with the apparatus, is smaller than the time interval for other observers. It is the smallest possible measured time between two events. The earthbound observer sees time intervals within the moving system as dilated (i.e., lengthened) relative to how the observer moving relative to Earth sees them within the moving system. Alternatively, according to the earthbound observer, less time passes between events within the moving frame. Note that the shortest elapsed time between events is in the inertial frame in which the observer sees the events (e.g., the emission and arrival of the light signal) occur at the same point.

This time effect is real and is not caused by inaccurate clocks or improper measurements. Time-interval measurements of the same event differ for observers in relative motion. The dilation of time is an intrinsic property of

time itself. All clocks moving relative to an observer, including biological clocks, such as a person's heartbeat, or aging, are observed to run more slowly compared with a clock that is stationary relative to the observer.

Note that if the relative velocity is much less than the speed of light ($v \ll c$), then v^2/c^2 is extremely small, and the elapsed times Δt and $\Delta \tau$ are nearly equal. At low velocities, physics based on modern relativity approaches classical physics—everyday experiences involve very small relativistic effects. However, for speeds near the speed of light, v^2/c^2 is close to one, so $\sqrt{1 - v^2/c^2}$ is very small and Δt becomes significantly larger than $\Delta \tau$.

Half-Life of a Muon

There is considerable experimental evidence that the equation $\Delta t = \gamma \Delta \tau$ is correct. One example is found in cosmic ray particles that continuously rain down on Earth from deep space. Some collisions of these particles with nuclei in the upper atmosphere result in short-lived particles called muons. The half-life (amount of time for half of a material to decay) of a muon is $1.52 \mu\text{s}$ when it is at rest relative to the observer who measures the half-life. This is the proper time interval $\Delta \tau$. This short time allows very few muons to reach Earth's surface and be detected if Newtonian assumptions about time and space were correct. However, muons produced by cosmic ray particles have a range of velocities, with some moving near the speed of light. It has been found that the muon's half-life as measured by an earthbound observer (Δt) varies with velocity exactly as predicted by the equation $\Delta t = \gamma \Delta \tau$. The faster the muon moves, the longer it lives. We on Earth see the muon last much longer than its half-life predicts within its own rest frame. As viewed from our frame, the muon decays more slowly than it does when at rest relative to us. A far larger fraction of muons reach the ground as a result.

Before we present the first example of solving a problem in relativity, we state a strategy you can use as a guideline for these calculations.

Note:**Relativity**

1. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). Look in particular for information on relative velocity v .
2. Identify exactly what needs to be determined in the problem (identify the unknowns).
3. Make certain you understand the conceptual aspects of the problem before making any calculations (express the answer as an equation). Decide, for example, which observer sees time dilated or length contracted before working with the equations or using them to carry out the calculation. If you have thought about who sees what, who is moving with the event being observed, who sees proper time, and so on, you will find it much easier to determine if your calculation is reasonable.
4. Determine the primary type of calculation to be done to find the unknowns identified above (do the calculation). You will find the section summary helpful in determining whether a length contraction, relativistic kinetic energy, or some other concept is involved.

Note that you should not round off during the calculation. As noted in the text, you must often perform your calculations to many digits to see the desired effect. You may round off at the very end of the problem solution, but do not use a rounded number in a subsequent calculation. Also, check the answer to see if it is reasonable: Does it make sense? This may be more difficult for relativity, which has few everyday examples to provide experience with what is reasonable. But you can look for velocities greater than c or relativistic effects that are in the wrong direction (such as a time contraction where a dilation was expected).

Example:**Time Dilation in a High-Speed Vehicle**

The Hypersonic Technology Vehicle 2 (HTV-2) is an experimental rocket vehicle capable of traveling at 21,000 km/h (5830 m/s). If an electronic clock in the HTV-2 measures a time interval of exactly 1-s duration, what would observers on Earth measure the time interval to be?

Strategy

Apply the time dilation formula to relate the proper time interval of the signal in HTV-2 to the time interval measured on the ground.

Solution

- a. Identify the knowns: $\Delta\tau = 1 \text{ s}$; $v = 5830 \text{ m/s}$.
- b. Identify the unknown: Δt .
- c. Express the answer as an equation:

Equation:

$$\Delta t = \gamma \Delta\tau = \frac{\Delta\tau}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

- d. Do the calculation. Use the expression for γ to determine Δt from $\Delta\tau$:

Equation:

$$\begin{aligned}\Delta t &= \frac{1 \text{ s}}{\sqrt{1 - \left(\frac{5830 \text{ m/s}}{3.00 \times 10^8 \text{ m/s}}\right)^2}} \\ &= 1.000000000189 \text{ s} \\ &= 1 \text{ s} + 1.89 \times 10^{-10} \text{ s}.\end{aligned}$$

Significance

The very high speed of the HTV-2 is still only 10^{-5} times the speed of light. Relativistic effects for the HTV-2 are negligible for almost all purposes, but are not zero.

Example:

What Speeds are Relativistic?

How fast must a vehicle travel for 1 second of time measured on a passenger's watch in the vehicle to differ by 1% for an observer measuring it from the ground outside?

Strategy

Use the time dilation formula to find v/c for the given ratio of times.

Solution

- a. Identify the known:

Equation:

$$\frac{\Delta\tau}{\Delta t} = \frac{1}{1.01}.$$

- b. Identify the unknown: v/c .

- c. Express the answer as an equation:

Equation:

$$\begin{aligned}\Delta t &= \gamma \Delta\tau = \frac{1}{\sqrt{1-v^2/c^2}} \Delta\tau \\ \frac{\Delta\tau}{\Delta t} &= \sqrt{1-v^2/c^2} \\ \left(\frac{\Delta\tau}{\Delta t}\right)^2 &= 1 - \frac{v^2}{c^2} \\ \frac{v}{c} &= \sqrt{1 - (\Delta\tau/\Delta t)^2}.\end{aligned}$$

- d. Do the calculation:

Equation:

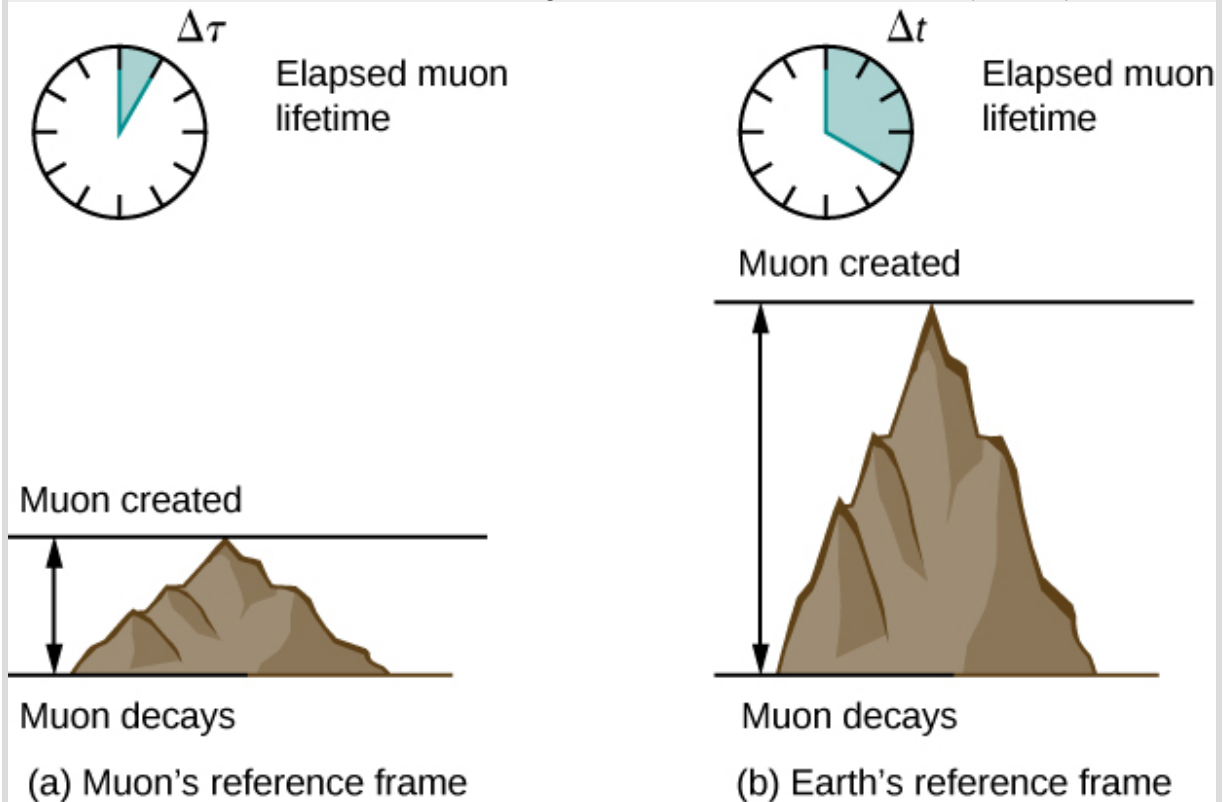
$$\begin{aligned}\frac{v}{c} &= \sqrt{1 - (1/1.01)^2} \\ &= 0.14.\end{aligned}$$

Significance

The result shows that an object must travel at very roughly 10% of the speed of light for its motion to produce significant relativistic time dilation effects.

Example:**Calculating Δt for a Relativistic Event**

Suppose a cosmic ray colliding with a nucleus in Earth's upper atmosphere produces a muon that has a velocity $v = 0.950c$. The muon then travels at constant velocity and lives $2.20 \mu\text{s}$ as measured in the muon's frame of reference. (You can imagine this as the muon's internal clock.) How long does the muon live as measured by an earthbound observer ([link](#))?



A muon in Earth's atmosphere lives longer as measured by an earthbound observer than as measured by the muon's internal clock.

As we will discuss later, in the muon's reference frame, it travels a shorter distance than measured in Earth's reference frame.

Strategy

A clock moving with the muon measures the proper time of its decay process, so the time we are given is $\Delta\tau = 2.20 \mu\text{s}$. The earthbound observer measures Δt as given by the equation $\Delta t = \gamma \Delta\tau$. Because the velocity is given, we can calculate the time in Earth's frame of reference.

Solution

- Identify the knowns: $v = 0.950c$, $\Delta\tau = 2.20\mu\text{s}$.
- Identify the unknown: Δt .
- Express the answer as an equation. Use:

Equation:

$$\Delta t = \gamma \Delta \tau$$

with

Equation:

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

- Do the calculation. Use the expression for γ to determine Δt from $\Delta\tau$

Equation:

$$\begin{aligned}\Delta t &= \gamma \Delta \tau \\ &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \Delta \tau \\ &= \frac{2.20\mu\text{s}}{\sqrt{1 - (0.950)^2}} \\ &= 7.05 \mu\text{s}.\end{aligned}$$

Remember to keep extra significant figures until the final answer.

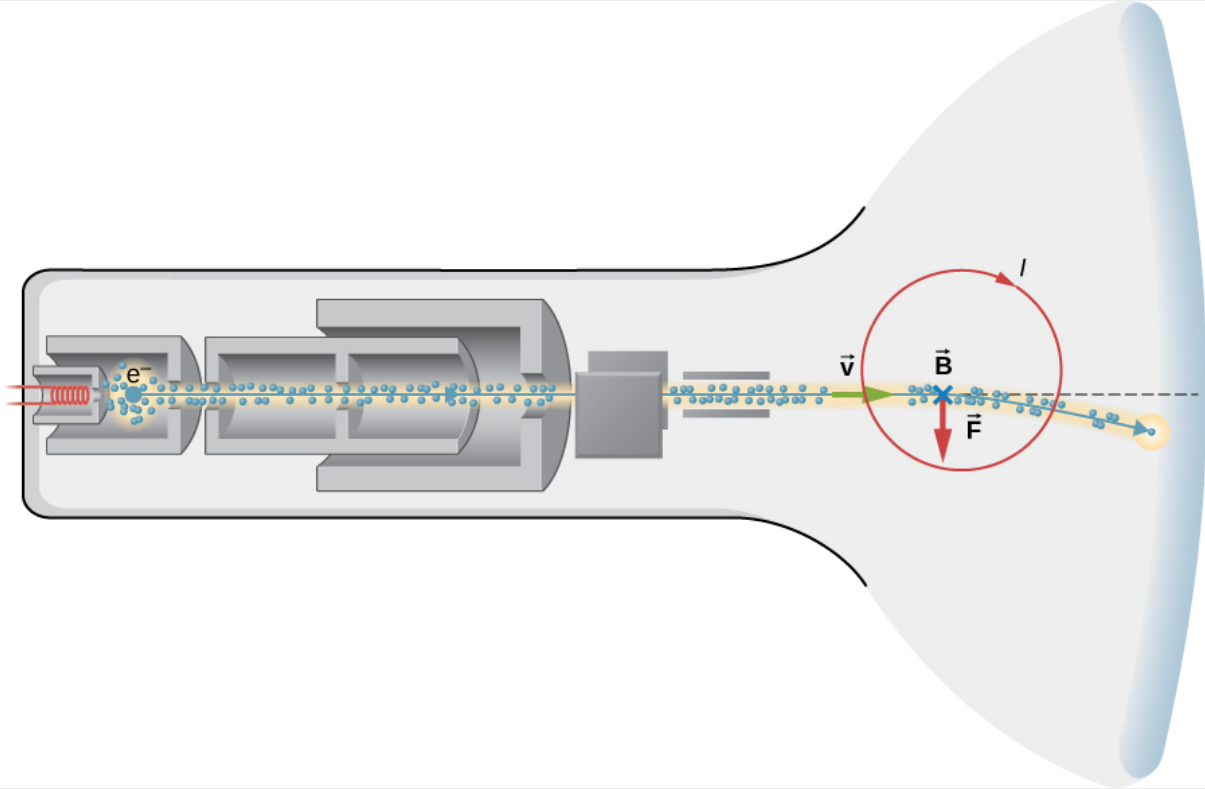
Significance

One implication of this example is that because $\gamma = 3.20$ at 95.0% of the speed of light ($v = 0.950c$), the relativistic effects are significant. The two time intervals differ by a factor of 3.20, when classically they would be the same. Something moving at $0.950c$ is said to be highly relativistic.

Example:

Relativistic Television

A non-flat screen, older-style television display ([\[link\]](#)) works by accelerating electrons over a short distance to relativistic speed, and then using electromagnetic fields to control where the electron beam strikes a fluorescent layer at the front of the tube. Suppose the electrons travel at $6.00 \times 10^7 \text{ m/s}$ through a distance of 0.200 m from the start of the beam to the screen. (a) What is the time of travel of an electron in the rest frame of the television set? (b) What is the electron's time of travel in its own rest frame?



The electron beam in a cathode ray tube television display.

Strategy for (a)

(a) Calculate the time from $vt = d$. Even though the speed is relativistic, the calculation is entirely in one frame of reference, and relativity is therefore not involved.

Solution

a. Identify the knowns:

Equation:

$$v = 6.00 \times 10^7 \text{ m/s}; d = 0.200 \text{ m}.$$

b. Identify the unknown: the time of travel Δt .

c. Express the answer as an equation:

Equation:

$$\Delta t = \frac{d}{v}.$$

d. Do the calculation:

Equation:

$$\begin{aligned} t &= \frac{0.200 \text{ m}}{6.00 \times 10^7 \text{ m/s}} \\ &= 3.33 \times 10^{-9} \text{ s}. \end{aligned}$$

Significance

The time of travel is extremely short, as expected. Because the calculation is entirely within a single frame of reference, relativity is not involved, even though the electron speed is close to c .

Strategy for (b)

(b) In the frame of reference of the electron, the vacuum tube is moving and the electron is stationary. The electron-emitting cathode leaves the electron and the front of the vacuum tube strikes the electron with the electron at the same location. Therefore we use the time dilation formula to relate the proper time in the electron rest frame to the time in the television frame.

Solution

a. Identify the knowns (from part a):

Equation:

$$\Delta t = 3.33 \times 10^{-9} \text{ s}; v = 6.00 \times 10^7 \text{ m/s}; d = 0.200 \text{ m}.$$

b. Identify the unknown: τ .

c. Express the answer as an equation:

Equation:

$$\Delta t = \gamma \Delta \tau = \frac{\Delta \tau}{\sqrt{1-v^2/c^2}}$$

$$\Delta \tau = \Delta t \sqrt{1-v^2/c^2}.$$

d. Do the calculation:

Equation:

$$\begin{aligned}\Delta \tau &= (3.33 \times 10^{-9} \text{ s}) \sqrt{1 - \left(\frac{6.00 \times 10^7 \text{ m/s}}{3.00 \times 10^8 \text{ m/s}} \right)^2} \\ &= 3.26 \times 10^{-9} \text{ s}.\end{aligned}$$

Significance

The time of travel is shorter in the electron frame of reference. Because the problem requires finding the time interval measured in different reference frames for the same process, relativity is involved. If we had tried to calculate the time in the electron rest frame by simply dividing the 0.200 m by the speed, the result would be slightly incorrect because of the relativistic speed of the electron.

Note:

Exercise:

Problem: Check Your Understanding What is γ if $v = 0.650c$?

Solution:

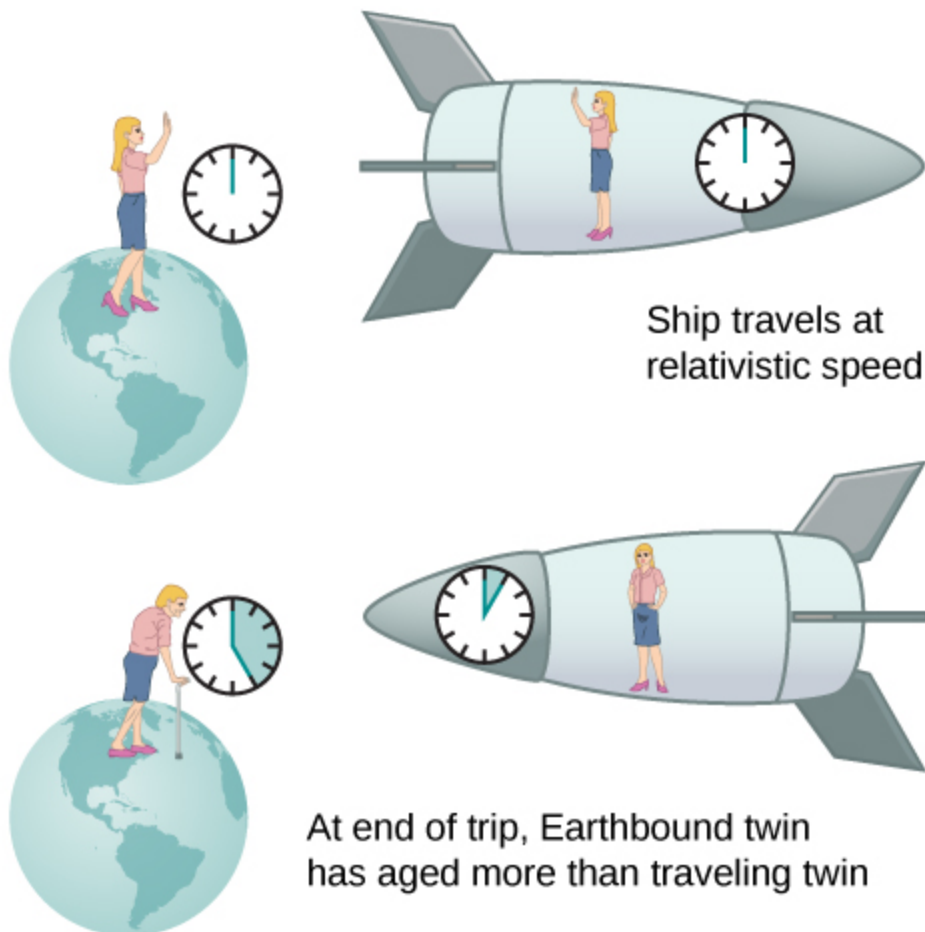
$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{1}{\sqrt{1-\frac{(0.650c)^2}{c^2}}} = 1.32$$

The Twin Paradox

An intriguing consequence of time dilation is that a space traveler moving at a high velocity relative to Earth would age less than the astronaut's earthbound twin. This is often known as the twin paradox. Imagine the astronaut moving at such a velocity that $\gamma = 30.0$, as in [\[link\]](#). A trip that takes 2.00 years in her frame would take 60.0 years in the earthbound twin's frame. Suppose the astronaut travels 1.00 year to another star system, briefly explores the area, and then travels 1.00 year back. An astronaut who was 40 years old at the start of the trip would be 42 when the spaceship returns. Everything on Earth, however, would have aged 60.0 years. The earthbound twin, if still alive, would be 100 years old.

The situation would seem different to the astronaut in [\[link\]](#). Because motion is relative, the spaceship would seem to be stationary and Earth would appear to move. (This is the sensation you have when flying in a jet.) Looking out the window of the spaceship, the astronaut would see time slow down on Earth by a factor of $\gamma = 30.0$. Seen from the spaceship, the earthbound sibling will have aged only $2/30$, or 0.07, of a year, whereas the astronaut would have aged 2.00 years.

At start of trip, both twins are same age



The twin paradox consists of the conflicting conclusions about which twin ages more as a result of a long space journey at relativistic speed.

The paradox here is that the two twins cannot both be correct. As with all paradoxes, conflicting conclusions come from a false premise. In fact, the astronaut's motion is significantly different from that of the earthbound twin. The astronaut accelerates to a high velocity and then accelerates opposite to the motion to view the star system. To return to Earth, she again accelerates and decelerates. The spacecraft is not in a single inertial frame to which the time dilation formula can be directly applied. That is, the astronaut twin changes inertial references. The earthbound twin does not

experience these accelerations and remains in the same inertial frame. Thus, the situation is not symmetric, and it is incorrect to claim that the astronaut observes the same effects as her twin. The lack of symmetry between the twins will be still more evident when we analyze the journey later in this chapter in terms of the path the astronaut follows through four-dimensional space-time.

In 1971, American physicists Joseph Hafele and Richard Keating verified time dilation at low relative velocities by flying extremely accurate atomic clocks around the world on commercial aircraft. They measured elapsed time to an accuracy of a few nanoseconds and compared it with the time measured by clocks left behind. Hafele and Keating's results were within experimental uncertainties of the predictions of relativity. Both special and general relativity had to be taken into account, because gravity and accelerations were involved as well as relative motion.

Note:

Exercise:

Problem:

Check Your Understanding a. A particle travels at $1.90 \times 10^8 \text{ m/s}$ and lives $2.10 \times 10^{-8} \text{ s}$ when at rest relative to an observer. How long does the particle live as viewed in the laboratory?

Solution:

$$\text{a. } \Delta t = \frac{\Delta \tau}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{2.10 \times 10^{-8} \text{ s}}{\sqrt{1 - \frac{(1.90 \times 10^8 \text{ m/s})^2}{(3.00 \times 10^8 \text{ m/s})^2}}} = 2.71 \times 10^{-8} \text{ s.}$$

Exercise:

Problem:

b. Spacecraft *A* and *B* pass in opposite directions at a relative speed of $4.00 \times 10^7 \text{ m/s}$. An internal clock in spacecraft *A* causes it to emit a radio signal for 1.00 s. The computer in spacecraft *B* corrects for the beginning and end of the signal having traveled different distances, to calculate the time interval during which ship *A* was emitting the signal. What is the time interval that the computer in spacecraft *B* calculates?

Solution:

b. Only the relative speed of the two spacecraft matters because there is no absolute motion through space. The signal is emitted from a fixed location in the frame of reference of *A*, so the proper time interval of its emission is $\tau = 1.00 \text{ s}$. The duration of the signal measured from frame of reference *B* is then

$$\Delta t = \frac{\Delta \tau}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1.00 \text{ s}}{\sqrt{1 - \frac{(4.00 \times 10^7 \text{ m/s})^2}{(3.00 \times 10^8 \text{ m/s})^2}}} = 1.01 \text{ s}.$$

Summary

- Two events are defined to be simultaneous if an observer measures them as occurring at the same time. They are not necessarily simultaneous to all observers—simultaneity is not absolute.
- Time dilation is the lengthening of the time interval between two events when seen in a moving inertial frame rather than the rest frame of the events (in which the events occur at the same location).
- Observers moving at a relative velocity v do not measure the same elapsed time between two events. Proper time $\Delta \tau$ is the time measured in the reference frame where the start and end of the time interval occur at the same location. The time interval Δt measured by an observer who sees the frame of events moving at speed v is related to the proper time interval $\Delta \tau$ of the events by the equation:

Equation:

$$\Delta t = \frac{\Delta \tau}{\sqrt{1 - \frac{v^2}{c^2}}} = \gamma \Delta \tau,$$

where

Equation:

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

- The premise of the twin paradox is faulty because the traveling twin is accelerating. The journey is not symmetrical for the two twins.
- Time dilation is usually negligible at low relative velocities, but it does occur, and it has been verified by experiment.
- The proper time is the shortest measure of any time interval. Any observer who is moving relative to the system being observed measures a time interval longer than the proper time.

Conceptual Questions

Exercise:

Problem:

(a) Does motion affect the rate of a clock as measured by an observer moving with it? (b) Does motion affect how an observer moving relative to a clock measures its rate?

Exercise:

Problem:

To whom does the elapsed time for a process seem to be longer, an observer moving relative to the process or an observer moving with the process? Which observer measures the interval of proper time?

Solution:

The observer moving with the process sees its interval of proper time, which is the shortest seen by any observer.

Exercise:**Problem:**

(a) How could you travel far into the future of Earth without aging significantly? (b) Could this method also allow you to travel into the past?

Problems**Exercise:**

Problem: (a) What is γ if $v = 0.250c$? (b) If $v = 0.500c$?

Solution:

a. 1.0328; b. 1.15

Exercise:

Problem: (a) What is γ if $v = 0.100c$? (b) If $v = 0.900c$?

Exercise:**Problem:**

Particles called π -mesons are produced by accelerator beams. If these particles travel at $2.70 \times 10^8 \text{ m/s}$ and live $2.60 \times 10^{-8} \text{ s}$ when at rest relative to an observer, how long do they live as viewed in the laboratory?

Solution:

$$5.96 \times 10^{-8} \text{ s}$$

Exercise:**Problem:**

Suppose a particle called a kaon is created by cosmic radiation striking the atmosphere. It moves by you at $0.980c$, and it lives $1.24 \times 10^{-8} \text{ s}$ when at rest relative to an observer. How long does it live as you observe it?

Exercise:**Problem:**

A neutral π -meson is a particle that can be created by accelerator beams. If one such particle lives $1.40 \times 10^{-16} \text{ s}$ as measured in the laboratory, and $0.840 \times 10^{-16} \text{ s}$ when at rest relative to an observer, what is its velocity relative to the laboratory?

Solution:

$$0.800c$$

Exercise:**Problem:**

A neutron lives 900 s when at rest relative to an observer. How fast is the neutron moving relative to an observer who measures its life span to be 2065 s?

Exercise:**Problem:**

If relativistic effects are to be less than 1%, then γ must be less than 1.01. At what relative velocity is $\gamma = 1.01$?

Solution:

$$0.140c$$

Exercise:

Problem:

If relativistic effects are to be less than 3%, then γ must be less than 1.03. At what relative velocity is $\gamma = 1.03$?

Glossary

proper time

$\Delta\tau$ is the time interval measured by an observer who sees the beginning and end of the process that the time interval measures occur at the same location

time dilation

lengthening of the time interval between two events when seen in a moving inertial frame rather than the rest frame of the events (in which the events occur at the same location)

Length Contraction

By the end of this section, you will be able to:

- Explain how simultaneity and length contraction are related.
- Describe the relation between length contraction and time dilation and use it to derive the length-contraction equation.

The length of the train car in [\[link\]](#) is the same for all the passengers. All of them would agree on the simultaneous location of the two ends of the car and obtain the same result for the distance between them. But simultaneous events in one inertial frame need not be simultaneous in another. If the train could travel at relativistic speeds, an observer on the ground would see the simultaneous locations of the two endpoints of the car at a different distance apart than observers inside the car. Measured distances need not be the same for different observers when relativistic speeds are involved.



People might describe distances differently, but at relativistic speeds, the distances really are different.
(credit: “russavia”/Flickr)

Proper Length

Two observers passing each other always see the same value of their relative speed. Even though time dilation implies that the train passenger and the observer standing alongside the tracks measure different times for the train to pass, they still agree that relative speed, which is distance divided by elapsed time, is the same. If an observer on the ground and one on the train measure a different time for the length of the train to pass the ground observer, agreeing on their relative speed means they must also see different distances traveled.

The muon discussed in [\[link\]](#) illustrates this concept ([\[link\]](#)). To an observer on Earth, the muon travels at $0.950c$ for $7.05 \mu\text{s}$ from the time it is produced until it decays. Therefore, it travels a distance relative to Earth of:

Equation:

$$L_0 = v\Delta t = (0.950)(3.00 \times 10^8 \text{ m/s})(7.05 \times 10^{-6} \text{ s}) = 2.01 \text{ km}.$$

In the muon frame, the lifetime of the muon is $2.20 \mu\text{s}$. In this frame of reference, the Earth, air, and ground have only enough time to travel:

Equation:

$$L = v\Delta\tau = (0.950)(3.00 \times 10^8 \text{ m/s})(2.20 \times 10^{-6} \text{ s}) \text{ km} = 0.627 \text{ km}.$$

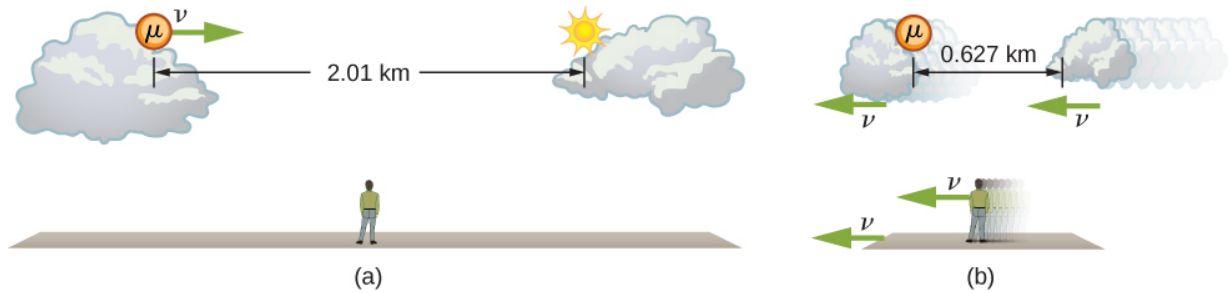
The distance between the same two events (production and decay of a muon) depends on who measures it and how they are moving relative to it.

Note:

Proper Length

Proper length L_0 is the distance between two points measured by an observer who is at rest relative to both of the points.

The earthbound observer measures the proper length L_0 because the points at which the muon is produced and decays are stationary relative to Earth. To the muon, Earth, air, and clouds are moving, so the distance L it sees is not the proper length.



(a) The earthbound observer sees the muon travel 2.01 km. (b) The same path has length 0.627 km seen from the muon's frame of reference. The Earth, air, and clouds are moving relative to the muon in its frame, and have smaller lengths along the direction of travel.

Length Contraction

To relate distances measured by different observers, note that the velocity relative to the earthbound observer in our muon example is given by

Equation:

$$v = \frac{L_0}{\Delta t}.$$

The time relative to the earthbound observer is Δt , because the object being timed is moving relative to this observer. The velocity relative to the moving observer is given by

Equation:

$$v = \frac{L}{\Delta\tau}.$$

The moving observer travels with the muon and therefore observes the proper time $\Delta\tau$. The two velocities are identical; thus,

Equation:

$$\frac{L_0}{\Delta t} = \frac{L}{\Delta\tau}.$$

We know that $\Delta t = \gamma\Delta\tau$. Substituting this equation into the relationship above gives

Equation:

$$L = \frac{L_0}{\gamma}.$$

Substituting for γ gives an equation relating the distances measured by different observers.

Note:

Length Contraction

Length contraction is the decrease in the measured length of an object from its proper length when measured in a reference frame that is moving with respect to the object:

Equation:

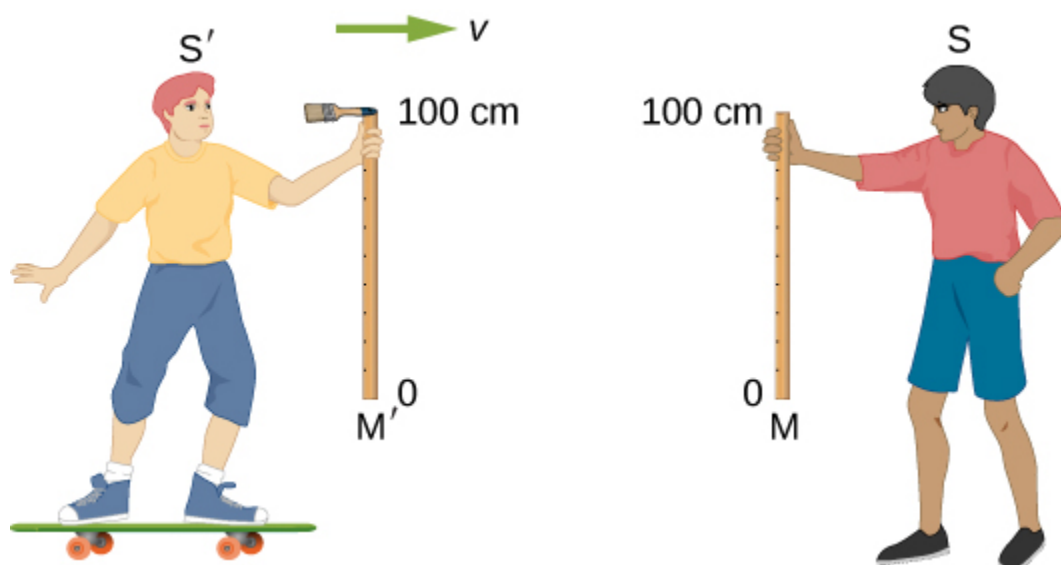
$$L = L_0 \sqrt{1 - \frac{v^2}{c^2}}$$

where L_0 is the length of the object in its rest frame, and L is the length in the frame moving with velocity v .

If we measure the length of anything moving relative to our frame, we find its length L to be smaller than the proper length L_0 that would be measured if the object were stationary. For example, in the muon's rest frame, the distance Earth moves between where the muon was produced and where it decayed is shorter than the distance traveled as seen from the Earth's frame. Those points are fixed relative to Earth but are moving relative to the muon. Clouds and other objects are also contracted along the direction of motion as seen from muon's rest frame.

Thus, two observers measure different distances along their direction of relative motion, depending on which one is measuring distances between objects at rest.

But what about distances measured in a direction perpendicular to the relative motion? Imagine two observers moving along their x -axes and passing each other while holding meter sticks vertically in the y -direction. [\[link\]](#) shows two meter sticks M and M' that are at rest in the reference frames of two boys S and S' , respectively. A small paintbrush is attached to the top (the 100-cm mark) of stick M' . Suppose that S' is moving to the right at a very high speed v relative to S , and the sticks are oriented so that they are perpendicular, or transverse, to their relative velocity vector. The sticks are held so that as they pass each other, their lower ends (the 0-cm marks) coincide. Assume that when S looks at his stick M afterwards, he finds a line painted on it, just below the top of the stick. Because the brush is attached to the top of the other boy's stick M' , S can only conclude that stick M' is less than 1.0 m long.



Meter sticks M and M' are stationary in the reference frames of observers S and S' , respectively. As the sticks pass, a small brush attached to the 100-cm mark of M' paints a line on M .

Now when the boys approach each other, S' , like S , sees a meter stick moving toward him with speed v . Because their situations are symmetric, each boy must make the same measurement of the stick in the other frame. So, if S measures stick M' to be less than 1.0 m long, S' must measure stick M to be also less than 1.0 m long, and S' must see his paintbrush pass over the top of stick M and not paint a line on it. In other words, after the same event, one boy sees a painted line on a stick, while the other does not see such a line on that same stick!

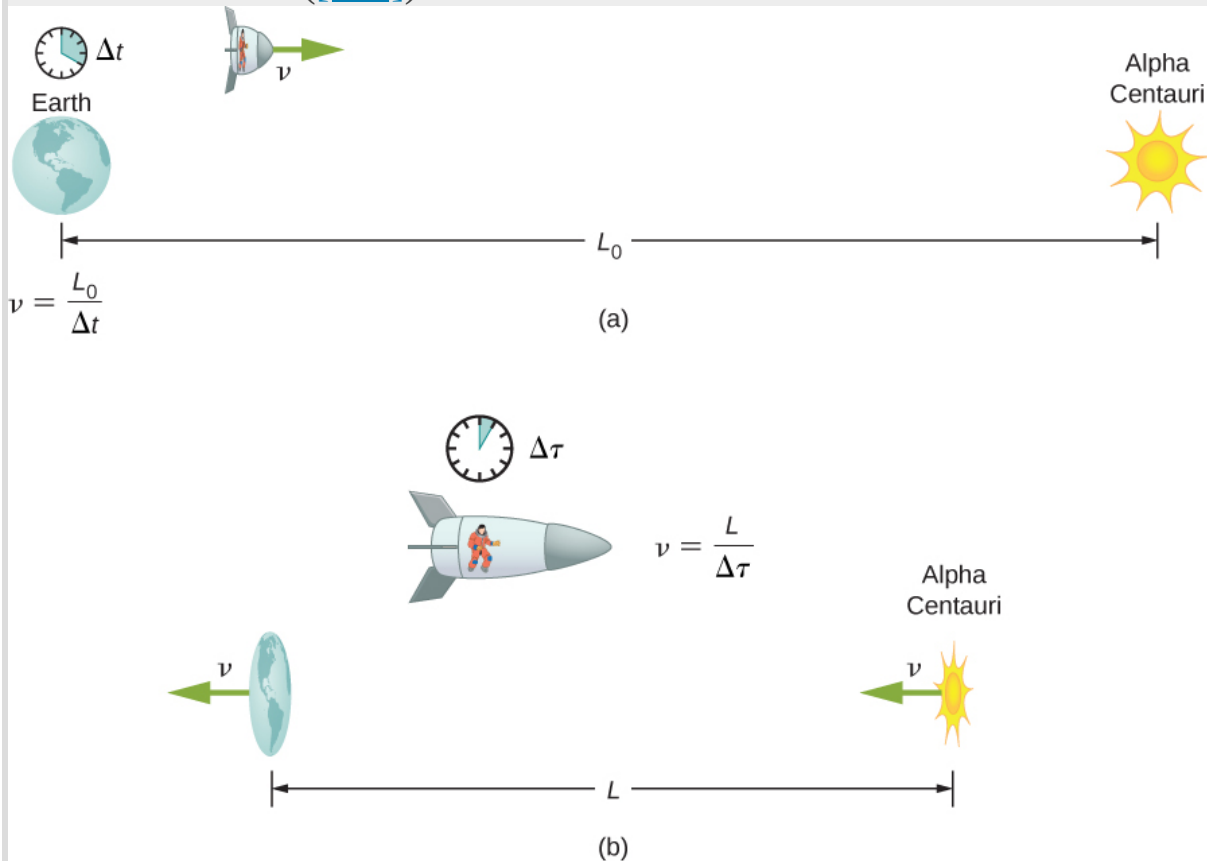
Einstein's first postulate requires that the laws of physics (as, for example, applied to painting) predict that S and S' , who are both in inertial frames, make the same observations; that is, S and S' must either both see a line painted on stick M , or both not see that line. We are therefore forced to conclude our original assumption that S saw a line painted below the top of his stick was wrong! Instead, S finds the line painted right at the 100-cm mark on M . Then both boys will agree that a line is painted on M , and they will also agree that both sticks are exactly 1 m long. We conclude then that

measurements of a transverse *length must be the same in different inertial frames*.

Example:

Calculating Length Contraction

Suppose an astronaut, such as the twin in the twin paradox discussion, travels so fast that $\gamma = 30.00$. (a) The astronaut travels from Earth to the nearest star system, Alpha Centauri, 4.300 light years (ly) away as measured by an earthbound observer. How far apart are Earth and Alpha Centauri as measured by the astronaut? (b) In terms of c , what is the astronaut's velocity relative to Earth? You may neglect the motion of Earth relative to the sun ([\[link\]](#)).



(a) The earthbound observer measures the proper distance between Earth and Alpha Centauri. (b) The astronaut observes a length contraction because Earth and Alpha Centauri move relative to her

ship. She can travel this shorter distance in a smaller time (her proper time) without exceeding the speed of light.

Strategy

First, note that a light year (ly) is a convenient unit of distance on an astronomical scale—it is the distance light travels in a year. For part (a), the 4.300-ly distance between Alpha Centauri and Earth is the proper distance L_0 , because it is measured by an earthbound observer to whom both stars are (approximately) stationary. To the astronaut, Earth and Alpha Centauri are moving past at the same velocity, so the distance between them is the contracted length L . In part (b), we are given γ , so we can find v by rearranging the definition of γ to express v in terms of c .

Solution for (a)

For part (a):

- Identify the knowns: $L_0 = 4.300 \text{ ly}$; $\gamma = 30.00$.
- Identify the unknown: L .
- Express the answer as an equation: $L = \frac{L_0}{\gamma}$.
- Do the calculation:

Equation:

$$\begin{aligned} L &= \frac{L_0}{\gamma} \\ &= \frac{4.300 \text{ ly}}{30.00} \\ &= 0.1433 \text{ ly}. \end{aligned}$$

Solution for (b)

For part (b):

- Identify the known: $\gamma = 30.00$.
- Identify the unknown: v in terms of c .
- Express the answer as an equation. Start with:

Equation:

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Then solve for the unknown v/c by first squaring both sides and then rearranging:

Equation:

$$\gamma^2 = \frac{1}{1 - \frac{v^2}{c^2}}$$

$$\frac{v^2}{c^2} = 1 - \frac{1}{\gamma^2}$$

$$\frac{v}{c} = \sqrt{1 - \frac{1}{\gamma^2}}.$$

d. Do the calculation:

Equation:

$$\begin{aligned} \frac{v}{c} &= \sqrt{1 - \frac{1}{\gamma^2}} \\ &= \sqrt{1 - \frac{1}{(30.00)^2}} \\ &= 0.99944 \end{aligned}$$

or

Equation:

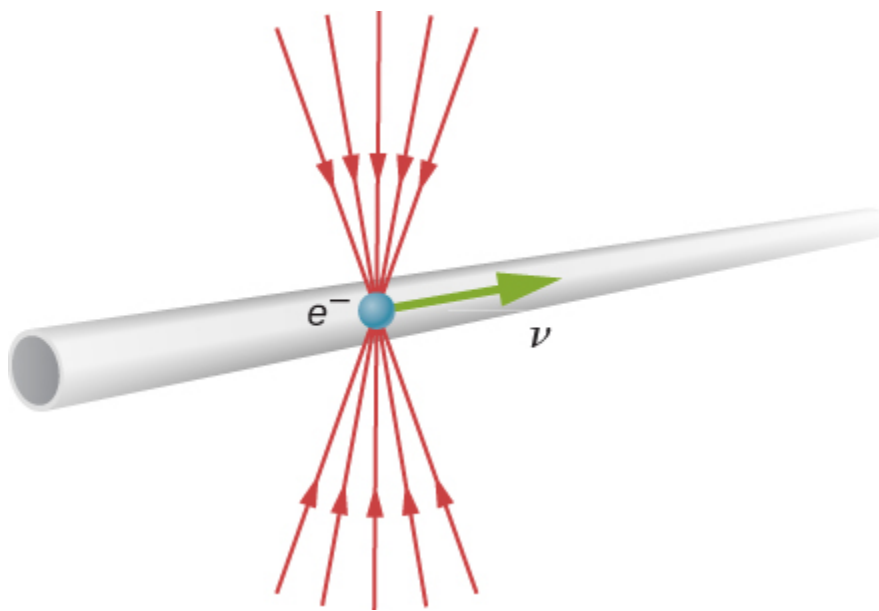
$$v = 0.9994 c.$$

Significance

Remember not to round off calculations until the final answer, or you could get erroneous results. This is especially true for special relativity calculations, where the differences might only be revealed after several decimal places. The relativistic effect is large here ($\gamma = 30.00$), and we see that v is approaching (not equaling) the speed of light. Because the distance as measured by the astronaut is so much smaller, the astronaut can travel it in much less time in her frame.

People traveling at extremely high velocities could cover very large distances (thousands or even millions of light years) and age only a few years on the way. However, like emigrants in past centuries who left their home, these people would leave the Earth they know forever. Even if they returned, thousands to millions of years would have passed on Earth, obliterating most of what now exists. There is also a more serious practical obstacle to traveling at such velocities; immensely greater energies would be needed to achieve such high velocities than classical physics predicts can be attained. This will be discussed later in the chapter.

Why don't we notice length contraction in everyday life? The distance to the grocery store does not seem to depend on whether we are moving or not. Examining the equation $L = L_0 \sqrt{1 - \frac{v^2}{c^2}}$, we see that at low velocities ($v \ll c$), the lengths are nearly equal, which is the classical expectation. But length contraction is real, if not commonly experienced. For example, a charged particle such as an electron traveling at relativistic velocity has electric field lines that are compressed along the direction of motion as seen by a stationary observer ([\[link\]](#)). As the electron passes a detector, such as a coil of wire, its field interacts much more briefly, an effect observed at particle accelerators such as the 3-km-long Stanford Linear Accelerator (SLAC). In fact, to an electron traveling down the beam pipe at SLAC, the accelerator and Earth are all moving by and are length contracted. The relativistic effect is so great that the accelerator is only 0.5 m long to the electron. It is actually easier to get the electron beam down the pipe, because the beam does not have to be as precisely aimed to get down a short pipe as it would to get down a pipe 3 km long. This, again, is an experimental verification of the special theory of relativity.



The electric field lines of a high-velocity charged particle are compressed along the direction of motion by length contraction, producing an observably different signal as the particle goes through a coil.

Note:

Exercise:

Problem:

Check Your Understanding A particle is traveling through Earth's atmosphere at a speed of $0.750c$. To an earthbound observer, the distance it travels is 2.50 km. How far does the particle travel as viewed from the particle's reference frame?

Solution:

$$L = L_0 \sqrt{1 - \frac{v^2}{c^2}} = (2.50 \text{ km}) \sqrt{1 - \frac{(0.750c)^2}{c^2}} = 1.65 \text{ km}$$

Summary

- All observers agree upon relative speed.
- Distance depends on an observer's motion. Proper length L_0 is the distance between two points measured by an observer who is at rest relative to both of the points.
- Length contraction is the decrease in observed length of an object from its proper length L_0 to length L when its length is observed in a reference frame where it is traveling at speed v .
- The proper length is the longest measurement of any length interval. Any observer who is moving relative to the system being observed measures a length shorter than the proper length.

Conceptual Questions

Exercise:

Problem:

To whom does an object seem greater in length, an observer moving with the object or an observer moving relative to the object? Which observer measures the object's proper length?

Solution:

The length of an object is greatest to an observer who is moving with the object, and therefore measures its proper length.

Exercise:

Problem:

Relativistic effects such as time dilation and length contraction are present for cars and airplanes. Why do these effects seem strange to us?

Exercise:

Problem:

Suppose an astronaut is moving relative to Earth at a significant fraction of the speed of light. (a) Does he observe the rate of his clocks to have slowed? (b) What change in the rate of earthbound clocks does he see? (c) Does his ship seem to him to shorten? (d) What about the distance between two stars that lie in the direction of his motion? (e) Do he and an earthbound observer agree on his velocity relative to Earth?

Solution:

a. No, not within the astronaut's own frame of reference. b. He sees Earth clocks to be in their rest frame moving by him, and therefore sees them slowed. c. No, not within the astronaut's own frame of reference. d. Yes, he measures the distance between the two stars to be shorter. e. The two observers agree on their relative speed.

Problems**Exercise:****Problem:**

A spaceship, 200 m long as seen on board, moves by the Earth at $0.970c$. What is its length as measured by an earthbound observer?

Solution:

48.6 m

Exercise:**Problem:**

How fast would a 6.0 m-long sports car have to be going past you in order for it to appear only 5.5 m long?

Exercise:

Problem:

(a) How far does the muon in [\[link\]](#) travel according to the earthbound observer? (b) How far does it travel as viewed by an observer moving with it? Base your calculation on its velocity relative to the Earth and the time it lives (proper time). (c) Verify that these two distances are related through length contraction $\gamma = 3.20$.

Solution:

Using the values given in [\[link\]](#): a. 0.627 km; b. 2.00 km; c. 2.00 km

Exercise:**Problem:**

(a) How long would the muon in [\[link\]](#) have lived as observed on Earth if its velocity was $0.0500c$? (b) How far would it have traveled as observed on Earth? (c) What distance is this in the muon's frame?

Exercise:**Problem:**

Unreasonable Results A spaceship is heading directly toward Earth at a velocity of $0.800c$. The astronaut on board claims that he can send a canister toward the Earth at $1.20c$ relative to Earth. (a) Calculate the velocity the canister must have relative to the spaceship. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

a. $10.0c$; b. The resulting speed of the canister is greater than c , an impossibility. c. It is unreasonable to assume that the canister will move toward the earth at $1.20c$.

Glossary

length contraction

decrease in observed length of an object from its proper length L_0 to length L when its length is observed in a reference frame where it is traveling at speed v

proper length

L_0 ; the distance between two points measured by an observer who is at rest relative to both of the points; for example, earthbound observers measure proper length when measuring the distance between two points that are stationary relative to Earth

The Lorentz Transformation

- Describe the Galilean transformation of classical mechanics, relating the position, time, velocities, and accelerations measured in different inertial frames
- Derive the corresponding Lorentz transformation equations, which, in contrast to the Galilean transformation, are consistent with special relativity
- Explain the Lorentz transformation and many of the features of relativity in terms of four-dimensional space-time

We have used the postulates of relativity to examine, in particular examples, how observers in different frames of reference measure different values for lengths and the time intervals. We can gain further insight into how the postulates of relativity change the Newtonian view of time and space by examining the transformation equations that give the space and time coordinates of events in one inertial reference frame in terms of those in another. We first examine how position and time coordinates transform between inertial frames according to the view in Newtonian physics. Then we examine how this has to be changed to agree with the postulates of relativity. Finally, we examine the resulting Lorentz transformation equations and some of their consequences in terms of four-dimensional space-time diagrams, to support the view that the consequences of special relativity result from the properties of time and space itself, rather than electromagnetism.

The Galilean Transformation Equations

An **event** is specified by its location and time (x, y, z, t) relative to one particular inertial frame of reference S . As an example, (x, y, z, t) could denote the position of a particle at time t , and we could be looking at these positions for many different times to follow the motion of the particle. Suppose a second frame of reference S' moves with velocity v with respect to the first. For simplicity, assume this relative velocity is along the x -axis. The relation between the time and coordinates in the two frames of reference is then

Equation:

$$x = x' + vt', \quad y = y', \quad z = z'.$$

Implicit in these equations is the assumption that time measurements made by observers in both S and S' are the same. That is,

Equation:

$$t = t'.$$

These four equations are known collectively as the **Galilean transformation**.

We can obtain the Galilean velocity and acceleration transformation equations by differentiating these equations with respect to time. We use u for the velocity of a particle throughout this chapter to distinguish it from v , the relative velocity of two reference frames. Note that, for the Galilean transformation, the increment of time used in differentiating to calculate the particle velocity is the same in both frames, $dt = dt'$. Differentiation yields

Equation:

$$u_x = u'_x + v, \quad u_y = u'_y, \quad u_z = u'_z$$

and

Equation:

$$a_x = a'_x, \quad a_y = a'_y, \quad a_z = a'_z.$$

We denote the velocity of the particle by u rather than v to avoid confusion with the velocity v of one frame of reference with respect to the other. Velocities in each frame differ by the velocity that one frame has as seen from the other frame. Observers in both frames of reference measure the same value of the acceleration. Because the mass is unchanged by the transformation, and distances between points are unchanged, observers in both frames see the same forces $F = ma$ acting between objects and the same form of Newton's second and third laws in all inertial frames. The laws of mechanics are consistent with the first postulate of relativity.

The Lorentz Transformation Equations

The Galilean transformation nevertheless violates Einstein's postulates, because the velocity equations state that a pulse of light moving with speed c along the x -axis would travel at speed $c - v$ in the other inertial frame. Specifically, the spherical pulse has radius $r = ct$ at time t in the unprimed frame, and also has radius $r' = ct'$ at time t' in the primed frame. Expressing these relations in Cartesian coordinates gives

Equation:

$$\begin{aligned}x^2 + y^2 + z^2 - c^2 t^2 &= 0 \\x'^2 + y'^2 + z'^2 - c^2 t'^2 &= 0.\end{aligned}$$

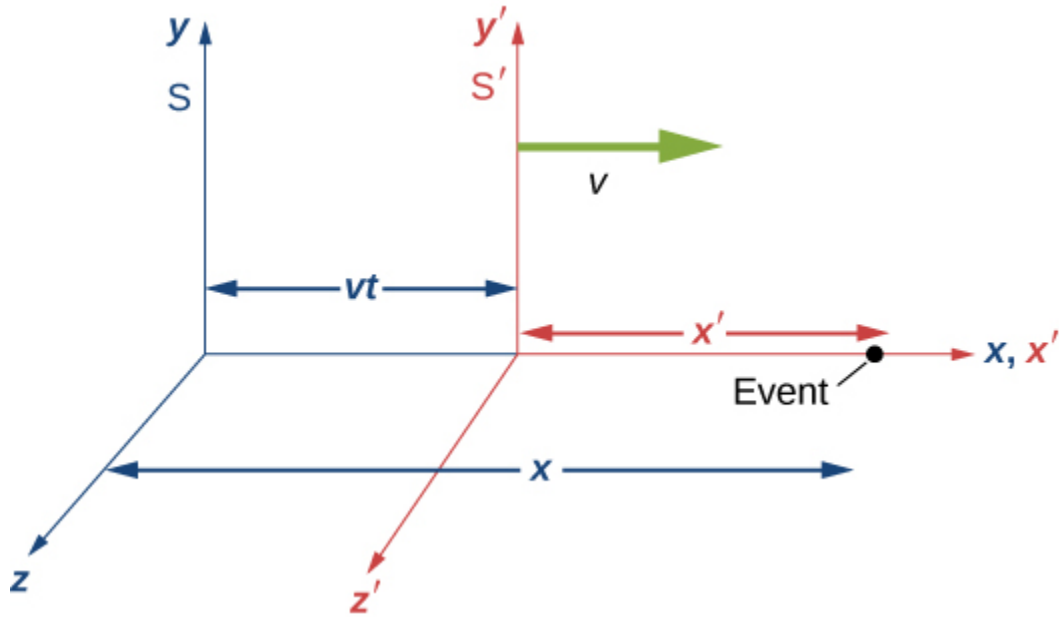
The left-hand sides of the two expressions can be set equal because both are zero. Because $y = y'$ and $z = z'$, we obtain

Equation:

$$x^2 - c^2 t^2 = x'^2 - c^2 t'^2.$$

This cannot be satisfied for nonzero relative velocity v of the two frames if we assume the Galilean transformation results in $t = t'$ with $x = x' + vt'$.

To find the correct set of transformation equations, assume the two coordinate systems S and S' in [\[link\]](#). First suppose that an event occurs at $(x', 0, 0, t')$ in S' and at $(x, 0, 0, t)$ in S , as depicted in the figure.



An event occurs at $(x, 0, 0, t)$ in S and at $(x', 0, 0, t')$ in S' .
The Lorentz transformation equations relate events in the two systems.

Suppose that at the instant that the origins of the coordinate systems in S and S' coincide, a flash bulb emits a spherically spreading pulse of light starting from the origin. At time t , an observer in S finds the origin of S' to be at $x = vt$. With the help of a friend in S' , the S observer also measures the distance from the event to the origin of S' and finds it to be $x'\sqrt{1 - v^2/c^2}$. This follows because we have already shown the postulates of relativity to imply length contraction. Thus the position of the event in S is

Equation:

$$x = vt + x'\sqrt{1 - v^2/c^2}$$

and

Equation:

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}.$$

The postulates of relativity imply that the equation relating distance and time of the spherical wave front:

Equation:

$$x^2 + y^2 + z^2 - c^2 t^2 = 0$$

must apply both in terms of primed and unprimed coordinates, which was shown above to lead to [\[link\]](#):

Equation:

$$x^2 - c^2 t^2 = x'^2 - c^2 t'^2.$$

We combine this with the equation relating x and x' to obtain the relation between t and t' :

Equation:

$$t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}.$$

The equations relating the time and position of the events as seen in S are then

Equation:

$$\begin{aligned} t &= \frac{t' + vx'/c^2}{\sqrt{1 - v^2/c^2}} \\ x &= \frac{x' + vt'}{\sqrt{1 - v^2/c^2}} \\ y &= y' \\ z &= z'. \end{aligned}$$

This set of equations, relating the position and time in the two inertial frames, is known as the **Lorentz transformation**. They are named in honor of H.A. Lorentz (1853–1928), who first proposed them. Interestingly, he justified the transformation on what was eventually discovered to be a fallacious hypothesis. The correct theoretical basis is Einstein’s special theory of relativity.

The reverse transformation expresses the variables in S in terms of those in S' . Simply interchanging the primed and unprimed variables and substituting gives:

Equation:

$$\begin{aligned}t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} \\x' &= \frac{x - vt}{\sqrt{1 - v^2/c^2}} \\y' &= y \\z' &= z.\end{aligned}$$

Example:

Using the Lorentz Transformation for Time

Spacecraft S' is at rest, eventually heading toward Alpha Centauri, when Spacecraft S passes it at relative speed $c/2$. The captain of S' sends a radio signal that lasts 1.2 s according to that ship’s clock. Use the Lorentz transformation to find the time interval of the signal measured by the communications officer of spaceship S .

Solution

- Identify the known: $\Delta t' = t_2' - t_1' = 1.2 \text{ s}$; $\Delta x' = x_2' - x_1' = 0$.
- Identify the unknown: $\Delta t = t_2 - t_1$.
- Express the answer as an equation. The time signal starts as (x', t_1') and stops at (x', t_2') . Note that the x' coordinate of both events is the same because the clock is at rest in S' . Write the first Lorentz transformation equation in terms of $\Delta t = t_2 - t_1$, $\Delta x = x_2 - x_1$, and similarly for the primed coordinates, as:

Equation:

$$\Delta t = \frac{\Delta t' + v\Delta x'/c^2}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Because the position of the clock in S' is fixed, $\Delta x' = 0$, and the time interval Δt becomes:

Equation:

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

d. Do the calculation.

With $\Delta t' = 1.2$ s this gives:

Equation:

$$\Delta t = \frac{1.2 \text{ s}}{\sqrt{1 - \left(\frac{1}{2}\right)^2}} = 1.4 \text{ s}.$$

Note that the Lorentz transformation reproduces the time dilation equation.

Example:

Using the Lorentz Transformation for Length

A surveyor measures a street to be $L = 100$ m long in Earth frame S . Use the Lorentz transformation to obtain an expression for its length measured from a spaceship S' , moving by at speed $0.20c$, assuming the x coordinates of the two frames coincide at time $t = 0$.

Solution

- Identify the known: $L = 100$ m; $v = 0.20c$; $\Delta\tau = 0$.
- Identify the unknown: L' .

- c. Express the answer as an equation. The surveyor in frame S has measured the two ends of the stick simultaneously, and found them at rest at x_2 and x_1 a distance $L = x_2 - x_1 = 100$ m apart. The spaceship crew measures the simultaneous location of the ends of the sticks in their frame. To relate the lengths recorded by observers in S' and S, respectively, write the second of the four Lorentz transformation equations as:

Equation:

$$x_2 - x_1 = \frac{x'_2 + vt}{\sqrt{1 - (v^2/c^2)}} - \frac{x'_1 + vt}{\sqrt{1 - (v^2/c^2)}}$$

$$x_2 - x_1 = \frac{x'_2 - x'_1}{\sqrt{1 - (v^2/c^2)}}.$$

- d. Do the calculation. Because $x_2 - x_1 = 100$ m, the length of the moving stick is equal to:

Equation:

$$L' = x'_2 - x'_1$$

$$= (L) \sqrt{1 - (v^2/c^2)}$$

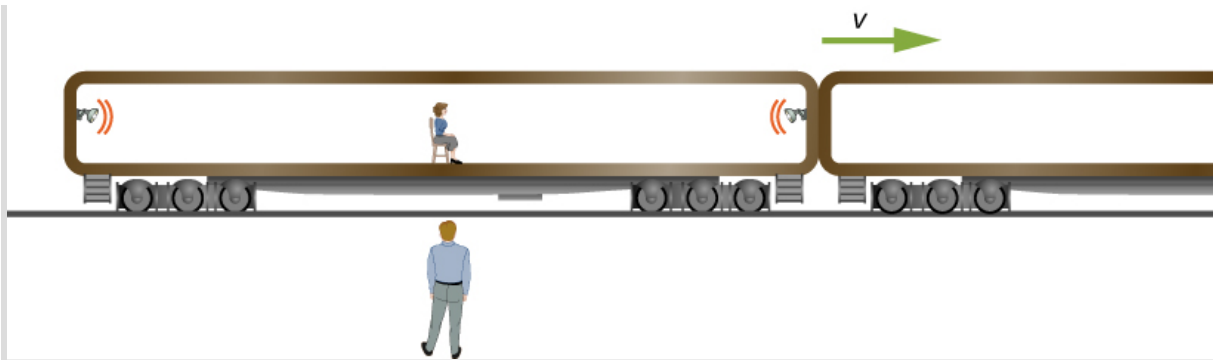
$$= (100 \text{ m}) \sqrt{1 - (0.20)^2}$$

$$L' = 98.0 \text{ m}.$$

Example:

Lorentz Transformation and Simultaneity

The observer shown in [\[link\]](#) standing by the railroad tracks sees the two bulbs flash simultaneously at both ends of the 26 m long passenger car when the middle of the car passes him at a speed of $c/2$. Find the separation in time between when the bulbs flashed as seen by the train passenger seated in the middle of the car.



An person watching a train go by observes two bulbs flash simultaneously at opposite ends of a passenger car. There is another passenger inside of the car observing the same flashes but from a different perspective.

Solution

- a. Identify the known: $\Delta t = 0$.

Note that the spatial separation of the two events is between the two lamps, not the distance of the lamp to the passenger.

- b. Identify the unknown: $\Delta t' = t'_2 - t'_1$.

Again, note that the time interval is between the flashes of the lamps, not between arrival times for reaching the passenger.

- c. Express the answer as an equation:

Equation:

$$\Delta t = \frac{\Delta t' + v \Delta x' / c^2}{\sqrt{1 - v^2 / c^2}}.$$

- d. Do the calculation:

Equation:

$$\begin{aligned} 0 &= \frac{\Delta t' + \frac{c}{2} (26 \text{ m}) / c^2}{\sqrt{1 - v^2 / c^2}} \\ \Delta t' &= -\frac{26 \text{ m/s}}{2c} = -\frac{26 \text{ m/s}}{2(3.00 \times 10^8 \text{ m/s})} \\ \Delta t' &= -4.33 \times 10^{-8} \text{ s.} \end{aligned}$$

Significance

The sign indicates that the event with the larger x_2' , namely, the flash from the right, is seen to occur first in the S' frame, as found earlier for this example, so that $t_2 < t_1$.

Space-time

Relativistic phenomena can be analyzed in terms of events in a four-dimensional space-time. When phenomena such as the twin paradox, time dilation, length contraction, and the dependence of simultaneity on relative motion are viewed in this way, they are seen to be characteristic of the nature of space and time, rather than specific aspects of electromagnetism.

In three-dimensional space, positions are specified by three coordinates on a set of Cartesian axes, and the displacement of one point from another is given by:

Equation:

$$(\Delta x, \Delta y, \Delta z) = (x_2 - x_1, y_2 - y_1, z_2 - z_1).$$

The distance Δr between the points is

Equation:

$$\Delta r^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2.$$

The distance Δr is invariant under a rotation of axes. If a new set of Cartesian axes rotated around the origin relative to the original axes are used, each point in space will have new coordinates in terms of the new axes, but the distance $\Delta r'$ given by

Equation:

$$\Delta r'^2 = (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2.$$

That has the same value that Δr^2 had. Something similar happens with the Lorentz transformation in space-time.

Define the separation between two events, each given by a set of x, y, z , and ct along a four-dimensional Cartesian system of axes in space-time, as

Equation:

$$(\Delta x, \Delta y, \Delta z, c\Delta t) = (x_2 - x_1, y_2 - y_1, z_2 - z_1, c(t_2 - t_1)).$$

Also define the space-time interval Δs between the two events as

Equation:

$$\Delta s^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (c\Delta t)^2.$$

If the two events have the same value of ct in the frame of reference considered, Δs would correspond to the distance Δr between points in space.

The path of a particle through space-time consists of the events (x, y, z, ct) specifying a location at each time of its motion. The path through space-time is called the **world line** of the particle. The world line of a particle that remains at rest at the same location is a straight line that is parallel to the time axis. If the particle moves at constant velocity parallel to the x -axis, its world line would be a sloped line $x = vt$, corresponding to a simple displacement vs. time graph. If the particle accelerates, its world line is curved. The increment of s along the world line of the particle is given in differential form as

Equation:

$$ds^2 = (dx)^2 + (dy)^2 + (dz)^2 - c^2(dt)^2.$$

Just as the distance Δr is invariant under rotation of the space axes, the space-time interval:

Equation:

$$\Delta s^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (c\Delta t)^2.$$

is invariant under the Lorentz transformation. This follows from the postulates of relativity, and can be seen also by substitution of the previous Lorentz transformation equations into the expression for the space-time interval:

Equation:

$$\begin{aligned}\Delta s^2 &= (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (c\Delta t)^2 \\ &= \left(\frac{\Delta x' + v\Delta t'}{\sqrt{1-v^2/c^2}} \right)^2 + (\Delta y')^2 + (\Delta z')^2 - \left(c \frac{\Delta t' + \frac{v\Delta x'}{c^2}}{\sqrt{1-v^2/c^2}} \right)^2 \\ &= (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2 - (c\Delta t')^2 \\ &= \Delta s'^2.\end{aligned}$$

In addition, the Lorentz transformation changes the coordinates of an event in time and space similarly to how a three-dimensional rotation changes old coordinates into new coordinates:

Equation:

Lorentz transformation

$(x, t \text{ coordinates}):$

$$x' = (\gamma)x + (-\beta\gamma)ct$$

$$ct' = (-\beta\gamma)x + (\gamma)ct$$

Axis – rotation around z -axis

$(x, y \text{ coordinates}):$

$$x' = (\cos \theta)x + (\sin \theta)y$$

$$y' = (-\sin \theta)x + (\cos \theta)y$$

where $\gamma = \frac{1}{\sqrt{1-\beta^2}}$; $\beta = v/c$.

Lorentz transformations can be regarded as generalizations of spatial rotations to space-time. However, there are some differences between a three-dimensional axis rotation and a Lorentz transformation involving the time axis, because of differences in how the metric, or rule for measuring the displacements Δr and Δs , differ. Although Δr is invariant under spatial rotations and Δs is invariant also under Lorentz transformation, the

Lorentz transformation involving the time axis does not preserve some features, such as the axes remaining perpendicular or the length scale along each axis remaining the same.

Note that the quantity Δs^2 can have either sign, depending on the coordinates of the space-time events involved. For pairs of events that give it a negative sign, it is useful to define $c^2\Delta\tau^2$ as $-\Delta s^2$. The significance of $c^2\Delta\tau$ as just defined follows by noting that in a frame of reference where the two events occur at the same location, we have $\Delta x = \Delta y = \Delta z = 0$ and therefore (from the equation for $\Delta s^2 = -c^2\Delta\tau^2$):

Equation:

$$c^2\Delta\tau^2 = -\Delta s^2 = (c\Delta t)^2.$$

Therefore $c^2\Delta\tau$ is the time interval $c^2\Delta t$ in the frame of reference where both events occur at the same location. It is the same interval of proper time discussed earlier. It also follows from the relation between Δs and that $c^2\Delta\tau$ that because Δs is Lorentz invariant, the proper time is also Lorentz invariant. All observers in all inertial frames agree on the proper time intervals between the same two events.

Note:

Exercise:

Problem:

Check Your Understanding Show that if a time increment dt elapses for an observer who sees the particle moving with velocity v , it corresponds to a proper time particle increment for the particle of $d\tau = \gamma dt$.

Solution:

Start with the definition of the proper time increment:

$$d\tau = \sqrt{-(ds)^2/c^2} = \sqrt{dt^2 - (dx^2 + dy^2 + dz^2)/c^2}.$$

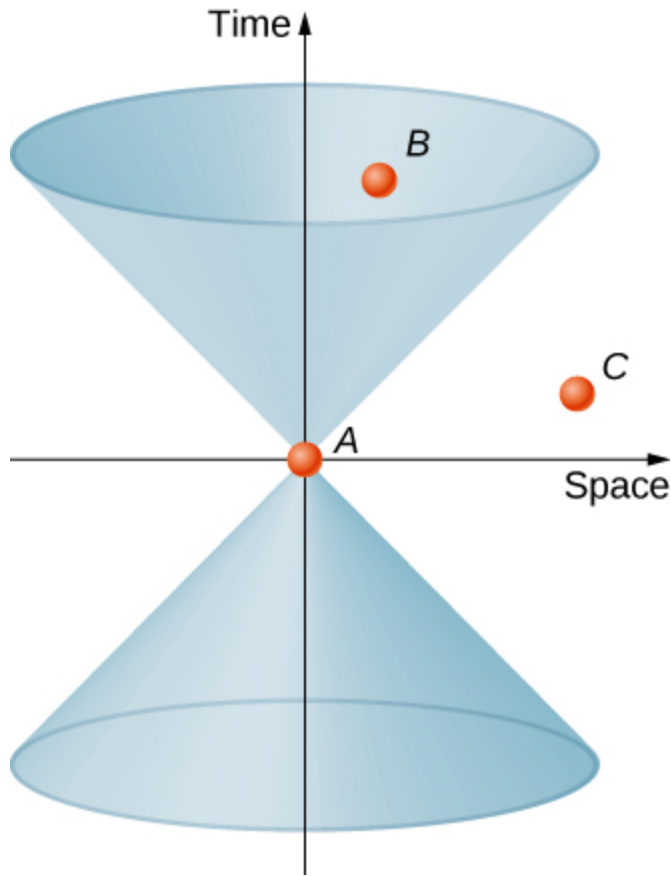
where (dx, dy, dz, cdt) are measured in the inertial frame of an observer who does not necessarily see that particle at rest. This therefore becomes

$$\begin{aligned}
 d\tau &= \sqrt{-(ds)^2/c^2} = \sqrt{dt^2 - [(dx)^2 + (dy)^2 + (dz)^2]/c^2} \\
 &= dt \sqrt{1 - \left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 \right] / c^2} \\
 &= dt \sqrt{1 - v^2/c^2} \\
 dt &= \gamma d\tau.
 \end{aligned}$$

The light cone

We can deal with the difficulty of visualizing and sketching graphs in four dimensions by imagining the three spatial coordinates to be represented collectively by a horizontal axis, and the vertical axis to be the ct -axis. Starting with a particular event in space-time as the origin of the space-time graph shown, the world line of a particle that remains at rest at the initial location of the event at the origin then is the time axis. Any plane through the time axis parallel to the spatial axes contains all the events that are simultaneous with each other and with the intersection of the plane and the time axis, as seen in the rest frame of the event at the origin.

It is useful to picture a light cone on the graph, formed by the world lines of all light beams passing through the origin event A , as shown in [\[link\]](#). The light cone, according to the postulates of relativity, has sides at an angle of 45° if the time axis is measured in units of ct , and, according to the postulates of relativity, the light cone remains the same in all inertial frames. Because the event A is arbitrary, every point in the space-time diagram has a light cone associated with it.



The light cone consists of all the world lines followed by light from the event A at the vertex of the cone.

Consider now the world line of a particle through space-time. Any world line outside of the cone, such as one passing from A through C, would involve speeds greater than c , and would therefore not be possible. Events such as C that lie outside the light cone are said to have a space-like separation from event A. They are characterized in one dimension by:

Equation:

$$\Delta s_{AC}^2 = (x_A - x_C)^2 + (y_A - y_C)^2 + (z_A - z_C)^2 - (c\Delta t)^2 > 0.$$

An event like B that lies in the upper cone is reachable without exceeding the speed of light in vacuum, and is characterized in one dimension by

Equation:

$$\Delta s_{AB}^2 = (x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 - (c\Delta t)^2 < 0.$$

The event is said to have a time-like separation from A . Time-like events that fall into the upper half of the light cone occur at greater values of t than the time of the event A at the vertex and are in the future relative to A .

Events that have time-like separation from A and fall in the lower half of the light cone are in the past, and can affect the event at the origin. The region outside the light cone is labeled as neither past nor future, but rather as “elsewhere.”

For any event that has a space-like separation from the event at the origin, it is possible to choose a time axis that will make the two events occur at the same time, so that the two events are simultaneous in some frame of reference. Therefore, which of the events with space-like separation comes before the other in time also depends on the frame of reference of the observer. Since space-like separations can be traversed only by exceeding the speed of light; this violation of which event can cause the other provides another argument for why particles cannot travel faster than the speed of light, as well as potential material for science fiction about time travel. Similarly for any event with time-like separation from the event at the origin, a frame of reference can be found that will make the events occur at the same location. Because the relations

Equation:

$$\Delta s_{AC}^2 = (x_A - x_C)^2 + (y_A - y_C)^2 + (z_A - z_C)^2 - (c\Delta t)^2 > 0$$

and

Equation:

$$\Delta s_{AB}^2 = (x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 - (c\Delta t)^2 < 0.$$

are Lorentz invariant, whether two events are time-like and can be made to occur at the same place or space-like and can be made to occur at the same time is the same for all observers. All observers in different inertial frames of reference agree on whether two events have a time-like or space-like separation.

The twin paradox seen in space-time

The twin paradox discussed earlier involves an astronaut twin traveling at near light speed to a distant star system, and returning to Earth. Because of time dilation, the space twin is predicted to age much less than the earthbound twin. This seems paradoxical because we might have expected at first glance for the relative motion to be symmetrical and naively thought it possible to also argue that the earthbound twin should age less.

To analyze this in terms of a space-time diagram, assume that the origin of the axes used is fixed in Earth. The world line of the earthbound twin is then along the time axis.

The world line of the astronaut twin, who travels to the distant star and then returns, must deviate from a straight line path in order to allow a return trip. As seen in [\[link\]](#), the circumstances of the two twins are not at all symmetrical. Their paths in space-time are of manifestly different length. Specifically, the world line of the earthbound twin has length $2c\Delta t$, which then gives the proper time that elapses for the earthbound twin as $2\Delta t$. The distance to the distant star system is $\Delta x = v\Delta t$. The proper time that elapses for the space twin is $2\Delta\tau$ where

Equation:

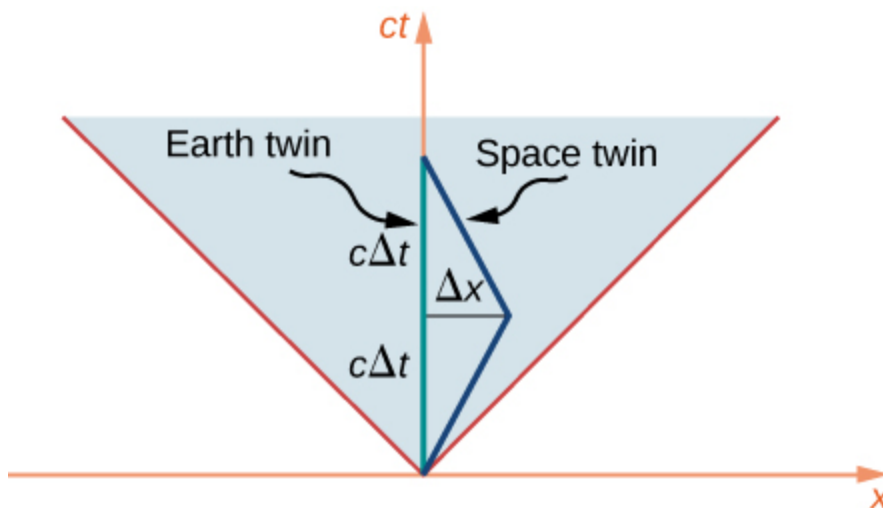
$$c^2\Delta\tau^2 = -\Delta s^2 = (c\Delta t)^2 - (\Delta x)^2.$$

This is considerably shorter than the proper time for the earthbound twin by the ratio

Equation:

$$\begin{aligned}\frac{c\Delta\tau}{c\Delta t} &= \sqrt{\frac{(c\Delta t)^2 - (\Delta x)^2}{(c\Delta t)^2}} = \sqrt{\frac{(c\Delta t)^2 - (v\Delta t)^2}{(c\Delta t)^2}} \\ &= \sqrt{1 - \frac{v^2}{c^2}} = \frac{1}{\gamma}.\end{aligned}$$

consistent with the time dilation formula. The twin paradox is therefore seen to be no paradox at all. The situation of the two twins is not symmetrical in the space-time diagram. The only surprise is perhaps that the seemingly longer path on the space-time diagram corresponds to the smaller proper time interval, because of how $\Delta\tau$ and Δs depend on Δx and Δt .



The space twin and the earthbound twin, in the twin paradox example, follow world lines of different length through space-time.

Lorentz transformations in space-time

We have already noted how the Lorentz transformation leaves

Equation:

$$\Delta s^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (c\Delta t)^2$$

unchanged and corresponds to a rotation of axes in the four-dimensional space-time. If the S and S' frames are in relative motion along their shared x-direction the space and time axes of S' are rotated by an angle α as seen from S, in the way shown in shown in [\[link\]](#), where:

Equation:

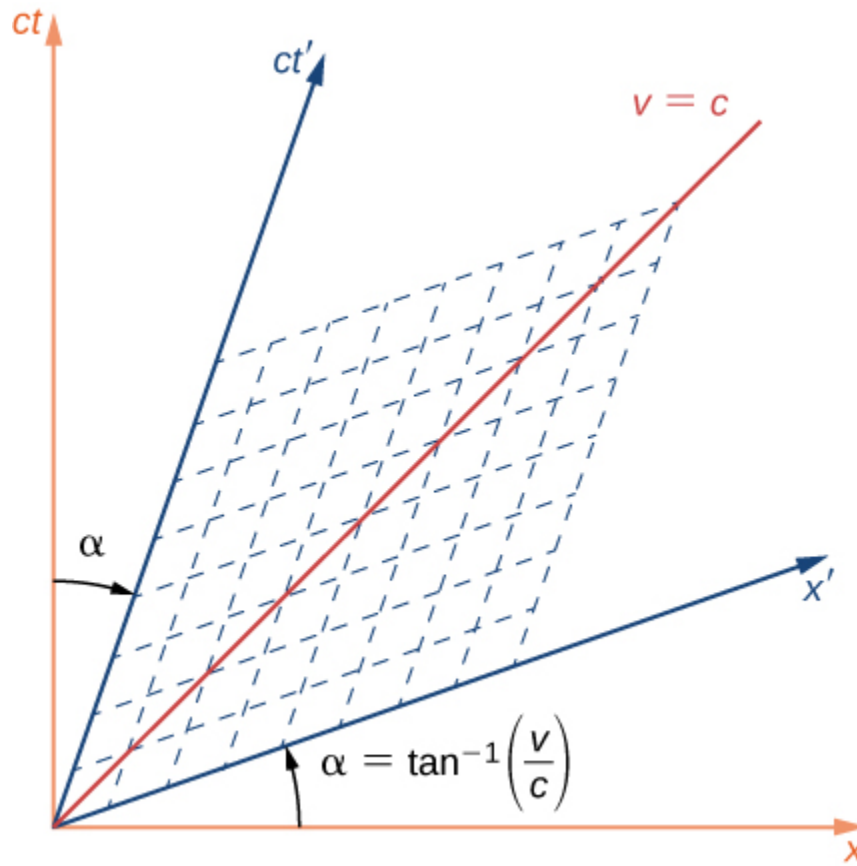
$$\tan\alpha = \frac{v}{c} = \beta.$$

This differs from a rotation in the usual three-dimension sense, insofar as the two space-time axes rotate toward each other symmetrically in a scissors-like way, as shown. The rotation of the time and space axes are both through the same angle. The mesh of dashed lines parallel to the two axes show how coordinates of an event would be read along the primed axes. This would be done by following a line parallel to the x' and one parallel to the t' -axis, as shown by the dashed lines. The length scale of both axes are changed by:

Equation:

$$ct' = ct \sqrt{\frac{1 + \beta^2}{1 - \beta^2}}; \quad x' = x \sqrt{\frac{1 + \beta^2}{1 - \beta^2}}.$$

The line labeled “ $v = c$ ” at 45° to the x-axis corresponds to the edge of the light cone, and is unaffected by the Lorentz transformation, in accordance with the second postulate of relativity. The “ $v = c$ ” line, and the light cone it represents, are the same for both the S and S' frame of reference.

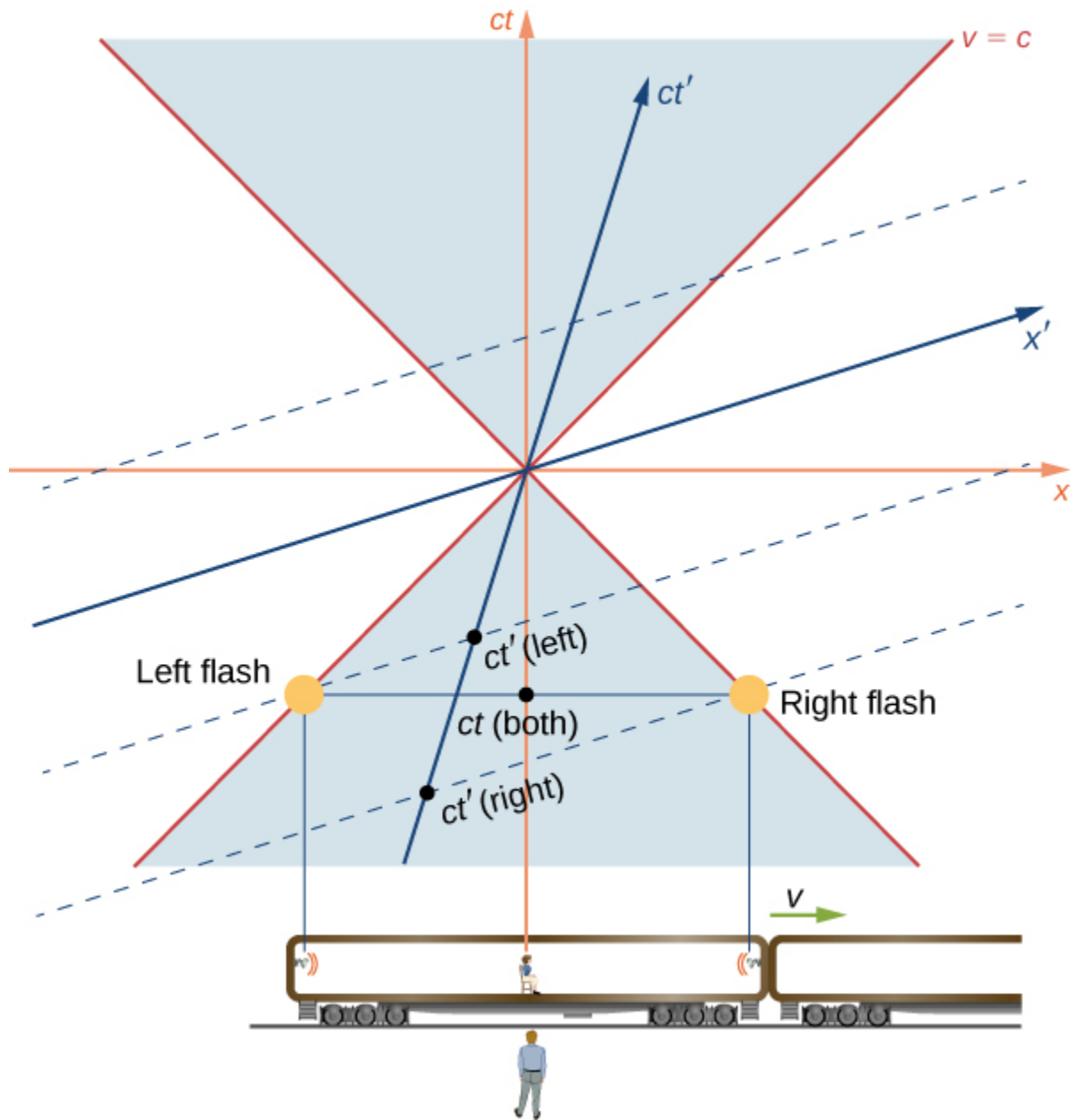


The Lorentz transformation results in new space and time axes rotated in a scissors-like way with respect to the original axes.

Simultaneity

Simultaneity of events at separated locations depends on the frame of reference used to describe them, as given by the scissors-like “rotation” to new time and space coordinates as described. If two events have the same t values in the unprimed frame of reference, they need not have the same values measured along the ct' -axis, and would then not be simultaneous in the primed frame.

As a specific example, consider the near-light-speed train in which flash lamps at the two ends of the car have flashed simultaneously in the frame of reference of an observer on the ground. The space-time graph is shown [\[link\]](#). The flashes of the two lamps are represented by the dots labeled “Left flash lamp” and “Right flash lamp” that lie on the light cone in the past. The world line of both pulses travel along the edge of the light cone to arrive at the observer on the ground simultaneously. Their arrival is the event at the origin. They therefore had to be emitted simultaneously in the unprimed frame, as represented by the point labeled as $t(\text{both})$. But time is measured along the ct' -axis in the frame of reference of the observer seated in the middle of the train car. So in her frame of reference, the emission event of the bulbs labeled as t' (left) and t' (right) were not simultaneous.



The train example revisited. The flashes occur at the same time $t(\text{both})$ along the time axis of the ground observer, but at different times, along the t' time axis of the passenger.

In terms of the space-time diagram, the two observers are merely using different time axes for the same events because they are in different inertial frames, and the conclusions of both observers are equally valid. As the

analysis in terms of the space-time diagrams further suggests, the property of how simultaneity of events depends on the frame of reference results from the properties of space and time itself, rather than from anything specifically about electromagnetism.

Summary

- The Galilean transformation equations describe how, in classical nonrelativistic mechanics, the position, velocity, and accelerations measured in one frame appear in another. Lengths remain unchanged and a single universal time scale is assumed to apply to all inertial frames.
- Newton's laws of mechanics obey the principle of having the same form in all inertial frames under a Galilean transformation, given by **Equation:**

$$x = x' + vt', \quad y = y', \quad z = z', \quad t = t'.$$

The concept that times and distances are the same in all inertial frames in the Galilean transformation, however, is inconsistent with the postulates of special relativity.

- The relativistically correct Lorentz transformation equations are **Equation:**

Lorentz transformation

$$t = \frac{t' + vx'/c^2}{\sqrt{1 - v^2/c^2}}$$

$$x = \frac{x' + vt'}{\sqrt{1 - v^2/c^2}}$$

$$y = y'$$

$$z = z'$$

Inverse Lorentz transformation

$$t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}$$

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}$$

$$y' = y$$

$$z' = z$$

We can obtain these equations by requiring an expanding spherical light signal to have the same shape and speed of growth, c , in both reference frames.

- Relativistic phenomena can be explained in terms of the geometrical properties of four-dimensional space-time, in which Lorentz

- transformations correspond to rotations of axes.
- The Lorentz transformation corresponds to a space-time axis rotation, similar in some ways to a rotation of space axes, but in which the invariant spatial separation is given by Δs rather than distances Δr , and that the Lorentz transformation involving the time axis does not preserve perpendicularity of axes or the scales along the axes.
 - The analysis of relativistic phenomena in terms of space-time diagrams supports the conclusion that these phenomena result from properties of space and time itself, rather than from the laws of electromagnetism.

Problems

Exercise:

Problem:

Describe the following physical occurrences as events, that is, in the form (x, y, z, t) : (a) A postman rings a doorbell of a house precisely at noon. (b) At the same time as the doorbell is rung, a slice of bread pops out of a toaster that is located 10 m from the door in the east direction from the door. (c) Ten seconds later, an airplane arrives at the airport, which is 10 km from the door in the east direction and 2 km to the south.

Exercise:

Problem:

Describe what happens to the angle $\alpha = \tan(v/c)$, and therefore to the transformed axes in [\[link\]](#), as the relative velocity v of the S and S' frames of reference approaches c .

Solution:

The angle α approaches 45° , and the t' - and x' -axes rotate toward the edge of the light cone.

Exercise:

Problem:

Describe the shape of the world line on a space-time diagram of (a) an object that remains at rest at a specific position along the x -axis; (b) an object that moves at constant velocity u in the x -direction; (c) an object that begins at rest and accelerates at a constant rate of in the positive x -direction.

Exercise:**Problem:**

A man standing still at a train station watches two boys throwing a baseball in a moving train. Suppose the train is moving east with a constant speed of 20 m/s and one of the boys throws the ball with a speed of 5 m/s with respect to himself toward the other boy, who is 5 m west from him. What is the velocity of the ball as observed by the man on the station?

Solution:

15 m/s east

Exercise:**Problem:**

When observed from the sun at a particular instant, Earth and Mars appear to move in opposite directions with speeds 108,000 km/h and 86,871 km/h, respectively. What is the speed of Mars at this instant when observed from Earth?

Exercise:**Problem:**

A man is running on a straight road perpendicular to a train track and away from the track at a speed of 12 m/s. The train is moving with a speed of 30 m/s with respect to the track. What is the speed of the man with respect to a passenger sitting at rest in the train?

Solution:

32 m/s

Exercise:**Problem:**

A man is running on a straight road that makes 30° with the train track. The man is running in the direction on the road that is away from the track at a speed of 12 m/s. The train is moving with a speed of 30 m/s with respect to the track. What is the speed of the man with respect to a passenger sitting at rest in the train?

Exercise:**Problem:**

In a frame at rest with respect to the billiard table, a billiard ball of mass m moving with speed v strikes another billiard ball of mass m at rest. The first ball comes to rest after the collision while the second ball takes off with speed v in the original direction of the motion of the first ball. This shows that momentum is conserved in this frame. (a) Now, describe the same collision from the perspective of a frame that is moving with speed v in the direction of the motion of the first ball. (b) Is the momentum conserved in this frame?

Solution:

a. The second ball approaches with velocity $-v$ and comes to rest while the other ball continues with velocity $-v$; b. This conserves momentum.

Exercise:

Problem:

In a frame at rest with respect to the billiard table, two billiard balls of same mass m are moving toward each other with the same speed v . After the collision, the two balls come to rest. (a) Show that momentum is conserved in this frame. (b) Now, describe the same collision from the perspective of a frame that is moving with speed v in the direction of the motion of the first ball. (c) Is the momentum conserved in this frame?

Exercise:**Problem:**

In a frame S , two events are observed: event 1: a pion is created at rest at the origin and event 2: the pion disintegrates after time τ . Another observer in a frame S' is moving in the positive direction along the positive x -axis with a constant speed v and observes the same two events in his frame. The origins of the two frames coincide at $t = t' = 0$. (a) Find the positions and timings of these two events in the frame S' (a) according to the Galilean transformation, and (b) according to the Lorentz transformation.

Solution:

$$\text{a. } \begin{matrix} t_1' = 0; & x_1' = 0; \\ t_2' = \tau; & x_2' = 0 \end{matrix}; \text{ b. } \begin{matrix} t_1' = 0; & x_1' = 0; \\ t_2' = \frac{\tau}{\sqrt{1-v^2/c^2}}; & x_2' = \frac{-v\tau}{\sqrt{1-v^2/c^2}} \end{matrix}$$

Glossary**event**

occurrence in space and time specified by its position and time coordinates (x, y, z, t) measured relative to a frame of reference

Galilean transformation

relation between position and time coordinates of the same events as seen in different reference frames, according to classical mechanics

Lorentz transformation

relation between position and time coordinates of the same events as seen in different reference frames, according to the special theory of relativity

world line

path through space-time

Relativistic Velocity Transformation

By the end of this section, you will be able to:

- Derive the equations consistent with special relativity for transforming velocities in one inertial frame of reference into another.
- Apply the velocity transformation equations to objects moving at relativistic speeds.
- Examine how the combined velocities predicted by the relativistic transformation equations compare with those expected classically.

Remaining in place in a kayak in a fast-moving river takes effort. The river current pulls the kayak along. Trying to paddle against the flow can move the kayak upstream relative to the water, but that only accounts for part of its velocity relative to the shore. The kayak's motion is an example of how velocities in Newtonian mechanics combine by vector addition. The kayak's velocity is the vector sum of its velocity relative to the water and the water's velocity relative to the riverbank. However, the relativistic addition of velocities is quite different.

Velocity Transformations

Imagine a car traveling at night along a straight road, as in [\[link\]](#). The driver sees the light leaving the headlights at speed c within the car's frame of reference. If the Galilean transformation applied to light, then the light from the car's headlights would approach the pedestrian at a speed $u = v + c$, contrary to Einstein's postulates.



According to experimental results and the second postulate of relativity, light from the car's headlights moves away from the car at speed c and toward the observer on the sidewalk at speed c .

Both the distance traveled and the time of travel are different in the two frames of reference, and they must differ in a way that makes the speed of light the same in all inertial frames. The correct rules for transforming velocities from one frame to another can be obtained from the Lorentz transformation equations.

Relativistic Transformation of Velocity

Suppose an object P is moving at constant velocity $\mathbf{u} = (u'_x, u'_y, u'_z)$ as measured in the S' frame. The S' frame is moving along its x' -axis at velocity v . In an increment of time dt' , the particle is displaced by dx' along the x' -axis. Applying the Lorentz transformation equations gives the corresponding increments of time and displacement in the unprimed axes:

Equation:

$$\begin{aligned}
dt &= \gamma (dt' + v dx' / c^2) \\
dx &= \gamma (dx' + v dt') \\
dy &= dy' \\
dz &= dz'.
\end{aligned}$$

The velocity components of the particle seen in the unprimed coordinate system are then

Equation:

$$\begin{aligned}
\frac{dx}{dt} &= \frac{\gamma(dx' + v dt')}{\gamma(dt' + v dx' / c^2)} = \frac{\frac{dx'}{dt'} + v}{1 + \frac{v}{c^2} \frac{dx'}{dt'}} \\
\frac{dy}{dt} &= \frac{dy'}{\gamma(dt' + v dx' / c^2)} = \frac{\frac{dy'}{dt'}}{\gamma \left(1 + \frac{v}{c^2} \frac{dx'}{dt'}\right)} \\
\frac{dz}{dt} &= \frac{dz'}{\gamma(dt' + v dx' / c^2)} = \frac{\frac{dz'}{dt'}}{\gamma \left(1 + \frac{v}{c^2} \frac{dx'}{dt'}\right)}.
\end{aligned}$$

We thus obtain the equations for the velocity components of the object as seen in frame S :

Equation:

$$u_x = \left(\frac{u'_x + v}{1 + v u'_x / c^2} \right), \quad u_y = \left(\frac{u'_y / \gamma}{1 + v u'_x / c^2} \right), \quad u_z = \left(\frac{u'_z / \gamma}{1 + v u'_x / c^2} \right).$$

Compare this with how the Galilean transformation of classical mechanics says the velocities transform, by adding simply as vectors:

Equation:

$$u_x = u'_x + v, \quad u_y = u'_y, \quad u_z = u'_z.$$

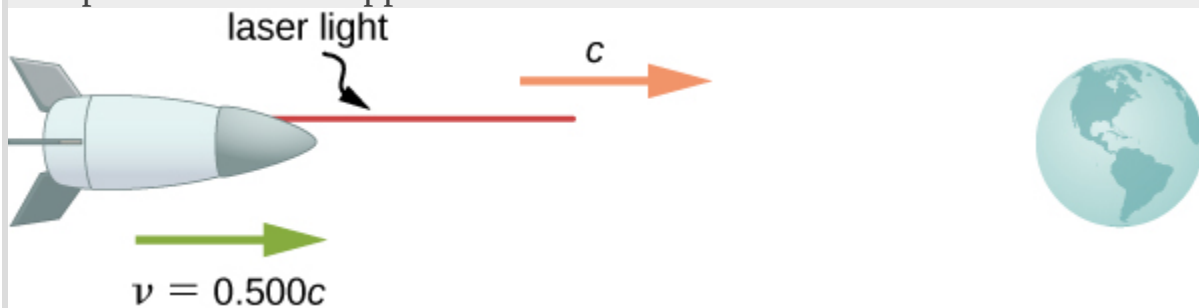
When the relative velocity of the frames is much smaller than the speed of light, that is, when $v \ll c$, the special relativity velocity addition law reduces to the Galilean velocity law. When the speed v of S' relative to S is

comparable to the speed of light, the **relativistic velocity addition** law gives a much smaller result than the **classical (Galilean) velocity addition** does.

Example:

Velocity Transformation Equations for Light

Suppose a spaceship heading directly toward Earth at half the speed of light sends a signal to us on a laser-produced beam of light ([\[link\]](#)). Given that the light leaves the ship at speed c as observed from the ship, calculate the speed at which it approaches Earth.



How fast does a light signal approach Earth if sent from a spaceship traveling at $0.500c$?

Strategy

Because the light and the spaceship are moving at relativistic speeds, we cannot use simple velocity addition. Instead, we determine the speed at which the light approaches Earth using relativistic velocity addition.

Solution

- Identify the knowns: $v = 0.500c$; $u' = c$.
- Identify the unknown: u .
- Express the answer as an equation: $u = \frac{v+u'}{1+\frac{vu'}{c^2}}$.
- Do the calculation:

Equation:

$$\begin{aligned}
 u &= \frac{v+u'}{1+\frac{vu'}{c^2}} \\
 &= \frac{0.500c+c}{1+\frac{(0.500c)(c)}{c^2}} \\
 &= \frac{(0.500+1)c}{\left(\frac{c^2+0.500c^2}{c^2}\right)} \\
 &= c.
 \end{aligned}$$

Significance

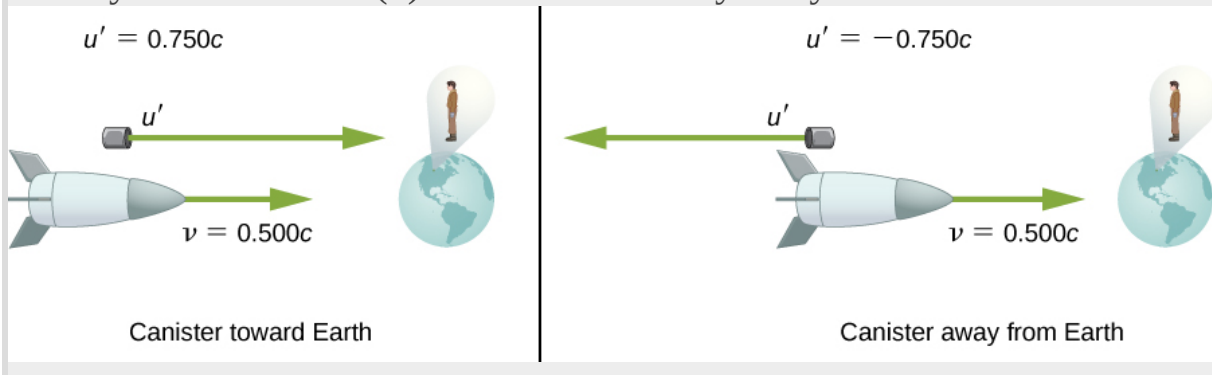
Relativistic velocity addition gives the correct result. Light leaves the ship at speed c and approaches Earth at speed c . The speed of light is independent of the relative motion of source and observer, whether the observer is on the ship or earthbound.

Velocities cannot add to greater than the speed of light, provided that v is less than c and u' does not exceed c . The following example illustrates that relativistic velocity addition is not as symmetric as classical velocity addition.

Example:

Relativistic Package Delivery

Suppose the spaceship in the previous example approaches Earth at half the speed of light and shoots a canister at a speed of $0.750c$ ([link](#)). (a) At what velocity does an earthbound observer see the canister if it is shot directly toward Earth? (b) If it is shot directly away from Earth?



A canister is fired at $0.7500c$ toward Earth or away from Earth.

Strategy

Because the canister and the spaceship are moving at relativistic speeds, we must determine the speed of the canister by an earthbound observer using relativistic velocity addition instead of simple velocity addition.

Solution for (a)

- Identify the knowns: $v = 0.500c$; $u' = 0.750c$.
- Identify the unknown: u .
- Express the answer as an equation: $u = \frac{v+u'}{1+\frac{vu'}{c^2}}$.
- Do the calculation:

Equation:

$$\begin{aligned}u &= \frac{v+u'}{1+\frac{vu'}{c^2}} \\&= \frac{0.500c+0.750c}{1+\frac{(0.500c)(0.750c)}{c^2}} \\&= 0.909c.\end{aligned}$$

Solution for (b)

- Identify the knowns: $v = 0.500c$; $u' = -0.750c$.
- Identify the unknown: u .
- Express the answer as an equation: $u = \frac{v+u'}{1+\frac{vu'}{c^2}}$.
- Do the calculation:

Equation:

$$\begin{aligned}u &= \frac{v+u'}{1+\frac{vu'}{c^2}} \\&= \frac{0.500c+(-0.750c)}{1+\frac{(0.500c)(-0.750c)}{c^2}} \\&= -0.400c.\end{aligned}$$

Significance

The minus sign indicates a velocity away from Earth (in the opposite direction from v), which means the canister is heading toward Earth in part (a) and away in part (b), as expected. But relativistic velocities do not add as simply as they do classically. In part (a), the canister does approach Earth faster, but at less than the vector sum of the velocities, which would give $1.250c$. In part (b), the canister moves away from Earth at a velocity of $-0.400c$, which is *faster* than the $-0.250c$ expected classically. The differences in velocities are not even symmetric: In part (a), an observer on Earth sees the canister and the ship moving apart at a speed of $0.409c$, and at a speed of $0.900c$ in part (b).

Note:

Exercise:

Problem:

Check Your Understanding Distances along a direction perpendicular to the relative motion of the two frames are the same in both frames. Why then are velocities perpendicular to the x -direction different in the two frames?

Solution:

Although displacements perpendicular to the relative motion are the same in both frames of reference, the time interval between events differ, and differences in dt and dt' lead to different velocities seen from the two frames.

Summary

- With classical velocity addition, velocities add like regular numbers in one-dimensional motion: $u = v + u'$, where v is the velocity between

two observers, u is the velocity of an object relative to one observer, and u' is the velocity relative to the other observer.

- Velocities cannot add to be greater than the speed of light.
- Relativistic velocity addition describes the velocities of an object moving at a relativistic velocity.

Problems

Exercise:

Problem:

If two spaceships are heading directly toward each other at $0.800c$, at what speed must a canister be shot from the first ship to approach the other at $0.999c$ as seen by the second ship?

Exercise:

Problem:

Two planets are on a collision course, heading directly toward each other at $0.250c$. A spaceship sent from one planet approaches the second at $0.750c$ as seen by the second planet. What is the velocity of the ship relative to the first planet?

Solution:

$0.615c$

Exercise:

Problem:

When a missile is shot from one spaceship toward another, it leaves the first at $0.950c$ and approaches the other at $0.750c$. What is the relative velocity of the two ships?

Exercise:

Problem:

What is the relative velocity of two spaceships if one fires a missile at the other at $0.750c$ and the other observes it to approach at $0.950c$?

Solution:

$0.696c$

Exercise:**Problem:**

Prove that for any relative velocity v between two observers, a beam of light sent from one to the other will approach at speed c (provided that v is less than c , of course).

Exercise:**Problem:**

Show that for any relative velocity v between two observers, a beam of light projected by one directly away from the other will move away at the speed of light (provided that v is less than c , of course).

Solution:

(Proof)

Glossary

classical (Galilean) velocity addition

method of adding velocities when $v \ll c$; velocities add like regular numbers in one-dimensional motion: $u = v + u'$, where v is the velocity between two observers, u is the velocity of an object relative to one observer, and u' is the velocity relative to the other observer

relativistic velocity addition

method of adding velocities of an object moving at a relativistic speeds

Doppler Effect for Light

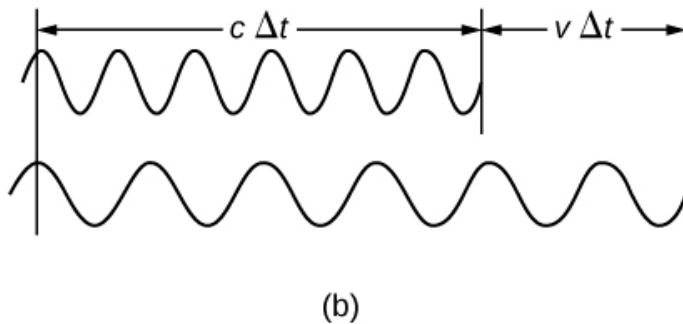
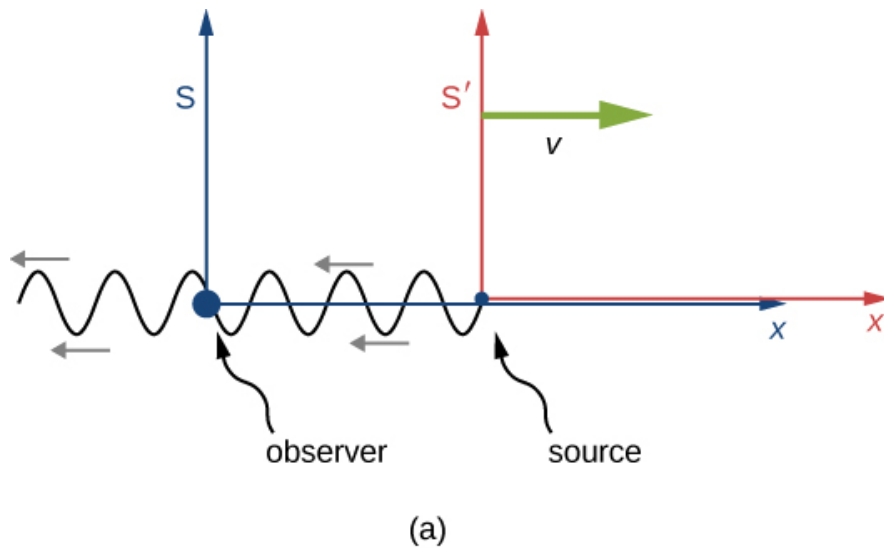
By the end of this section, you will be able to:

- Explain the origin of the shift in frequency and wavelength of the observed wavelength when observer and source moved toward or away from each other
- Derive an expression for the relativistic Doppler shift
- Apply the Doppler shift equations to real-world examples

As discussed in the chapter on sound, if a source of sound and a listener are moving farther apart, the listener encounters fewer cycles of a wave in each second, and therefore lower frequency, than if their separation remains constant. For the same reason, the listener detects a higher frequency if the source and listener are getting closer. The resulting Doppler shift in detected frequency occurs for any form of wave. For sound waves, however, the equations for the Doppler shift differ markedly depending on whether it is the source, the observer, or the air, which is moving. Light requires no medium, and the Doppler shift for light traveling in vacuum depends only on the relative speed of the observer and source.

The Relativistic Doppler Effect

Suppose an observer in S sees light from a source in S' moving away at velocity v ([\[link\]](#)). The wavelength of the light could be measured within S' —for example, by using a mirror to set up standing waves and measuring the distance between nodes. These distances are proper lengths with S' as their rest frame, and change by a factor $\sqrt{1 - v^2/c^2}$ when measured in the observer's frame S , where the ruler measuring the wavelength in S' is seen as moving.



- (a) When a light wave is emitted by a source fixed in the moving inertial frame S' , the observer in S sees the wavelength measured in S' to be shorter by a factor $\sqrt{1 - v^2/c^2}$. (b) Because the observer sees the source moving away within S , the wave pattern reaching the observer in S is also stretched by the factor $(c\Delta t + v\Delta t) / (c\Delta t) = 1 + v/c$.

If the source were stationary in S , the observer would see a length $c\Delta t$ of the wave pattern in time Δt . But because of the motion of S' relative to S , considered solely within S , the observer sees the wave pattern, and therefore the wavelength, stretched out by a factor of **Equation:**

$$\frac{c\Delta t_{\text{period}} + v\Delta t_{\text{period}}}{c\Delta t_{\text{period}}} = 1 + \frac{v}{c}$$

as illustrated in (b) of [\[link\]](#). The overall increase from both effects gives

Equation:

$$\lambda_{\text{obs}} = \lambda_{\text{src}} \left(1 + \frac{v}{c}\right) \sqrt{\frac{1}{1 - \frac{v^2}{c^2}}} = \lambda_{\text{src}} \left(1 + \frac{v}{c}\right) \sqrt{\frac{1}{\left(1 + \frac{v}{c}\right) \left(1 - \frac{v}{c}\right)}} = \lambda_{\text{src}} \sqrt{\frac{\left(1 + \frac{v}{c}\right)}{\left(1 - \frac{v}{c}\right)}}$$

where λ_{src} is the wavelength of the light seen by the source in S' and λ_{obs} is the wavelength that the observer detects within S .

Red Shifts and Blue Shifts

The observed wavelength λ_{obs} of electromagnetic radiation is longer (called a “red shift”) than that emitted by the source when the source moves away from the observer. Similarly, the wavelength is shorter (called a “blue shift”) when the source moves toward the observer. The amount of change is determined by

Equation:

$$\lambda_{\text{obs}} = \lambda_s \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}$$

where λ_s is the wavelength in the frame of reference of the source, and v is the relative velocity of the two frames S and S' . The velocity v is positive for motion away from an observer and negative for motion toward an observer. In terms of source frequency and observed frequency, this equation can be written as

Equation:

$$f_{\text{obs}} = f_s \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}}$$

Notice that the signs are different from those of the wavelength equation.

Example:

Calculating a Doppler Shift

Suppose a galaxy is moving away from Earth at a speed $0.825c$. It emits radio waves with a wavelength of

0.525 m. What wavelength would we detect on Earth?

Strategy

Because the galaxy is moving at a relativistic speed, we must determine the Doppler shift of the radio waves using the relativistic Doppler shift instead of the classical Doppler shift.

Solution

a. Identify the knowns: $u = 0.825c$; $\lambda_s = 0.525 \text{ m}$.

b. Identify the unknown: λ_{obs} .

c. Express the answer as an equation:

Equation:

$$\lambda_{\text{obs}} = \lambda_s \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}.$$

d. Do the calculation:

Equation:

$$\begin{aligned}\lambda_{\text{obs}} &= \lambda_s \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}} \\ &= (0.525 \text{ m}) \sqrt{\frac{1 + \frac{0.825c}{c}}{1 - \frac{0.825c}{c}}} \\ &= 1.70 \text{ m}.\end{aligned}$$

Significance

Because the galaxy is moving away from Earth, we expect the wavelengths of radiation it emits to be redshifted. The wavelength we calculated is 1.70 m, which is redshifted from the original wavelength of 0.525 m. You will see in [Particle Physics and Cosmology](#) that detecting redshifted radiation led to present-day understanding of the origin and evolution of the universe.

Note:

Exercise:

Problem:

Check Your Understanding Suppose a space probe moves away from Earth at a speed $0.350c$. It sends a radio-wave message back to Earth at a frequency of 1.50 GHz. At what frequency is the message received on Earth?

Solution:

We can substitute the data directly into the equation for relativistic Doppler frequency:

$$f_{\text{obs}} = f_s \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}} = (1.50 \text{ GHz}) \sqrt{\frac{1 - \frac{0.350c}{c}}{1 + \frac{0.350c}{c}}} = 1.04 \text{ GHz}.$$

The relativistic Doppler effect has applications ranging from Doppler radar storm monitoring to providing information on the motion and distance of stars. We describe some of these applications in the exercises.

Summary

- An observer of electromagnetic radiation sees relativistic Doppler effects if the source of the radiation is moving relative to the observer. The wavelength of the radiation is longer (called a red shift) than that emitted by the source when the source moves away from the observer and shorter (called a blue shift) when the source moves toward the observer. The shifted wavelength is described by the equation:

Equation:

$$\lambda_{\text{obs}} = \lambda_s \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}.$$

where λ_{obs} is the observed wavelength, λ_s is the source wavelength, and v is the relative velocity of the source to the observer.

Conceptual Questions

Exercise:

Problem:

Explain the meaning of the terms “red shift” and “blue shift” as they relate to the relativistic Doppler effect.

Exercise:

Problem:

What happens to the relativistic Doppler effect when relative velocity is zero? Is this the expected result?

Solution:

There is no measured change in wavelength or frequency in this case. The relativistic Doppler effect depends only on the relative velocity of the source and the observer, not any speed relative to a medium for the light waves.

Exercise:

Problem:

Is the relativistic Doppler effect consistent with the classical Doppler effect in the respect that λ_{obs} is larger for motion away?

Exercise:

Problem:

All galaxies farther away than about 50×10^6 ly exhibit a red shift in their emitted light that is proportional to distance, with those farther and farther away having progressively greater red shifts. What does this imply, assuming that the only source of red shift is relative motion?

Solution:

It shows that the stars are getting more distant from Earth, that the universe is expanding, and doing so at an accelerating rate, with greater velocity for more distant stars.]

Problems

Exercise:

Problem:

A highway patrol officer uses a device that measures the speed of vehicles by bouncing radar off them and measuring the Doppler shift. The outgoing radar has a frequency of 100 GHz and the returning echo has a frequency 15.0 kHz higher. What is the velocity of the vehicle? Note that there are two Doppler shifts in echoes. Be certain not to round off until the end of the problem, because the effect is small.

Relativistic Momentum

By the end of this section, you will be able to:

- Define relativistic momentum in terms of mass and velocity
- Show how relativistic momentum relates to classical momentum
- Show how conservation of relativistic momentum limits objects with mass to speeds less than c

Momentum is a central concept in physics. The broadest form of Newton's second law is stated in terms of momentum. Momentum is conserved whenever the net external force on a system is zero. This makes momentum conservation a fundamental tool for analyzing collisions ([\[link\]](#)). Much of what we know about subatomic structure comes from the analysis of collisions of accelerator-produced relativistic particles, and momentum conservation plays a crucial role in this analysis.



Momentum is an important concept for these football players from the University of California at Berkeley and the University of California at Davis. A player with the same velocity but greater mass collides with greater impact because his momentum is

greater. For objects moving at relativistic speeds, the effect is even greater.

The first postulate of relativity states that the laws of physics are the same in all inertial frames. Does the law of conservation of momentum survive this requirement at high velocities? It can be shown that the momentum calculated as merely $\vec{\mathbf{p}} = m \frac{d\vec{\mathbf{x}}}{dt}$, even if it is conserved in one frame of reference, may not be conserved in another after applying the Lorentz transformation to the velocities. The correct equation for momentum can be shown, instead, to be the classical expression in terms of the increment $d\tau$ of proper time of the particle, observed in the particle's rest frame:

Equation:

$$\begin{aligned}\vec{\mathbf{p}} &= m \frac{d\vec{\mathbf{x}}}{d\tau} = m \frac{d\vec{\mathbf{x}}}{dt} \frac{dt}{d\tau} \\ &= m \frac{d\vec{\mathbf{x}}}{dt} \frac{1}{\sqrt{1-u^2/c^2}} \\ &= \frac{m\vec{\mathbf{u}}}{\sqrt{1-u^2/c^2}} = \gamma m\vec{\mathbf{u}}.\end{aligned}$$

Note:

Relativistic Momentum

Relativistic momentum $\vec{\mathbf{p}}$ is classical momentum multiplied by the relativistic factor γ :

Equation:

$$\vec{\mathbf{p}} = \gamma m\vec{\mathbf{u}}$$

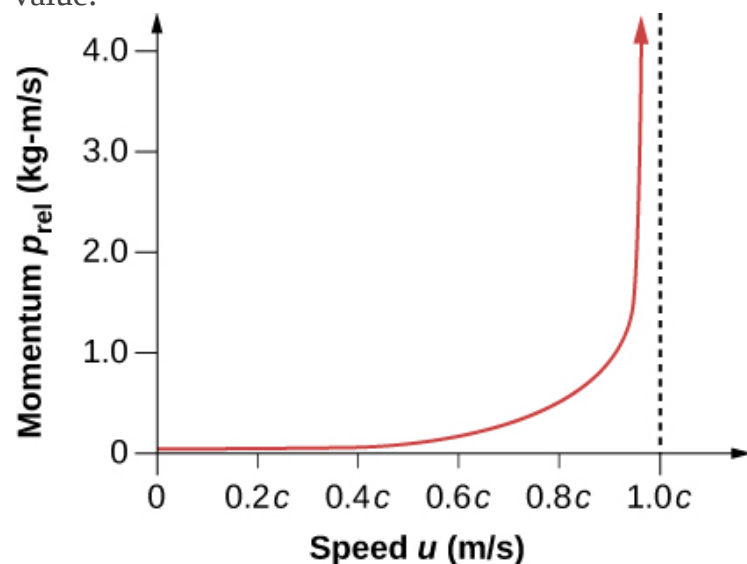
where m is the **rest mass** of the object, $\vec{\mathbf{u}}$ is its velocity relative to an observer, and γ is the relativistic factor:

Equation:

$$\gamma = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}}.$$

Note that we use u for velocity here to distinguish it from relative velocity v between observers. The factor γ that occurs here has the same form as the previous relativistic factor γ except that it is now in terms of the velocity of the particle u instead of the relative velocity v of two frames of reference.

With p expressed in this way, total momentum p_{tot} is conserved whenever the net external force is zero, just as in classical physics. Again we see that the relativistic quantity becomes virtually the same as the classical quantity at low velocities, where u/c is small and γ is very nearly equal to 1. Relativistic momentum has the same intuitive role as classical momentum. It is greatest for large masses moving at high velocities, but because of the factor γ , relativistic momentum approaches infinity as u approaches c ([link](#)). This is another indication that an object with mass cannot reach the speed of light. If it did, its momentum would become infinite—an unreasonable value.



Relativistic momentum approaches infinity as the velocity of an object approaches the speed of light.

The relativistically correct definition of momentum as $p = \gamma m u$ is sometimes taken to imply that mass varies with velocity: $m_{\text{var}} = \gamma m$, particularly in older textbooks. However, note that m is the mass of the object as measured by a person at rest relative to the object. Thus, m is defined to be the rest mass, which could be measured at rest, perhaps using gravity. When a mass is moving relative to an observer, the only way that its mass can be determined is through collisions or other means involving momentum. Because the mass of a moving object cannot be determined independently

of momentum, the only meaningful mass is rest mass. Therefore, when we use the term “mass,” assume it to be identical to “rest mass.”

Relativistic momentum is defined in such a way that conservation of momentum holds in all inertial frames. Whenever the net external force on a system is zero, relativistic momentum is conserved, just as is the case for classical momentum. This has been verified in numerous experiments.

Note:

Exercise:

Problem:

Check Your Understanding What is the momentum of an electron traveling at a speed $0.985c$? The rest mass of the electron is $9.11 \times 10^{-31} \text{ kg}$.

Solution:

Substitute the data into the given equation:

$$p = \gamma mu = \frac{mu}{\sqrt{1 - \frac{u^2}{c^2}}} = \frac{(9.11 \times 10^{-31} \text{ kg})(0.985)(3.00 \times 10^8 \text{ m/s})}{\sqrt{1 - \frac{(0.985c)^2}{c^2}}} = 1.56 \times 10^{-21} \text{ kg}\cdot\text{m/s}.$$

Summary

- The law of conservation of momentum is valid for relativistic momentum whenever the net external force is zero. The relativistic momentum is $p = \gamma mu$, where m is the rest mass of the object, u is its velocity relative to an observer, and the relativistic factor is $\gamma = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}}$.
- At low velocities, relativistic momentum is equivalent to classical momentum.
- Relativistic momentum approaches infinity as u approaches c . This implies that an object with mass cannot reach the speed of light.

Conceptual Questions

Exercise:

Problem:

How does modern relativity modify the law of conservation of momentum?

Exercise:**Problem:**

Is it possible for an external force to be acting on a system and relativistic momentum to be conserved? Explain.

Solution:

Yes. This can happen if the external force is balanced by other externally applied forces, so that the net external force is zero.

Problems**Exercise:****Problem:**

Find the momentum of a helium nucleus having a mass of 6.68×10^{-27} kg that is moving at $0.200c$.

Solution:

$$4.09 \times 10^{-19} \text{ kg} \cdot \text{m/s}$$

Exercise:

Problem: What is the momentum of an electron traveling at $0.980c$?

Exercise:**Problem:**

(a) Find the momentum of a 1.00×10^9 -kg asteroid heading towards Earth at 30.0 km/s. (b) Find the ratio of this momentum to the classical momentum. (Hint: Use the approximation that $\gamma = 1 + (1/2)v^2/c^2$ at low velocities.)

Solution:

a. $3.000000015 \times 10^{13} \text{ kg} \cdot \text{m/s}$; b. 1.000000005

Exercise:**Problem:**

(a) What is the momentum of a 2000-kg satellite orbiting at 4.00 km/s? (b) Find the ratio of this momentum to the classical momentum. (Hint: Use the approximation that $\gamma = 1 + (1/2)v^2/c^2$ at low velocities.)

Exercise:**Problem:**

What is the velocity of an electron that has a momentum of $3.04 \times 10^{-21} \text{ kg} \cdot \text{m/s}$? Note that you must calculate the velocity to at least four digits to see the difference from c .

Solution:

$$2.988 \times 10^8 \text{ m/s}$$

Exercise:**Problem:**

Find the velocity of a proton that has a momentum of $4.48 \times 10^{-19} \text{ kg} \cdot \text{m/s}$.

Glossary

relativistic momentum

\vec{p} , the momentum of an object moving at relativistic velocity; $\vec{p} = \gamma m \vec{u}$

rest mass

mass of an object as measured by an observer at rest relative to the object

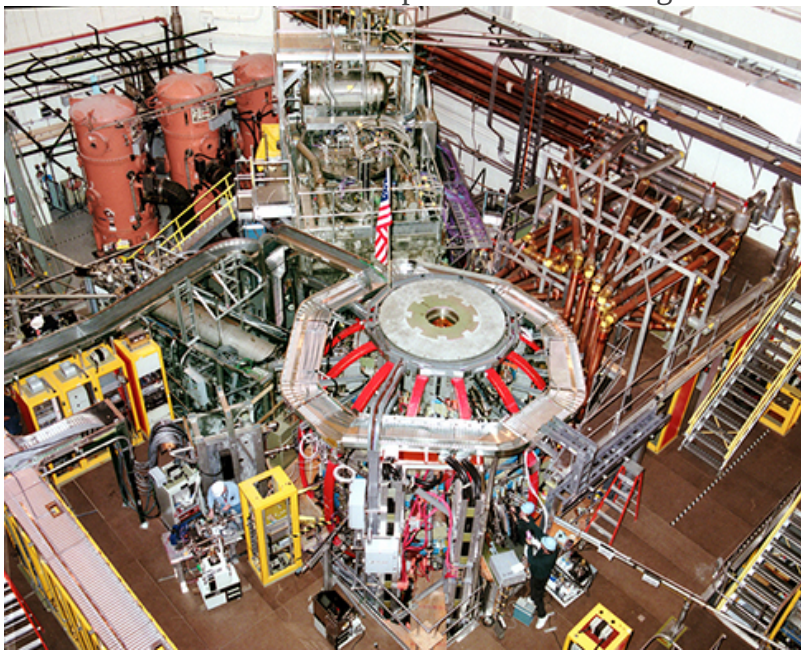
Relativistic Energy

By the end of this section, you will be able to:

- Explain how the work-energy theorem leads to an expression for the relativistic kinetic energy of an object
- Show how the relativistic energy relates to the classical kinetic energy, and sets a limit on the speed of any object with mass
- Describe how the total energy of a particle is related to its mass and velocity
- Explain how relativity relates to energy-mass equivalence, and some of the practical implications of energy-mass equivalence

The tokamak in [\[link\]](#) is a form of experimental fusion reactor, which can change mass to energy. Nuclear reactors are proof of the relationship between energy and matter.

Conservation of energy is one of the most important laws in physics. Not only does energy have many important forms, but each form can be converted to any other. We know that classically, the total amount of energy in a system remains constant. Relativistically, energy is still conserved, but energy-mass equivalence must now be taken into account, for example, in the reactions that occur within a nuclear reactor. Relativistic energy is intentionally defined so that it is conserved in all inertial frames, just as is the case for relativistic momentum. As a consequence, several fundamental quantities are related in ways not known in classical physics. All of these relationships have been verified by experimental results and have fundamental consequences. The altered definition of energy contains some of the most fundamental and spectacular new insights into nature in recent history.



The National Spherical Torus Experiment (NSTX) is a fusion reactor in which hydrogen isotopes undergo fusion to produce helium. In this process, a relatively

small mass of fuel is converted into a large amount of energy. (credit: Princeton Plasma Physics Laboratory)

Kinetic Energy and the Ultimate Speed Limit

The first postulate of relativity states that the laws of physics are the same in all inertial frames. Einstein showed that the law of conservation of energy of a particle is valid relativistically, but for energy expressed in terms of velocity and mass in a way consistent with relativity.

Consider first the relativistic expression for the kinetic energy. We again use u for velocity to distinguish it from relative velocity v between observers. Classically, kinetic energy is related to mass and speed by the familiar expression $K = \frac{1}{2} mu^2$. The corresponding relativistic expression for kinetic energy can be obtained from the work-energy theorem. This theorem states that the net work on a system goes into kinetic energy. Specifically, if a force, expressed as $\vec{F} = \frac{d\vec{p}}{dt} = m \frac{d(\gamma\vec{u})}{dt}$, accelerates a particle from rest to its final velocity, the work done on the particle should be equal to its final kinetic energy. In mathematical form, for one-dimensional motion:

Equation:

$$\begin{aligned} K &= \int F dx = \int m \frac{d}{dt} (\gamma u) dx \\ &= m \int \frac{d(\gamma u)}{dt} \frac{dx}{dt} dt = m \int u \frac{d}{dt} \left(\frac{u}{\sqrt{1 - (u/c)^2}} \right) dt. \end{aligned}$$

Integrate this by parts to obtain

Equation:

$$\begin{aligned}
K &= \left. \frac{mu^2}{\sqrt{1-(u/c)^2}} \right|_0^u - m \int \frac{u}{\sqrt{1-(u/c)^2}} \frac{du}{dt} dt \\
&= \frac{mu^2}{\sqrt{1-(u/c)^2}} - m \int \frac{u}{\sqrt{1-(u/c)^2}} du \\
&= \frac{mu^2}{\sqrt{1-(u/c)^2}} - mc^2 \left(\sqrt{1-(u/c)^2} \right) \Big|_0^u \\
&= \frac{mu^2}{\sqrt{1-(u/c)^2}} + \frac{mc^2}{\sqrt{1-(u/c)^2}} - mc^2 \\
&= mc^2 \left[\frac{(u^2/c^2)+1-(u^2/c^2)}{\sqrt{1-(u/c)^2}} \right] - mc^2 \\
K &= \frac{mc^2}{\sqrt{1-(u/c)^2}} - mc^2.
\end{aligned}$$

Note:

Relativistic Kinetic Energy

Relativistic kinetic energy of any particle of mass m is

Equation:

$$K_{\text{rel}} = (\gamma - 1)mc^2.$$

When an object is motionless, its speed is $u = 0$ and

Equation:

$$\gamma = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} = 1$$

so that $K_{\text{rel}} = 0$ at rest, as expected. But the expression for relativistic kinetic energy (such as total energy and rest energy) does not look much like the classical $\frac{1}{2} mu^2$. To show that the expression for K_{rel} reduces to the classical expression for kinetic energy at low speeds, we use the binomial expansion to obtain an approximation for $(1 + \varepsilon)^n$ valid for small ε :

Equation:

$$(1 + \varepsilon)^n = 1 + n\varepsilon + \frac{n(n-1)}{2!}\varepsilon^2 + \frac{n(n-1)(n-2)}{3!}\varepsilon^3 + \dots \approx 1 + n\varepsilon$$

by neglecting the very small terms in ε^2 and higher powers of ε . Choosing $\varepsilon = -u^2/c^2$ and $n = -\frac{1}{2}$ leads to the conclusion that γ at nonrelativistic speeds, where $\varepsilon = u/c$ is small, satisfies

Equation:

$$\gamma = (1 - u^2/c^2)^{-1/2} \approx 1 + \frac{1}{2} \left(\frac{u^2}{c^2} \right).$$

A binomial expansion is a way of expressing an algebraic quantity as a sum of an infinite series of terms. In some cases, as in the limit of small speed here, most terms are very small. Thus, the expression derived here for γ is not exact, but it is a very accurate approximation. Therefore, at low speed:

Equation:

$$\gamma - 1 = \frac{1}{2} \left(\frac{u^2}{c^2} \right).$$

Entering this into the expression for relativistic kinetic energy gives

Equation:

$$K_{\text{rel}} = \left[\frac{1}{2} \left(\frac{u^2}{c^2} \right) \right] mc^2 = \frac{1}{2} mu^2 = K_{\text{class}}.$$

That is, relativistic kinetic energy becomes the same as classical kinetic energy when $u \ll c$.

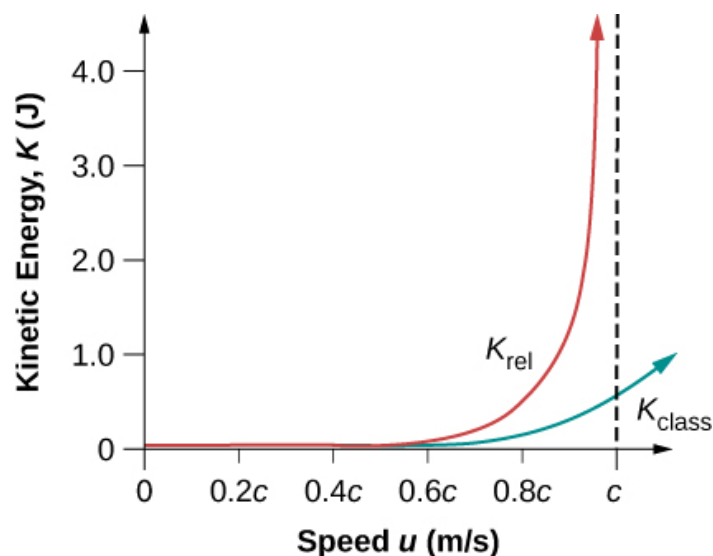
It is even more interesting to investigate what happens to kinetic energy when the speed of an object approaches the speed of light. We know that γ becomes infinite as u approaches c , so that K_{rel} also becomes infinite as the velocity approaches the speed of light ([link](#)). The increase in K_{rel} is far larger than in K_{class} as v approaches c . An infinite amount of work (and, hence, an infinite amount of energy input) is required to accelerate a mass to the speed of light.

Note:

The Speed of Light

No object with mass can attain the **speed of light**.

The speed of light is the ultimate speed limit for any particle having mass. All of this is consistent with the fact that velocities less than c always add to less than c . Both the relativistic form for kinetic energy and the ultimate speed limit being c have been confirmed in detail in numerous experiments. No matter how much energy is put into accelerating a mass, its velocity can only approach—not reach—the speed of light.



This graph of K_{rel} versus velocity shows how kinetic energy increases without bound as velocity approaches the speed of light. Also shown is K_{class} , the classical kinetic energy.

Example:

Comparing Kinetic Energy

An electron has a velocity $v = 0.990c$. (a) Calculate the kinetic energy in MeV of the electron. (b) Compare this with the classical value for kinetic energy at this velocity. (The mass of an electron is $9.11 \times 10^{-31} \text{ kg}$.)

Strategy

The expression for relativistic kinetic energy is always correct, but for (a), it must be used because the velocity is highly relativistic (close to c). First, we calculate the relativistic factor γ , and then use it to determine the relativistic kinetic energy. For (b), we calculate the classical kinetic energy (which would be close to the relativistic value if v were less than a few percent of c) and see that it is not the same.

Solution for (a)

For part (a):

- Identify the knowns: $v = 0.990c$; $m = 9.11 \times 10^{-31}\text{kg}$.
- Identify the unknown: K_{rel} .
- Express the answer as an equation: $K_{\text{rel}} = (\gamma - 1)mc^2$ with $\gamma = \frac{1}{\sqrt{1-u^2/c^2}}$.
- Do the calculation. First calculate γ . Keep extra digits because this is an intermediate calculation:

Equation:

$$\begin{aligned}\gamma &= \frac{1}{\sqrt{1-\frac{u^2}{c^2}}} \\ &= \frac{1}{\sqrt{1-\frac{(0.990c)^2}{c^2}}} \\ &= 7.0888.\end{aligned}$$

Now use this value to calculate the kinetic energy:

Equation:

$$\begin{aligned}K_{\text{rel}} &= (\gamma - 1)mc^2 \\ &= (7.0888 - 1)(9.11 \times 10^{-31} \text{ kg})(3.00 \times 10^8 \text{ m/s}^2) \\ &= 4.9922 \times 10^{-13} \text{ J}.\end{aligned}$$

- Convert units:

Equation:

$$\begin{aligned}K_{\text{rel}} &= (4.9922 \times 10^{-13} \text{ J}) \left(\frac{1 \text{ MeV}}{1.60 \times 10^{-13} \text{ J}} \right) \\ &= 3.12 \text{ MeV}.\end{aligned}$$

Solution for (b)

For part (b):

- List the knowns: $v = 0.990c$; $m = 9.11 \times 10^{-31}\text{kg}$.
- List the unknown: K_{rel} .
- Express the answer as an equation: $K_{\text{class}} = \frac{1}{2} mu^2$.
- Do the calculation:

Equation:

$$\begin{aligned}K_{\text{class}} &= \frac{1}{2} mu^2 \\ &= \frac{1}{2}(9.11 \times 10^{-31} \text{ kg})(0.990)^2(3.00 \times 10^8 \text{ m/s})^2 \\ &= 4.0179 \times 10^{-14} \text{ J}.\end{aligned}$$

- Convert units:

Equation:

$$K_{\text{class}} = 4.0179 \times 10^{-14} \text{ J} \left(\frac{1 \text{ MeV}}{1.60 \times 10^{-13} \text{ J}} \right) \\ = 0.251 \text{ MeV}.$$

Significance

As might be expected, because the velocity is 99.0% of the speed of light, the classical kinetic energy differs significantly from the correct relativistic value. Note also that the classical value is much smaller than the relativistic value. In fact, $K_{\text{rel}}/K_{\text{class}} = 12.4$ in this case. This illustrates how difficult it is to get a mass moving close to the speed of light. Much more energy is needed than predicted classically. Ever-increasing amounts of energy are needed to get the velocity of a mass a little closer to that of light. An energy of 3 MeV is a very small amount for an electron, and it can be achieved with present-day particle accelerators. SLAC, for example, can accelerate electrons to over $50 \times 10^9 \text{ eV} = 50,000 \text{ MeV}$.

Is there any point in getting v a little closer to c than 99.0% or 99.9%? The answer is yes. We learn a great deal by doing this. The energy that goes into a high-velocity mass can be converted into any other form, including into entirely new particles. In the Large Hadron Collider in [\[link\]](#), charged particles are accelerated before entering the ring-like structure. There, two beams of particles are accelerated to their final speed of about 99.7% the speed of light in opposite directions, and made to collide, producing totally new species of particles. Most of what we know about the substructure of matter and the collection of exotic short-lived particles in nature has been learned this way. Patterns in the characteristics of these previously unknown particles hint at a basic substructure for all matter. These particles and some of their characteristics will be discussed in a later chapter on particle physics.



The European Organization for Nuclear Research (called CERN after its French name) operates the largest particle accelerator in the world, straddling the border between France and Switzerland. (credit: modification of work by NASA)

Total Relativistic Energy

The expression for kinetic energy can be rearranged to:

Equation:

$$E = \frac{mc^2}{\sqrt{1 - u^2/c^2}} = K + mc^2.$$

Einstein argued in a separate article, also later published in 1905, that if the energy of a particle changes by ΔE , its mass changes by $\Delta m = \Delta E/c^2$. Abundant experimental evidence since then confirms that mc^2 corresponds to the energy that the particle of mass m has when at rest. For example, when a neutral pion of mass m at rest decays into two photons, the photons have zero mass but are observed to have total energy corresponding to

mc^2 for the pion. Similarly, when a particle of mass m decays into two or more particles with smaller total mass, the observed kinetic energy imparted to the products of the decay corresponds to the decrease in mass. Thus, E is the total relativistic energy of the particle, and mc^2 is its rest energy.

Note:

Total Energy

Total energy E of a particle is

Equation:

$$E = \gamma mc^2$$

where m is mass, c is the speed of light, $\gamma = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}}$, and u is the velocity of the mass relative to an observer.

Note:

Rest Energy

Rest energy of an object is

Equation:

$$E_0 = mc^2.$$

This is the correct form of Einstein's most famous equation, which for the first time showed that energy is related to the mass of an object at rest. For example, if energy is stored in the object, its rest mass increases. This also implies that mass can be destroyed to release energy. The implications of these first two equations regarding relativistic energy are so broad that they were not completely recognized for some years after Einstein published them in 1905, nor was the experimental proof that they are correct widely recognized at first. Einstein, it should be noted, did understand and describe the meanings and implications of his theory.

Example:

Calculating Rest Energy

Calculate the rest energy of a 1.00-g mass.

Strategy

One gram is a small mass—less than one-half the mass of a penny. We can multiply this mass, in SI units, by the speed of light squared to find the equivalent rest energy.

Solution

- a. Identify the knowns: $m = 1.00 \times 10^{-3} \text{ kg}$; $c = 3.00 \times 10^8 \text{ m/s}$.
- b. Identify the unknown: E_0 .
- c. Express the answer as an equation: $E_0 = mc^2$.
- d. Do the calculation:

Equation:

$$\begin{aligned} E_0 &= mc^2 = (1.00 \times 10^{-3} \text{ kg})(3.00 \times 10^8 \text{ m/s})^2 \\ &= 9.00 \times 10^{13} \text{ kg} \cdot \text{m}^2/\text{s}^2. \end{aligned}$$

- e. Convert units. Noting that $1 \text{ kg} \cdot \text{m}^2/\text{s}^2 = 1 \text{ J}$, we see the rest energy is:

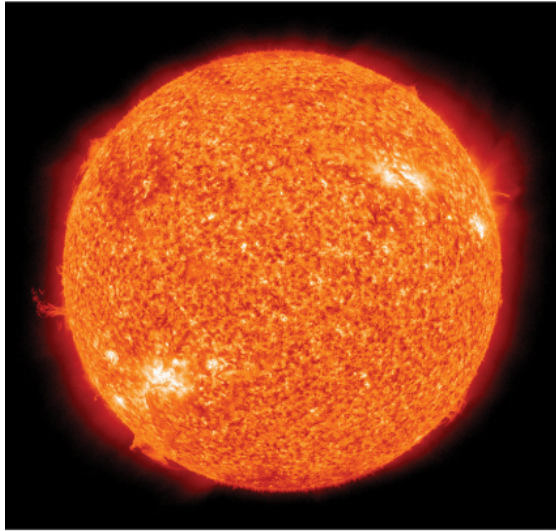
Equation:

$$E_0 = 9.00 \times 10^{13} \text{ J}.$$

Significance

This is an enormous amount of energy for a 1.00-g mass. Rest energy is large because the speed of light c is a large number and c^2 is a very large number, so that mc^2 is huge for any macroscopic mass. The $9.00 \times 10^{13} \text{ J}$ rest mass energy for 1.00 g is about twice the energy released by the Hiroshima atomic bomb and about 10,000 times the kinetic energy of a large aircraft carrier.

Today, the practical applications of *the conversion of mass into another form of energy*, such as in nuclear weapons and nuclear power plants, are well known. But examples also existed when Einstein first proposed the correct form of relativistic energy, and he did describe some of them. Nuclear radiation had been discovered in the previous decade, and it had been a mystery as to where its energy originated. The explanation was that, in some nuclear processes, a small amount of mass is destroyed and energy is released and carried by nuclear radiation. But the amount of mass destroyed is so small that it is difficult to detect that any is missing. Although Einstein proposed this as the source of energy in the radioactive salts then being studied, it was many years before there was broad recognition that mass could be and, in fact, commonly is, converted to energy ([link](#)).



(a)



(b)

(a) The sun and (b) the Susquehanna Steam Electric Station both convert mass into energy—the sun via nuclear fusion, and the electric station via nuclear fission. (credit a: modification of work by NASA/SDO (AIA))

Because of the relationship of rest energy to mass, we now consider mass to be a form of energy rather than something separate. There had not been even a hint of this prior to Einstein's work. Energy-mass equivalence is now known to be the source of the sun's energy, the energy of nuclear decay, and even one of the sources of energy keeping Earth's interior hot.

Stored Energy and Potential Energy

What happens to energy stored in an object at rest, such as the energy put into a battery by charging it, or the energy stored in a toy gun's compressed spring? The energy input becomes part of the total energy of the object and thus increases its rest mass. All stored and potential energy becomes mass in a system. In seeming contradiction, the principle of conservation of mass (meaning total mass is constant) was one of the great laws verified by nineteenth-century science. Why was it not noticed to be incorrect? The following example helps answer this question.

Example:

Calculating Rest Mass

A car battery is rated to be able to move 600 ampere-hours ($A \cdot h$) of charge at 12.0 V. (a) Calculate the increase in rest mass of such a battery when it is taken from being fully

depleted to being fully charged, assuming none of the chemical reactants enter or leave the battery. (b) What percent increase is this, given that the battery's mass is 20.0 kg?

Strategy

In part (a), we first must find the energy stored as chemical energy E_{batt} in the battery, which equals the electrical energy the battery can provide. Because $E_{\text{batt}} = qV$, we have to calculate the charge q in $600 \text{ A} \cdot \text{h}$, which is the product of the current I and the time t . We then multiply the result by 12.0 V . We can then calculate the battery's increase in mass using $E_{\text{batt}} = (\Delta m)c^2$. Part (b) is a simple ratio converted into a percentage.

Solution for (a)

- Identify the knowns: $I \cdot t = 600 \text{ A} \cdot \text{h}$; $V = 12.0 \text{ V}$; $c = 3.00 \times 10^8 \text{ m/s}$.
- Identify the unknown: Δm .
- Express the answer as an equation:

Equation:

$$\begin{aligned} E_{\text{batt}} &= (\Delta m)c^2 \\ \Delta m &= \frac{E_{\text{batt}}}{c^2} \\ &= \frac{qV}{c^2} \\ &= \frac{(It)V}{c^2}. \end{aligned}$$

- Do the calculation:

Equation:

$$\Delta m = \frac{(600 \text{ A} \cdot \text{h})(12.0 \text{ V})}{(3.00 \times 10^8)^2}.$$

Write amperes A as coulombs per second (C/s), and convert hours into seconds:

Equation:

$$\begin{aligned} \Delta m &= \frac{(600 \text{ C/s} \cdot \text{h})\left(\frac{3600 \text{ s}}{1 \text{ h}}\right)(12.0 \text{ J/C})}{(3.00 \times 10^8 \text{ m/s})^2} \\ &= 2.88 \times 10^{-10} \text{ kg}. \end{aligned}$$

where we have used the conversion $1 \text{ kg} \cdot \text{m}^2/\text{s}^2 = 1 \text{ J}$.

Solution for (b)

For part (b):

- Identify the knowns: $\Delta m = 2.88 \times 10^{-10} \text{ kg}$; $m = 20.0 \text{ kg}$.
- Identify the unknown: % change.
- Express the answer as an equation: % increase $= \frac{\Delta m}{m} \times 100 \%$.
- Do the calculation:

Equation:

$$\begin{aligned}
 \% \text{ increase} &= \frac{\Delta m}{m} \times 100 \% \\
 &= \frac{2.88 \times 10^{-10} \text{ kg}}{20.0 \text{ kg}} \times 100 \% \\
 &= 1.44 \times 10^{-9} \%.
 \end{aligned}$$

Significance

Both the actual increase in mass and the percent increase are very small, because energy is divided by c^2 , a very large number. We would have to be able to measure the mass of the battery to a precision of a billionth of a percent, or 1 part in 10^{11} , to notice this increase. It is no wonder that the mass variation is not readily observed. In fact, this change in mass is so small that we may question how anyone could verify that it is real. The answer is found in nuclear processes in which the percentage of mass destroyed is large enough to be measured accurately. The mass of the fuel of a nuclear reactor, for example, is measurably smaller when its energy has been used. In that case, stored energy has been released (converted mostly into thermal energy to power electric generators) and the rest mass has decreased. A decrease in mass also occurs from using the energy stored in a battery, except that the stored energy is much greater in nuclear processes, making the change in mass measurable in practice as well as in theory.

Relativistic Energy and Momentum

We know classically that kinetic energy and momentum are related to each other, because:

Equation:

$$K_{\text{class}} = \frac{p^2}{2m} = \frac{(mu)^2}{2m} = \frac{1}{2} mu^2.$$

Relativistically, we can obtain a relationship between energy and momentum by algebraically manipulating their defining equations. This yields:

Equation:

$$E^2 = (pc)^2 + (mc^2)^2,$$

where E is the relativistic total energy, $E = mc^2 / \sqrt{1 - u^2/c^2}$, and p is the relativistic momentum. This relationship between relativistic energy and relativistic momentum is more complicated than the classical version, but we can gain some interesting new insights by examining it. First, total energy is related to momentum and rest mass. At rest, momentum is zero, and the equation gives the total energy to be the rest energy mc^2 (so this equation is consistent with the discussion of rest energy above). However, as the mass is accelerated, its momentum p increases, thus increasing the total energy. At sufficiently high velocities, the

rest energy term $(mc^2)^2$ becomes negligible compared with the momentum term $(pc)^2$; thus, $E = pc$ at extremely relativistic velocities.

If we consider momentum p to be distinct from mass, we can determine the implications of the equation $E^2 = (pc)^2 + (mc^2)^2$, for a particle that has no mass. If we take m to be zero in this equation, then $E = pc$, or $p = E/c$. Massless particles have this momentum. There are several massless particles found in nature, including photons (which are packets of electromagnetic radiation). Another implication is that a massless particle must travel at speed c and only at speed c . It is beyond the scope of this text to examine the relationship in the equation $E^2 = (pc)^2 + (mc^2)^2$ in detail, but you can see that the relationship has important implications in special relativity.

Note:

Exercise:

Problem:

Check Your Understanding What is the kinetic energy of an electron if its speed is $0.992c$?

Solution:

$$\begin{aligned} K_{\text{rel}} &= (\gamma - 1)mc^2 = \left(\frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} - 1 \right) mc^2 \\ &= \left(\frac{1}{\sqrt{1 - \frac{(0.992c)^2}{c^2}}} - 1 \right) (9.11 \times 10^{-31} \text{ kg})(3.00 \times 10^8 \text{ m/s})^2 = 5.67 \times 10^{-13} \text{ J} \end{aligned}$$

Summary

- The relativistic work-energy theorem is $W_{\text{net}} = E - E_0 = \gamma mc^2 - mc^2 = (\gamma - 1)mc^2$.
- Relativistically, $W_{\text{net}} = K_{\text{rel}}$ where K_{rel} is the relativistic kinetic energy.
- An object of mass m at velocity u has kinetic energy $K_{\text{rel}} = (\gamma - 1)mc^2$, where $\gamma = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}}$.
- At low velocities, relativistic kinetic energy reduces to classical kinetic energy.
- No object with mass can attain the speed of light, because an infinite amount of work and an infinite amount of energy input is required to accelerate a mass to the speed of light.

- Relativistic energy is conserved as long as we define it to include the possibility of mass changing to energy.
- The total energy of a particle with mass m traveling at speed u is defined as $E = \gamma mc^2$, where $\gamma = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}}$ and u denotes the velocity of the particle.
- The rest energy of an object of mass m is $E_0 = mc^2$, meaning that mass is a form of energy. If energy is stored in an object, its mass increases. Mass can be destroyed to release energy.
- We do not ordinarily notice the increase or decrease in mass of an object because the change in mass is so small for a large increase in energy. The equation $E^2 = (pc)^2 + (mc^2)^2$ relates the relativistic total energy E and the relativistic momentum p . At extremely high velocities, the rest energy mc^2 becomes negligible, and $E = pc$.

Key Equations

Time dilation	$\Delta t = \frac{\Delta \tau}{\sqrt{1-\frac{v^2}{c^2}}} = \gamma \tau$
Lorentz factor	$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$
Length contraction	$L = L_0 \sqrt{1 - \frac{v^2}{c^2}} = \frac{L_0}{\gamma}$
Galilean transformation	$x = x' + vt', \quad y = y', \quad z = z', \quad t = t'$
Lorentz transformation	$t = \frac{t' + vx'/c^2}{\sqrt{1-v^2/c^2}}$
	$x = \frac{x' + vt'}{\sqrt{1-v^2/c^2}}$
	$y = y'$
	$z = z'$
Inverse Lorentz transformation	$t' = \frac{t - vx/c^2}{\sqrt{1-v^2/c^2}}$

	$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}$
	$y' = y$
	$z' = z$
Space-time invariants	$(\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - c^2(\Delta t)^2$
	$(\Delta \tau)^2 = -(\Delta s)^2/c^2 = (\Delta t)^2 - \frac{[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2]}{c^2}$
Relativistic velocity addition	$u_x = \left(\frac{u'_x + v}{1 + vu'_x/c^2} \right), \quad u_y = \left(\frac{u'_y/\gamma}{1 + vu'_x/c^2} \right), \quad u_z = \left(\frac{u'_z/\gamma}{1 + vu'_x/c^2} \right)$
Relativistic Doppler effect for wavelength	$\lambda_{\text{obs}} = \lambda_s \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}$
Relativistic Doppler effect for frequency	$f_{\text{obs}} = f_s \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}}$
Relativistic momentum	$\vec{p} = \gamma m \vec{u} = \frac{m \vec{u}}{\sqrt{1 - \frac{u^2}{c^2}}}$
Relativistic total energy	$E = \gamma mc^2$, where $\gamma = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}}$
Relativistic kinetic energy	$K_{\text{rel}} = (\gamma - 1)mc^2$, where $\gamma = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}}$

Conceptual Questions

Exercise:

Problem:

How are the classical laws of conservation of energy and conservation of mass modified by modern relativity?

Exercise:

Problem:

What happens to the mass of water in a pot when it cools, assuming no molecules escape or are added? Is this observable in practice? Explain.

Solution:

Because it loses thermal energy, which is the kinetic energy of the random motion of its constituent particles, its mass decreases by an extremely small amount, as described by energy-mass equivalence.

Exercise:**Problem:**

Consider a thought experiment. You place an expanded balloon of air on weighing scales outside in the early morning. The balloon stays on the scales and you are able to measure changes in its mass. Does the mass of the balloon change as the day progresses? Discuss the difficulties in carrying out this experiment.

Exercise:**Problem:**

The mass of the fuel in a nuclear reactor decreases by an observable amount as it puts out energy. Is the same true for the coal and oxygen combined in a conventional power plant? If so, is this observable in practice for the coal and oxygen? Explain.

Solution:

Yes, in principle there would be a similar effect on mass for any decrease in energy, but the change would be so small for the energy changes in a chemical reaction that it would be undetectable in practice.

Exercise:**Problem:**

We know that the velocity of an object with mass has an upper limit of c . Is there an upper limit on its momentum? Its energy? Explain.

Exercise:

Problem: Given the fact that light travels at c , can it have mass? Explain.

Solution:

Not according to special relativity. Nothing with mass can attain the speed of light.

Exercise:

Problem:

If you use an Earth-based telescope to project a laser beam onto the moon, you can move the spot across the moon's surface at a velocity greater than the speed of light. Does this violate modern relativity? (Note that light is being sent from the Earth to the moon, not across the surface of the moon.)

Problems**Exercise:****Problem:**

What is the rest energy of an electron, given its mass is 9.11×10^{-31} kg? Give your answer in joules and MeV.

Solution:

0.512 MeV according to the number of significant figures stated. The exact value is closer to 0.511 MeV.

Exercise:**Problem:**

Find the rest energy in joules and MeV of a proton, given its mass is 1.67×10^{-27} kg.

Exercise:**Problem:**

If the rest energies of a proton and a neutron (the two constituents of nuclei) are 938.3 and 939.6 MeV, respectively, what is the difference in their mass in kilograms?

Solution:

2.3×10^{-30} kg; to two digits because the difference in rest mass energies is found to two digits

Exercise:**Problem:**

The Big Bang that began the universe is estimated to have released 10^{68} J of energy. How many stars could half this energy create, assuming the average star's mass is 4.00×10^{30} kg?

Exercise:

Problem:

A supernova explosion of a 2.00×10^{31} kg star produces 1.00×10^{44} J of energy.

(a) How many kilograms of mass are converted to energy in the explosion? (b) What is the ratio $\Delta m/m$ of mass destroyed to the original mass of the star?

Solution:

a. 1.11×10^{27} kg; b. 5.56×10^{-5}

Exercise:**Problem:**

(a) Using data from [Potential Energy of a System](#), calculate the mass converted to energy by the fission of 1.00 kg of uranium. (b) What is the ratio of mass destroyed to the original mass, $\Delta m/m$?

Exercise:**Problem:**

(a) Using data from [Potential Energy of a System](#), calculate the amount of mass converted to energy by the fusion of 1.00 kg of hydrogen. (b) What is the ratio of mass destroyed to the original mass, $\Delta m/m$? (c) How does this compare with $\Delta m/m$ for the fission of 1.00 kg of uranium?

Solution:

a. 7.1×10^{-3} kg; b. $7.1 \times 10^{-3} = 7.1 \times 10^{-3}$; c. $\frac{\Delta m}{m}$ is greater for hydrogen

Exercise:**Problem:**

There is approximately 10^{34} J of energy available from fusion of hydrogen in the world's oceans. (a) If 10^{33} J of this energy were utilized, what would be the decrease in mass of the oceans (ignoring the loss of mass from the leftover oxygen)? (b) How great a volume of water does this correspond to? (c) Comment on whether this is a significant fraction of the total mass of the oceans.

Exercise:**Problem:**

A muon has a rest mass energy of 105.7 MeV, and it decays into an electron and a massless particle. (a) If all the lost mass is converted into the electron's kinetic energy, find γ for the electron. (b) What is the electron's velocity?

Solution:

a. 208; b. 0.999988c; six digits used to show difference from c

Exercise:**Problem:**

A π -meson is a particle that decays into a muon and a massless particle. The π -meson has a rest mass energy of 139.6 MeV, and the muon has a rest mass energy of 105.7 MeV. Suppose the π -meson is at rest and all of the missing mass goes into the muon's kinetic energy. How fast will the muon move?

Exercise:**Problem:**

(a) Calculate the relativistic kinetic energy of a 1000-kg car moving at 30.0 m/s if the speed of light were only 45.0 m/s. (b) Find the ratio of the relativistic kinetic energy to classical.

Solution:

a. 6.92×10^5 J; b. 1.54

Exercise:**Problem:**

Alpha decay is nuclear decay in which a helium nucleus is emitted. If the helium nucleus has a mass of 6.80×10^{-27} kg and is given 5.00 MeV of kinetic energy, what is its velocity?

Exercise:**Problem:**

(a) Beta decay is nuclear decay in which an electron is emitted. If the electron is given 0.750 MeV of kinetic energy, what is its velocity? (b) Comment on how the high velocity is consistent with the kinetic energy as it compares to the rest mass energy of the electron.

Solution:

a. 0.914c; b. The rest mass energy of an electron is 0.511 MeV, so the kinetic energy is approximately 150% of the rest mass energy. The electron should be traveling close to the speed of light.

Additional Problems

Exercise:

Problem:

(a) At what relative velocity is $\gamma = 1.50$? (b) At what relative velocity is $\gamma = 100$?

Exercise:

Problem:

(a) At what relative velocity is $\gamma = 2.00$? (b) At what relative velocity is $\gamma = 10.0$?

Solution:

a. $0.866c$; b. $0.995c$

Exercise:

Problem:

Unreasonable Results (a) Find the value of γ required for the following situation. An earthbound observer measures 23.9 h to have passed while signals from a high-velocity space probe indicate that 24.0 h have passed on board. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem:

(a) How long does it take the astronaut in [\[link\]](#) to travel 4.30 ly at $0.99944c$ (as measured by the earthbound observer)? (b) How long does it take according to the astronaut? (c) Verify that these two times are related through time dilation with $\gamma = 30.00$ as given.

Solution:

a. 4.303 y to four digits to show any effect; b. 0.1434 y; c. $1/\sqrt{(1 - v^2/c^2)} = 29.88$.

Exercise:

Problem:

(a) How fast would an athlete need to be running for a 100-m race to look 100 yd long? (b) Is the answer consistent with the fact that relativistic effects are difficult to observe in ordinary circumstances? Explain.

Exercise:

Problem:

(a) Find the value of γ for the following situation. An astronaut measures the length of his spaceship to be 100 m, while an earthbound observer measures it to be 25.0 m. (b) What is the speed of the spaceship relative to Earth?

Solution:

a. 4.00; b. $v = 0.867c$

Exercise:**Problem:**

A clock in a spaceship runs one-tenth the rate at which an identical clock on Earth runs. What is the speed of the spaceship?

Exercise:**Problem:**

An astronaut has a heartbeat rate of 66 beats per minute as measured during his physical exam on Earth. The heartbeat rate of the astronaut is measured when he is in a spaceship traveling at $0.5c$ with respect to Earth by an observer (A) in the ship and by an observer (B) on Earth. (a) Describe an experimental method by which observer B on Earth will be able to determine the heartbeat rate of the astronaut when the astronaut is in the spaceship. (b) What will be the heartbeat rate(s) of the astronaut reported by observers A and B?

Solution:

a. A sends a radio pulse at each heartbeat to B, who knows their relative velocity and uses the time dilation formula to calculate the proper time interval between heartbeats from the observed signal. b. $(66 \text{ beats/min})\sqrt{1 - v^2/c^2} = 57.1 \text{ beats/min}$

Exercise:**Problem:**

A spaceship (A) is moving at speed $c/2$ with respect to another spaceship (B). Observers in A and B set their clocks so that the event at (x, y, z, t) of turning on a laser in spaceship B has coordinates $(0, 0, 0, 0)$ in A and also $(0, 0, 0, 0)$ in B. An observer at the origin of B turns on the laser at $t = 0$ and turns it off at $t = \tau$ in his time. What is the time duration between on and off as seen by an observer in A?

Exercise:

Problem:

Same two observers as in the preceding exercise, but now we look at two events occurring in spaceship A. A photon arrives at the origin of A at its time $t = 0$ and another photon arrives at $(x = 1.00 \text{ m}, 0, 0)$ at $t = 0$ in the frame of ship A. (a) Find the coordinates and times of the two events as seen by an observer in frame B. (b) In which frame are the two events simultaneous and in which frame are they are not simultaneous?

Solution:

a. first photon: $(0, 0, 0)$ at $t = t'$; second photon:

$$t' = \frac{-vx/c^2}{\sqrt{1-v^2/c^2}} = \frac{-(c/2)(1.00 \text{ m})/c^2}{\sqrt{0.75}} = -\frac{0.577 \text{ m}}{c} = 1.93 \times 10^{-9} \text{ s}$$

$$x' = \frac{x}{\sqrt{1-v^2/c^2}} = \frac{1.00 \text{ m}}{\sqrt{0.75}} = 1.15 \text{ m}$$

b. simultaneous in A, not simultaneous in B

Exercise:**Problem:**

Same two observers as in the preceding exercises. A rod of length 1 m is laid out on the x-axis in the frame of B from origin to $(x = 1.00 \text{ m}, 0, 0)$. What is the length of the rod observed by an observer in the frame of spaceship A?

Exercise:**Problem:**

An observer at origin of inertial frame S sees a flashbulb go off at $x = 150 \text{ km}$, $y = 15.0 \text{ km}$, and $z = 1.00 \text{ km}$ at time $t = 4.5 \times 10^{-4} \text{ s}$. At what time and position in the S' system did the flash occur, if S' is moving along shared x-direction with S at a velocity $v = 0.6c$?

Solution:

$$\begin{aligned} t' &= \frac{t - vx/c^2}{\sqrt{1-v^2/c^2}} = \frac{(4.5 \times 10^{-4} \text{ s}) - (0.6c)\left(\frac{150 \times 10^3 \text{ m}}{c^2}\right)}{\sqrt{1-(0.6)^2}} \\ &= 1.88 \times 10^{-4} \text{ s} \\ x' &= \frac{x - vt}{\sqrt{1-v^2/c^2}} = \frac{150 \times 10^3 \text{ m} - (0.60)(3.00 \times 10^8 \text{ m/s})(4.5 \times 10^{-4} \text{ s})}{\sqrt{1-(0.6)^2}} \\ &= 8.6 \times 10^4 \text{ m} = 86 \text{ km} \\ y &= y' = 15 \text{ km} \\ z &= z' = 1 \text{ km} \end{aligned}$$

Exercise:**Problem:**

An observer sees two events 1.5×10^{-8} s apart at a separation of 800 m. How fast must a second observer be moving relative to the first to see the two events occur simultaneously?

Exercise:**Problem:**

An observer standing by the railroad tracks sees two bolts of lightning strike the ends of a 500-m-long train simultaneously at the instant the middle of the train passes him at 50 m/s. Use the Lorentz transformation to find the time between the lightning strikes as measured by a passenger seated in the middle of the train.

Solution:

$$\Delta t = \frac{\Delta t' + v\Delta x'/c^2}{\sqrt{1-v^2/c^2}}$$

$$0 = \frac{\Delta t' + v(500 \text{ m})/c^2}{\sqrt{1-v^2/c^2}};$$

since $v \ll c$, we can ignore the term v^2/c^2 and find

$$\Delta t' = -\frac{(50 \text{ m/s})(500 \text{ m})}{(3.00 \times 10^8 \text{ m/s})^2} = -2.78 \times 10^{-13} \text{ s}$$

The breakdown of Newtonian simultaneity is negligibly small, but not exactly zero, at realistic train speeds of 50 m/s.

Exercise:**Problem:**

Two astronomical events are observed from Earth to occur at a time of 1 s apart and a distance separation of 1.5×10^9 m from each other. (a) Determine whether separation of the two events is space like or time like. (b) State what this implies about whether it is consistent with special relativity for one event to have caused the other?

Exercise:**Problem:**

Two astronomical events are observed from Earth to occur at a time of 0.30 s apart and a distance separation of 2.0×10^9 m from each other. How fast must a spacecraft travel from the site of one event toward the other to make the events occur at the same time when measured in the frame of reference of the spacecraft?

Solution:

$$\begin{aligned}\Delta t' &= \frac{\Delta t - v\Delta x/c^2}{\sqrt{1-v^2/c^2}} \\ 0 &= \frac{(0.30 \text{ s}) - \frac{(v)(2.0 \times 10^9 \text{ m})}{(3.00 \times 10^8 \text{ m/s})^2}}{\sqrt{1-v^2/c^2}} \\ v &= \frac{(0.30 \text{ s})}{(2.0 \times 10^9 \text{ m})} (3.00 \times 10^8 \text{ m/s})^2 \\ v &= 1.35 \times 10^7 \text{ m/s}\end{aligned}$$

Exercise:

Problem:

A spacecraft starts from being at rest at the origin and accelerates at a constant rate g , as seen from Earth, taken to be an inertial frame, until it reaches a speed of $c/2$. (a) Show that the increment of proper time is related to the elapsed time in Earth's frame by:

Equation:

$$d\tau = \sqrt{1 - v^2/c^2} dt.$$

- (b) Find an expression for the elapsed time to reach speed $c/2$ as seen in Earth's frame. (c) Use the relationship in (a) to obtain a similar expression for the elapsed proper time to reach $c/2$ as seen in the spacecraft, and determine the ratio of the time seen from Earth with that on the spacecraft to reach the final speed.

Exercise:

Problem:

(a) All but the closest galaxies are receding from our own Milky Way Galaxy. If a galaxy 12.0×10^9 ly away is receding from us at $0.900c$, at what velocity relative to us must we send an exploratory probe to approach the other galaxy at $0.990c$ as measured from that galaxy? (b) How long will it take the probe to reach the other galaxy as measured from Earth? You may assume that the velocity of the other galaxy remains constant. (c) How long will it then take for a radio signal to be beamed back? (All of this is possible in principle, but not practical.)

Solution:

Note that all answers to this problem are reported to five significant figures, to distinguish the results. a. $0.99947c$; b. 1.2064×10^{11} y; c. 1.2058×10^{11} y

Exercise:

Problem:

Suppose a spaceship heading straight toward the Earth at $0.750c$ can shoot a canister at $0.500c$ relative to the ship. (a) What is the velocity of the canister relative to Earth, if it is shot directly at Earth? (b) If it is shot directly away from Earth?

Exercise:**Problem:**

Repeat the preceding problem with the ship heading directly away from Earth.

Solution:

a. $-0.400c$; b. $-0.909c$

Exercise:**Problem:**

If a spaceship is approaching the Earth at $0.100c$ and a message capsule is sent toward it at $0.100c$ relative to Earth, what is the speed of the capsule relative to the ship?

Exercise:**Problem:**

(a) Suppose the speed of light were only 3000 m/s . A jet fighter moving toward a target on the ground at 800 m/s shoots bullets, each having a muzzle velocity of 1000 m/s . What are the bullets' velocity relative to the target? (b) If the speed of light was this small, would you observe relativistic effects in everyday life? Discuss.

Solution:

a. 1.65 km/s ; b. Yes, if the speed of light were this small, speeds that we can achieve in everyday life would be larger than 1% of the speed of light and we could observe relativistic effects much more often.

Exercise:**Problem:**

If a galaxy moving away from the Earth has a speed of 1000 km/s and emits 656 nm light characteristic of hydrogen (the most common element in the universe). (a) What wavelength would we observe on Earth? (b) What type of electromagnetic radiation is this? (c) Why is the speed of Earth in its orbit negligible here?

Exercise:

Problem:

A space probe speeding towards the nearest star moves at $0.250c$ and sends radio information at a broadcast frequency of 1.00 GHz . What frequency is received on Earth?

Solution:

775 MHz

Exercise:**Problem:**

Near the center of our galaxy, hydrogen gas is moving directly away from us in its orbit about a black hole. We receive 1900 nm electromagnetic radiation and know that it was 1875 nm when emitted by the hydrogen gas. What is the speed of the gas?

Exercise:**Problem:**

(a) Calculate the speed of a $1.00\text{-}\mu\text{g}$ particle of dust that has the same momentum as a proton moving at $0.999c$. (b) What does the small speed tell us about the mass of a proton compared to even a tiny amount of macroscopic matter?

Solution:

a. $1.12 \times 10^{-8}\text{ m/s}$; b. The small speed tells us that the mass of a protein is substantially smaller than that of even a tiny amount of macroscopic matter.

Exercise:**Problem:**

(a) Calculate γ for a proton that has a momentum of $1.00\text{ kg} \cdot \text{m/s}$. (b) What is its speed? Such protons form a rare component of cosmic radiation with uncertain origins.

Exercise:**Problem:**

Show that the relativistic form of Newton's second law is (a) $F = m \frac{du}{dt} \frac{1}{(1-u^2/c^2)^{3/2}}$;

(b) Find the force needed to accelerate a mass of 1 kg by 1 m/s^2 when it is traveling at a velocity of $c/2$.

Solution:

a.

$$\begin{aligned}
 F &= \frac{dp}{dt} = \frac{d}{dt} \left(\frac{mu}{\sqrt{1-u^2/c^2}} \right) \\
 &= \frac{du}{dt} \left(\frac{m}{\sqrt{1-u^2/c^2}} \right) - \frac{1}{2} \frac{mu^2}{(1-u^2/c^2)^{3/2}} 2 \frac{du}{dt} ; \\
 &= \frac{m}{(1-u^2/c^2)^{3/2}} \frac{du}{dt}
 \end{aligned}$$

b.

$$\begin{aligned}
 F &= \frac{m}{(1-u^2/c^2)^{3/2}} \frac{du}{dt} \\
 &= \frac{1 \text{ kg}}{\left(1 - \left(\frac{1}{2}\right)^2\right)^{3/2}} \left(1 \text{ m/s}^2\right) \\
 &= 1.53 \text{ N}
 \end{aligned}$$

Exercise:

Problem:

A positron is an antimatter version of the electron, having exactly the same mass. When a positron and an electron meet, they annihilate, converting all of their mass into energy. (a) Find the energy released, assuming negligible kinetic energy before the annihilation. (b) If this energy is given to a proton in the form of kinetic energy, what is its velocity? (c) If this energy is given to another electron in the form of kinetic energy, what is its velocity?

Exercise:

Problem:

What is the kinetic energy in MeV of a π -meson that lives $1.40 \times 10^{-16} \text{ s}$ as measured in the laboratory, and $0.840 \times 10^{-16} \text{ s}$ when at rest relative to an observer, given that its rest energy is 135 MeV?

Solution:

90.0 MeV

Exercise:

Problem:

Find the kinetic energy in MeV of a neutron with a measured life span of 2065 s, given its rest energy is 939.6 MeV, and rest life span is 900s.

Exercise:

Problem:

(a) Show that $(pc)^2/(mc^2)^2 = \gamma^2 - 1$. This means that at large velocities $pc \gg mc^2$. (b) Is $E \approx pc$ when $\gamma = 30.0$, as for the astronaut discussed in the twin paradox?

Solution:

a. $\gamma^2 - 1$; b. yes

Exercise:**Problem:**

One cosmic ray neutron has a velocity of $0.250c$ relative to the Earth. (a) What is the neutron's total energy in MeV? (b) Find its momentum. (c) Is $E \approx pc$ in this situation? Discuss in terms of the equation given in part (a) of the previous problem.

Exercise:**Problem:**

What is γ for a proton having a mass energy of 938.3 MeV accelerated through an effective potential of 1.0 TV (teravolt)?

Solution:

1.07×10^3

Exercise:**Problem:**

(a) What is the effective accelerating potential for electrons at the Stanford Linear Accelerator, if $\gamma = 1.00 \times 10^5$ for them? (b) What is their total energy (nearly the same as kinetic in this case) in GeV?

Exercise:**Problem:**

(a) Using data from [Potential Energy of a System](#), find the mass destroyed when the energy in a barrel of crude oil is released. (b) Given these barrels contain 200 liters and assuming the density of crude oil is 750kg/m^3 , what is the ratio of mass destroyed to original mass, $\Delta m/m$?

Solution:

a. $6.56 \times 10^{-8} \text{ kg}$; b. $m = (200 \text{ L}) (1 \text{ m}^3/1000 \text{ L}) (750 \text{ kg/m}^3) = 150 \text{ kg}$;
therefore, $\frac{\Delta m}{m} = 4.37 \times 10^{-10}$

Exercise:

Problem:

(a) Calculate the energy released by the destruction of 1.00 kg of mass. (b) How many kilograms could be lifted to a 10.0 km height by this amount of energy?

Exercise:

Problem:

A Van de Graaff accelerator utilizes a 50.0 MV potential difference to accelerate charged particles such as protons. (a) What is the velocity of a proton accelerated by such a potential? (b) An electron?

Solution:

a. $0.314c$; b. $0.99995c$ (Five digits used to show difference from c)

Exercise:

Problem:

Suppose you use an average of $500 \text{ kW} \cdot \text{h}$ of electric energy per month in your home. (a) How long would 1.00 g of mass converted to electric energy with an efficiency of 38.0% last you? (b) How many homes could be supplied at the $500 \text{ kW} \cdot \text{h}$ per month rate for one year by the energy from the described mass conversion?

Exercise:

Problem:

(a) A nuclear power plant converts energy from nuclear fission into electricity with an efficiency of 35.0%. How much mass is destroyed in one year to produce a continuous 1000 MW of electric power? (b) Do you think it would be possible to observe this mass loss if the total mass of the fuel is 10^4 kg ?

Solution:

a. 1.00 kg ; b. This much mass would be measurable, but probably not observable just by looking because it is 0.01% of the total mass.

Exercise:

Problem:

Nuclear-powered rockets were researched for some years before safety concerns became paramount. (a) What fraction of a rocket's mass would have to be destroyed to get it into a low Earth orbit, neglecting the decrease in gravity? (Assume an orbital altitude of 250 km, and calculate both the kinetic energy (classical) and the gravitational potential energy needed.) (b) If the ship has a mass of 1.00×10^5 kg (100 tons), what total yield nuclear explosion in tons of TNT is needed?

Exercise:**Problem:**

The sun produces energy at a rate of 3.85×10^{26} W by the fusion of hydrogen. About 0.7% of each kilogram of hydrogen goes into the energy generated by the Sun. (a) How many kilograms of hydrogen undergo fusion each second? (b) If the sun is 90.0% hydrogen and half of this can undergo fusion before the sun changes character, how long could it produce energy at its current rate? (c) How many kilograms of mass is the sun losing per second? (d) What fraction of its mass will it have lost in the time found in part (b)?

Solution:

a. 6.06×10^{11} kg/s; b. 4.67×10^{10} y; c. 4.27×10^9 kg; d. 0.32%

Exercise:**Problem:**

Show that $E^2 - p^2c^2$ for a particle is invariant under Lorentz transformations.

Glossary

relativistic kinetic energy

kinetic energy of an object moving at relativistic speeds

rest energy

energy stored in an object at rest: $E_0 = mc^2$

speed of light

ultimate speed limit for any particle having mass

total energy

sum of all energies for a particle, including rest energy and kinetic energy, given for a particle of mass m and speed u by $E = \gamma mc^2$, where $\gamma = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}}$

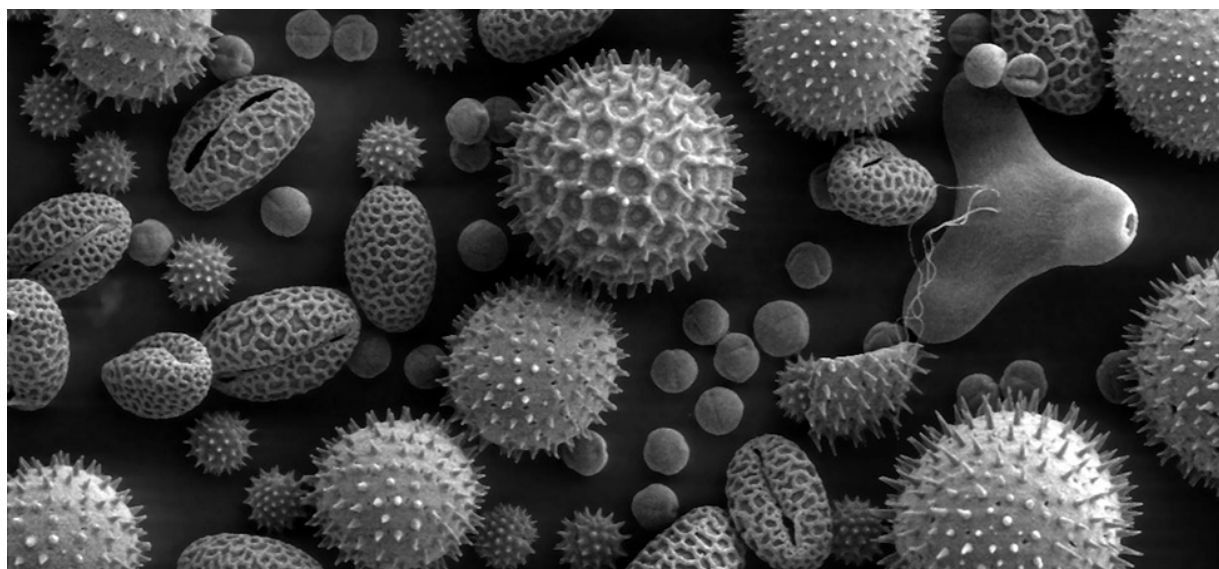
Introduction

class="introduction"

In this
image of
pollen taken
with an
electron
microscope,
the bean-
shaped
grains are
about 50 nm
long.

Electron
microscopes
can have a
much higher
resolving
power than
a
conventional light
microscope
because
electron
wavelengths
can be
100,000
times
shorter than
the
wavelengths
of visible-
light
photons.
(credit:

modification
of work by
Dartmouth
College
Electron
Microscope
Facility)



Two of the most revolutionary concepts of the twentieth century were the description of light as a collection of particles, and the treatment of particles as waves. These wave properties of matter have led to the discovery of technologies such as electron microscopy, which allows us to examine submicroscopic objects such as grains of pollen, as shown above.

In this chapter, you will learn about the energy quantum, a concept that was introduced in 1900 by the German physicist Max Planck to explain blackbody radiation. We discuss how Albert Einstein extended Planck's concept to a quantum of light (a "photon") to explain the photoelectric effect. We also show how American physicist Arthur H. Compton used the photon concept in 1923 to explain wavelength shifts observed in X-rays. After a discussion of Bohr's model of hydrogen, we describe how matter waves were postulated in 1924 by Louis-Victor de Broglie to justify Bohr's model and we examine the experiments conducted in 1923–1927 by Clinton

Davisson and Lester Germer that confirmed the existence of de Broglie's matter waves.

Blackbody Radiation

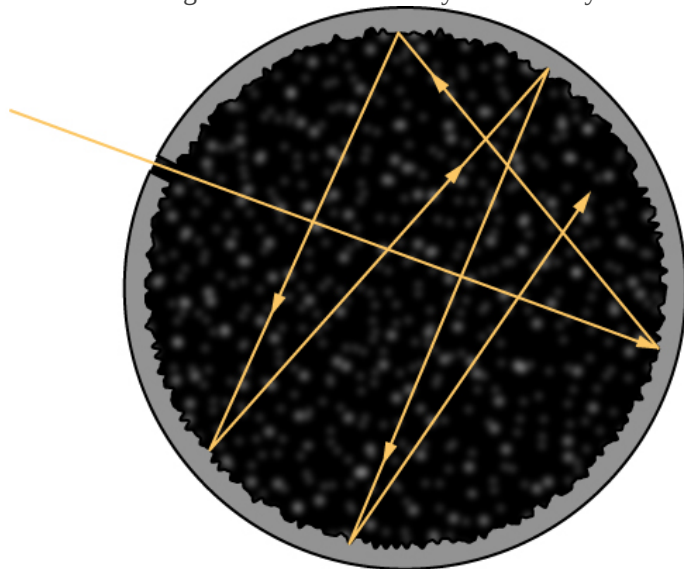
By the end of this section you will be able to:

- Apply Wien's and Stefan's laws to analyze radiation emitted by a blackbody
- Explain Planck's hypothesis of energy quanta

All bodies emit electromagnetic radiation over a range of wavelengths. In an earlier chapter, we learned that a cooler body radiates less energy than a warmer body. We also know by observation that when a body is heated and its temperature rises, the perceived wavelength of its emitted radiation changes from infrared to red, and then from red to orange, and so forth. As its temperature rises, the body glows with the colors corresponding to ever-smaller wavelengths of the electromagnetic spectrum. This is the underlying principle of the incandescent light bulb: A hot metal filament glows red, and when heating continues, its glow eventually covers the entire visible portion of the electromagnetic spectrum. The temperature (T) of the object that emits radiation, or the **emitter**, determines the wavelength at which the radiated energy is at its maximum. For example, the Sun, whose surface temperature is in the range between 5000 K and 6000 K, radiates most strongly in a range of wavelengths about 560 nm in the visible part of the electromagnetic spectrum. Your body, when at its normal temperature of about 300 K, radiates most strongly in the infrared part of the spectrum.

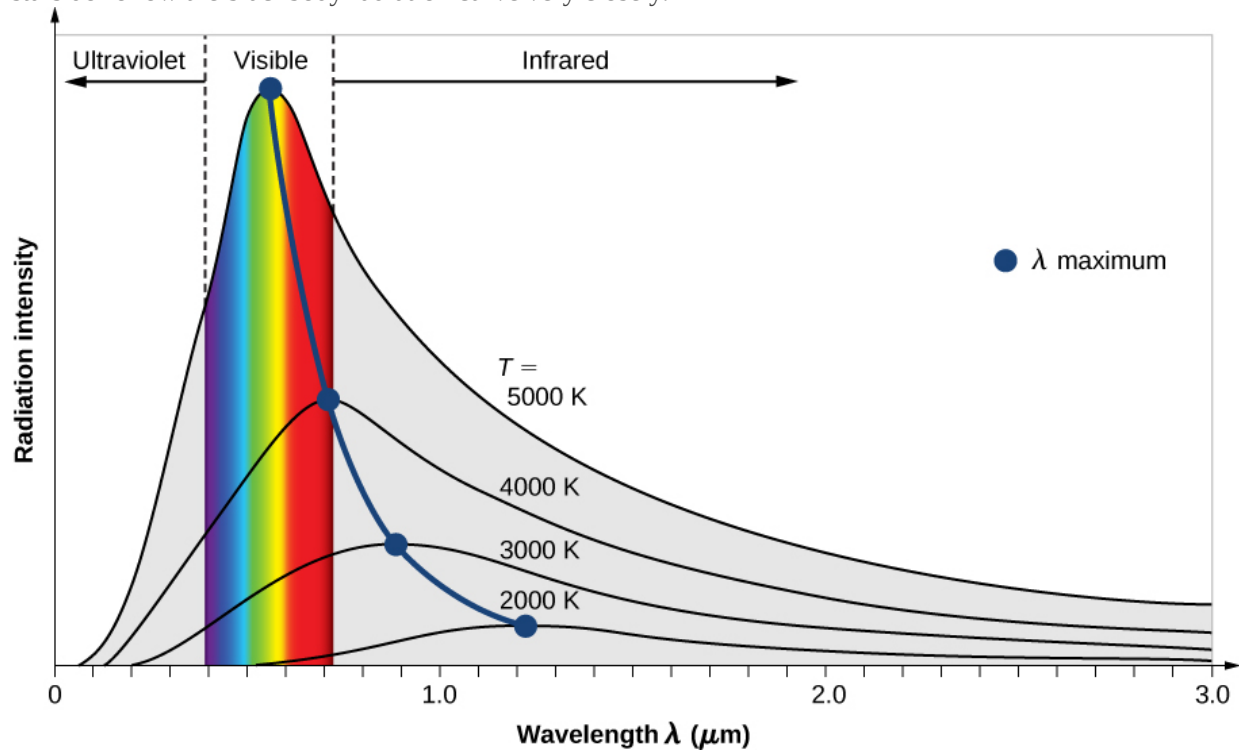
Radiation that is incident on an object is partially absorbed and partially reflected. At thermodynamic equilibrium, the rate at which an object absorbs radiation is the same as the rate at which it emits it. Therefore, a good **absorber** of radiation (any object that absorbs radiation) is also a good emitter. A perfect absorber absorbs all electromagnetic radiation incident on it; such an object is called a **blackbody**.

Although the blackbody is an idealization, because no physical object absorbs 100% of incident radiation, we can construct a close realization of a blackbody in the form of a small hole in the wall of a sealed enclosure known as a cavity radiator, as shown in [\[link\]](#). The inside walls of a cavity radiator are rough and blackened so that any radiation that enters through a tiny hole in the cavity wall becomes trapped inside the cavity. At thermodynamic equilibrium (at temperature T), the cavity walls absorb exactly as much radiation as they emit. Furthermore, inside the cavity, the radiation entering the hole is balanced by the radiation leaving it. The emission spectrum of a blackbody can be obtained by analyzing the light radiating from the hole. Electromagnetic waves emitted by a blackbody are called **blackbody radiation**.

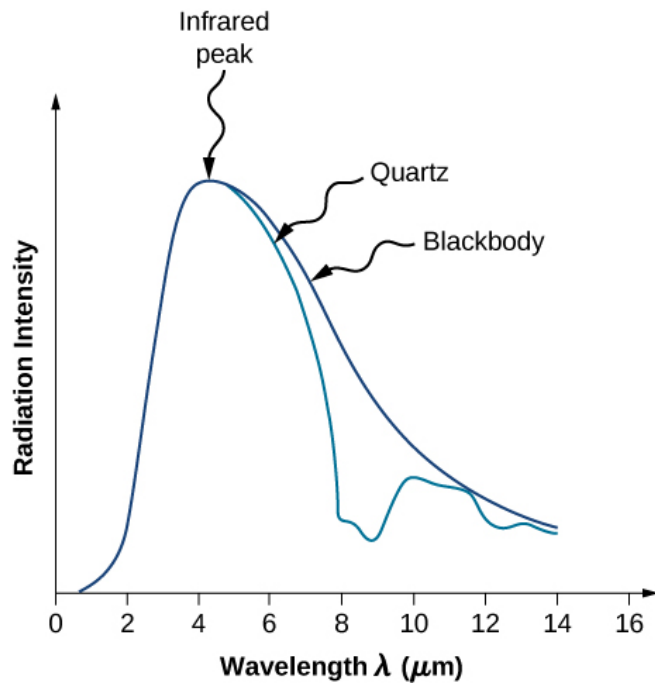


A blackbody is physically realized by a small hole in the wall of a cavity radiator.

The intensity $I(\lambda, T)$ of blackbody radiation depends on the wavelength λ of the emitted radiation and on the temperature T of the blackbody ([link](#)). The function $I(\lambda, T)$ is the **power intensity** that is radiated per unit wavelength; in other words, it is the power radiated per unit area of the hole in a cavity radiator per unit wavelength. According to this definition, $I(\lambda, T)d\lambda$ is the power per unit area that is emitted in the wavelength interval from λ to $\lambda + d\lambda$. The intensity distribution among wavelengths of radiation emitted by cavities was studied experimentally at the end of the nineteenth century. Generally, radiation emitted by materials only approximately follows the blackbody radiation curve ([link](#)); however, spectra of common stars do follow the blackbody radiation curve very closely.



The intensity of blackbody radiation versus the wavelength of the emitted radiation. Each curve corresponds to a different blackbody temperature, starting with a low temperature (the lowest curve) to a high temperature (the highest curve).



The spectrum of radiation emitted from a quartz surface (blue curve) and the blackbody radiation curve (black curve) at 600 K.

Two important laws summarize the experimental findings of blackbody radiation: *Wien's displacement law* and *Stefan's law*. Wien's displacement law is illustrated in [\[link\]](#) by the curve connecting the maxima on the intensity curves. In these curves, we see that the hotter the body, the shorter the wavelength corresponding to the emission peak in the radiation curve. Quantitatively, Wien's law reads

Note:
Equation:

$$\lambda_{\max} T = 2.898 \times 10^{-3} \text{m} \cdot \text{K}$$

where λ_{\max} is the position of the maximum in the radiation curve. In other words, λ_{\max} is the wavelength at which a blackbody radiates most strongly at a given temperature T . Note that in [\[link\]](#), the temperature is in kelvins. Wien's displacement law allows us to estimate the temperatures of distant stars by measuring the wavelength of radiation they emit.

Example:
Temperatures of Distant Stars

On a clear evening during the winter months, if you happen to be in the Northern Hemisphere and look up at the sky, you can see the constellation Orion (The Hunter). One star in this constellation, Rigel, flickers in

a blue color and another star, Betelgeuse, has a reddish color, as shown in [\[link\]](#). Which of these two stars is cooler, Betelgeuse or Rigel?

Strategy

We treat each star as a blackbody. Then according to Wien's law, its temperature is inversely proportional to the wavelength of its peak intensity. The wavelength $\lambda_{\text{max}}^{(\text{blue})}$ of blue light is shorter than the wavelength $\lambda_{\text{max}}^{(\text{red})}$ of red light. Even if we do not know the precise wavelengths, we can still set up a proportion.

Solution

Writing Wien's law for the blue star and for the red star, we have

Equation:

$$\lambda_{\text{max}}^{(\text{red})} T_{(\text{red})} = 2.898 \times 10^{-3} \text{m} \cdot \text{K} = \lambda_{\text{max}}^{(\text{blue})} T_{(\text{blue})}$$

When simplified, [\[link\]](#) gives

Equation:

$$T_{(\text{red})} = \frac{\lambda_{\text{max}}^{(\text{blue})}}{\lambda_{\text{max}}^{(\text{red})}} T_{(\text{blue})} < T_{(\text{blue})}$$

Therefore, Betelgeuse is cooler than Rigel.

Significance

Note that Wien's displacement law tells us that the higher the temperature of an emitting body, the shorter the wavelength of the radiation it emits. The qualitative analysis presented in this example is generally valid for any emitting body, whether it is a big object such as a star or a small object such as the glowing filament in an incandescent lightbulb.

Note:

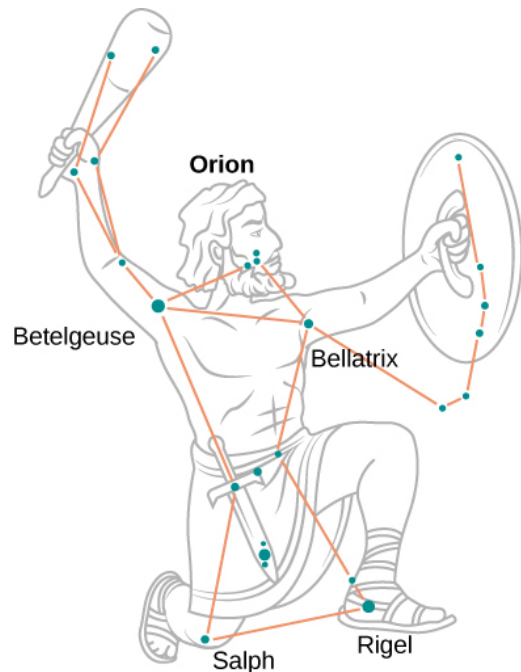
Exercise:

Problem:

Check Your Understanding The flame of a peach-scented candle has a yellowish color and the flame of a Bunsen's burner in a chemistry lab has a bluish color. Which flame has a higher temperature?

Solution:

Bunsen's burner



In the Orion constellation, the red star Betelgeuse, which usually takes on a yellowish tint, appears as the figure's right shoulder (in the upper left). The giant blue star on the bottom right is Rigel, which appears as the hunter's left foot. (credit left: modification of work by Matthew Spinelli, NASA APOD)

The second experimental relation is Stefan's law, which concerns the total power of blackbody radiation emitted across the entire spectrum of wavelengths at a given temperature. In [\[link\]](#), this total power is represented by the area under the blackbody radiation curve for a given T . As the temperature of a blackbody increases, the total emitted power also increases. Quantitatively, Stefan's law expresses this relation as

Note:
Equation:

$$P(T) = \sigma AT^4$$

where A is the surface area of a blackbody, T is its temperature (in kelvins), and σ is the **Stefan–Boltzmann constant**, $\sigma = 5.670 \times 10^{-8} \text{W}/(\text{m}^2 \cdot \text{K}^4)$. Stefan's law enables us to estimate how much energy a star is radiating by remotely measuring its temperature.

Example:
Power Radiated by Stars

A star such as our Sun will eventually evolve to a "red giant" star and then to a "white dwarf" star. A typical white dwarf is approximately the size of Earth, and its surface temperature is about $2.5 \times 10^4 \text{K}$. A typical red giant has a surface temperature of $3.0 \times 10^3 \text{K}$ and a radius $\sim 100,000$ times larger than that of a white

dwarf. What is the average radiated power per unit area and the total power radiated by each of these types of stars? How do they compare?

Strategy

If we treat the star as a blackbody, then according to Stefan's law, the total power that the star radiates is proportional to the fourth power of its temperature. To find the power radiated per unit area of the surface, we do not need to make any assumptions about the shape of the star because P/A depends only on temperature. However, to compute the total power, we need to make an assumption that the energy radiates through a spherical surface enclosing the star, so that the surface area is $A = 4\pi R^2$, where R is its radius.

Solution

A simple proportion based on Stefan's law gives

Equation:

$$\frac{P_{\text{dwarf}}/A_{\text{dwarf}}}{P_{\text{giant}}/A_{\text{giant}}} = \frac{\sigma T_{\text{dwarf}}^4}{\sigma T_{\text{giant}}^4} = \left(\frac{T_{\text{dwarf}}}{T_{\text{giant}}}\right)^4 = \left(\frac{2.5 \times 10^4}{3.0 \times 10^3}\right)^4 = 4820$$

The power emitted per unit area by a white dwarf is about 5000 times that the power emitted by a red giant. Denoting this ratio by $a = 4.8 \times 10^3$, [\[link\]](#) gives

Equation:

$$\frac{P_{\text{dwarf}}}{P_{\text{giant}}} = a \frac{A_{\text{dwarf}}}{A_{\text{giant}}} = a \frac{4\pi R_{\text{dwarf}}^2}{4\pi R_{\text{giant}}^2} = a \left(\frac{R_{\text{dwarf}}}{R_{\text{giant}}}\right)^2 = 4.8 \times 10^3 \left(\frac{R_{\text{dwarf}}}{10^5 R_{\text{dwarf}}}\right)^2 = 4.8 \times 10^{-7}$$

We see that the total power emitted by a white dwarf is a tiny fraction of the total power emitted by a red giant. Despite its relatively lower temperature, the overall power radiated by a red giant far exceeds that of the white dwarf because the red giant has a much larger surface area. To estimate the absolute value of the emitted power per unit area, we again use Stefan's law. For the white dwarf, we obtain

Equation:

$$\frac{P_{\text{dwarf}}}{A_{\text{dwarf}}} = \sigma T_{\text{dwarf}}^4 = 5.670 \times 10^{-8} \frac{\text{W}}{\text{m}^2 \cdot \text{K}^4} (2.5 \times 10^4 \text{K})^4 = 2.2 \times 10^{10} \text{W/m}^2$$

The analogous result for the red giant is obtained by scaling the result for a white dwarf:

Equation:

$$\frac{P_{\text{giant}}}{A_{\text{giant}}} = \frac{2.2 \times 10^{10}}{4.82 \times 10^3} \frac{\text{W}}{\text{m}^2} = 4.56 \times 10^6 \frac{\text{W}}{\text{m}^2} \cong 4.6 \times 10^6 \frac{\text{W}}{\text{m}^2}$$

Significance

To estimate the total power emitted by a white dwarf, in principle, we could use [\[link\]](#). However, to find its surface area, we need to know the average radius, which is not given in this example. Therefore, the solution stops here. The same is also true for the red giant star.

Note:

Exercise:

Problem:

Check Your Understanding An iron poker is being heated. As its temperature rises, the poker begins to glow—first dull red, then bright red, then orange, and then yellow. Use either the blackbody radiation curve or Wien's law to explain these changes in the color of the glow.

Solution:

The wavelength of the radiation maximum decreases with increasing temperature.

Note:**Exercise:****Problem:**

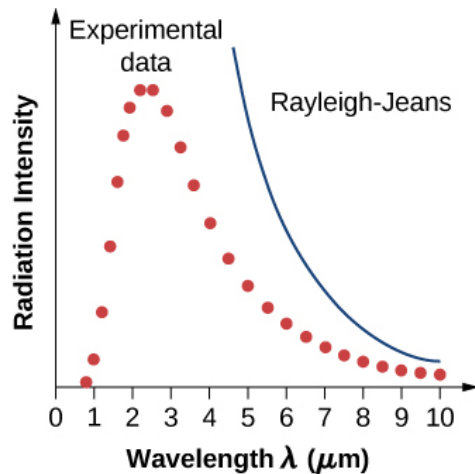
Check Your Understanding Suppose that two stars, α and β , radiate exactly the same total power. If the radius of star α is three times that of star β , what is the ratio of the surface temperatures of these stars? Which one is hotter?

Solution:

$T_\alpha / T_\beta = 1 / \sqrt{3} \cong 0.58$, so the star β is hotter.

The term “blackbody” was coined by Gustav R. Kirchhoff in 1862. The blackbody radiation curve was known experimentally, but its shape eluded physical explanation until the year 1900. The physical model of a blackbody at temperature T is that of the electromagnetic waves enclosed in a cavity (see [\[link\]](#)) and at thermodynamic equilibrium with the cavity walls. The waves can exchange energy with the walls. The objective here is to find the energy density distribution among various modes of vibration at various wavelengths (or frequencies). In other words, we want to know how much energy is carried by a single wavelength or a band of wavelengths. Once we know the energy distribution, we can use standard statistical methods (similar to those studied in a previous chapter) to obtain the blackbody radiation curve, Stefan’s law, and Wien’s displacement law. When the physical model is correct, the theoretical predictions should be the same as the experimental curves.

In a classical approach to the blackbody radiation problem, in which radiation is treated as waves (as you have studied in previous chapters), the modes of electromagnetic waves trapped in the cavity are in equilibrium and continually exchange their energies with the cavity walls. There is no physical reason why a wave should do otherwise: Any amount of energy can be exchanged, either by being transferred from the wave to the material in the wall or by being received by the wave from the material in the wall. This classical picture is the basis of the model developed by Lord Rayleigh and, independently, by Sir James Jeans. The result of this classical model for blackbody radiation curves is known as the *Rayleigh–Jeans law*. However, as shown in [\[link\]](#), the Rayleigh–Jeans law fails to correctly reproduce experimental results. In the limit of short wavelengths, the Rayleigh–Jeans law predicts infinite radiation intensity, which is inconsistent with the experimental results in which radiation intensity has finite values in the ultraviolet region of the spectrum. This divergence between the results of classical theory and experiments, which came to be called the *ultraviolet catastrophe*, shows how classical physics fails to explain the mechanism of blackbody radiation.



The ultraviolet catastrophe: The Rayleigh–Jeans law does not explain the observed blackbody emission spectrum.

The blackbody radiation problem was solved in 1900 by Max Planck. Planck used the same idea as the Rayleigh–Jeans model in the sense that he treated the electromagnetic waves between the walls inside the cavity classically, and assumed that the radiation is in equilibrium with the cavity walls. The innovative idea that Planck introduced in his model is the assumption that the cavity radiation originates from atomic oscillations inside the cavity walls, and that these oscillations can have only *discrete* values of energy. Therefore, the radiation trapped inside the cavity walls can exchange energy with the walls only in discrete amounts. Planck’s hypothesis of discrete energy values, which he called *quanta*, assumes that the oscillators inside the cavity walls have **quantized energies**. This was a brand new idea that went beyond the classical physics of the nineteenth century because, as you learned in a previous chapter, in the classical picture, the energy of an oscillator can take on any continuous value. Planck assumed that the energy of an oscillator (E_n) can have only discrete, or quantized, values:

Note:

Equation:

$$E_n = nhf, \text{ where } n = 1, 2, 3, \dots$$

In [\[link\]](#), f is the frequency of Planck’s oscillator. The natural number n that enumerates these discrete energies is called a **quantum number**. The physical constant h is called *Planck’s constant*:

Note:

Equation:

$$h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s} = 4.136 \times 10^{-15} \text{ eV} \cdot \text{s}$$

Each discrete energy value corresponds to a **quantum state of a Planck oscillator**. Quantum states are enumerated by quantum numbers. For example, when Planck's oscillator is in its first $n = 1$ quantum state, its energy is $E_1 = hf$; when it is in the $n = 2$ quantum state, its energy is $E_2 = 2hf$; when it is in the $n = 3$ quantum state, $E_3 = 3hf$; and so on.

Note that [\[link\]](#) shows that there are infinitely many quantum states, which can be represented as a sequence $\{hf, 2hf, 3hf, \dots, (n-1)hf, nhf, (n+1)hf, \dots\}$. Each two consecutive quantum states in this sequence are separated by an energy jump, $\Delta E = hf$. An oscillator in the wall can receive energy from the radiation in the cavity (absorption), or it can give away energy to the radiation in the cavity (emission). The absorption process sends the oscillator to a higher quantum state, and the emission process sends the oscillator to a lower quantum state. Whichever way this exchange of energy goes, the smallest amount of energy that can be exchanged is hf . There is no upper limit to how much energy can be exchanged, but whatever is exchanged must be an integer multiple of hf . If the energy packet does not have this exact amount, it is neither absorbed nor emitted at the wall of the blackbody.

Note:

Planck's Quantum Hypothesis

Planck's hypothesis of energy quanta states that the amount of energy emitted by the oscillator is carried by the quantum of radiation, ΔE :

Equation:

$$\Delta E = hf$$

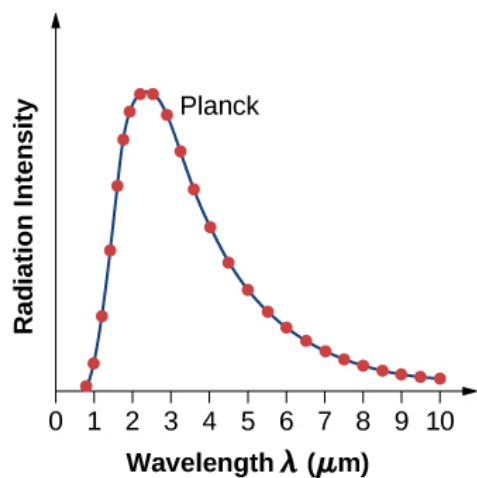
Recall that the frequency of electromagnetic radiation is related to its wavelength and to the speed of light by the fundamental relation $f\lambda = c$. This means that we can express [\[link\]](#) equivalently in terms of wavelength λ . When included in the computation of the energy density of a blackbody, Planck's hypothesis gives the following theoretical expression for the power intensity of emitted radiation per unit wavelength:

Note:

Equation:

$$I(\lambda, T) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1}$$

where c is the speed of light in vacuum and k_B is Boltzmann's constant, $k_B = 1.380 \times 10^{-23} \text{ J/K}$. The theoretical formula expressed in [\[link\]](#) is called *Planck's blackbody radiation law*. This law is in agreement with the experimental blackbody radiation curve (see [\[link\]](#)). In addition, Wien's displacement law and Stefan's law can both be derived from [\[link\]](#). To derive Wien's displacement law, we use differential calculus to find the maximum of the radiation intensity curve $I(\lambda, T)$. To derive Stefan's law and find the value of the Stefan–Boltzmann constant, we use integral calculus and integrate $I(\lambda, T)$ to find the total power radiated by a blackbody at one temperature in the entire spectrum of wavelengths from $\lambda = 0$ to $\lambda = \infty$. This derivation is left as an exercise later in this chapter.



Planck's theoretical result (continuous curve) and the experimental blackbody radiation curve (dots).

Example:

Planck's Quantum Oscillator

A quantum oscillator in the cavity wall in [\[link\]](#) is vibrating at a frequency of $5.0 \times 10^{14} \text{ Hz}$. Calculate the spacing between its energy levels.

Strategy

Energy states of a quantum oscillator are given by [\[link\]](#). The energy spacing ΔE is obtained by finding the energy difference between two adjacent quantum states for quantum numbers $n + 1$ and n .

Solution

We can substitute the given frequency and Planck's constant directly into the equation:

Equation:

$$\Delta E = E_{n+1} - E_n = (n + 1)hf - nhf = hf = (6.626 \times 10^{-34} \text{ J} \cdot \text{s})(5.0 \times 10^{14} \text{ Hz}) = 3.3 \times 10^{-19} \text{ J}$$

Significance

Note that we do not specify what kind of material was used to build the cavity. Here, a quantum oscillator is a theoretical model of an atom or molecule of material in the wall.

Note:

Exercise:

Problem:

Check Your Understanding A molecule is vibrating at a frequency of $5.0 \times 10^{14} \text{ Hz}$. What is the smallest spacing between its vibrational energy levels?

Solution:

$$3.3 \times 10^{-19} \text{ J}$$

Example:

Quantum Theory Applied to a Classical Oscillator

A 1.0-kg mass oscillates at the end of a spring with a spring constant of 1000 N/m. The amplitude of these oscillations is 0.10 m. Use the concept of quantization to find the energy spacing for this classical oscillator. Is the energy quantization significant for macroscopic systems, such as this oscillator?

Strategy

We use [\[link\]](#) as though the system were a quantum oscillator, but with the frequency f of the mass vibrating on a spring. To evaluate whether or not quantization has a significant effect, we compare the quantum energy spacing with the macroscopic total energy of this classical oscillator.

Solution

For the spring constant, $k = 1.0 \times 10^3 \text{ N/m}$, the frequency f of the mass, $m = 1.0 \text{ kg}$, is

Equation:

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}} = \frac{1}{2\pi} \sqrt{\frac{1.0 \times 10^3 \text{ N/m}}{1.0 \text{ kg}}} \simeq 5.0 \text{ Hz}$$

The energy quantum that corresponds to this frequency is

Equation:

$$\Delta E = hf = (6.626 \times 10^{-34} \text{ J} \cdot \text{s})(5.0 \text{ Hz}) = 3.3 \times 10^{-33} \text{ J}$$

When vibrations have amplitude $A = 0.10 \text{ m}$, the energy of oscillations is

Equation:

$$E = \frac{1}{2} k A^2 = \frac{1}{2} (1000 \text{ N/m})(0.1 \text{ m})^2 = 5.0 \text{ J}$$

Significance

Thus, for a classical oscillator, we have $\Delta E / E \approx 10^{-34}$. We see that the separation of the energy levels is immeasurably small. Therefore, for all practical purposes, the energy of a classical oscillator takes on continuous values. This is why classical principles may be applied to macroscopic systems encountered in everyday life without loss of accuracy.

Note:

Exercise:

Problem:

Check Your Understanding Would the result in [\[link\]](#) be different if the mass were not 1.0 kg but a tiny mass of $1.0 \mu\text{g}$, and the amplitude of vibrations were $0.10 \mu\text{m}$?

Solution:

No, because then $\Delta E / E \approx 10^{-21}$

When Planck first published his result, the hypothesis of energy quanta was not taken seriously by the physics community because it did not follow from any established physics theory at that time. It was perceived, even by Planck himself, as a useful mathematical trick that led to a good theoretical “fit” to the experimental curve. This perception was changed in 1905 when Einstein published his explanation of the photoelectric effect, in which he gave Planck’s energy quantum a new meaning: that of a particle of light.

Summary

- All bodies radiate energy. The amount of radiation a body emits depends on its temperature. The experimental Wien’s displacement law states that the hotter the body, the shorter the wavelength corresponding to the emission peak in the radiation curve. The experimental Stefan’s law states that the total power of radiation emitted across the entire spectrum of wavelengths at a given temperature is proportional to the fourth power of the Kelvin temperature of the radiating body.
- Absorption and emission of radiation are studied within the model of a blackbody. In the classical approach, the exchange of energy between radiation and cavity walls is continuous. The classical approach does not explain the blackbody radiation curve.
- To explain the blackbody radiation curve, Planck assumed that the exchange of energy between radiation and cavity walls takes place only in discrete quanta of energy. Planck’s hypothesis of energy quanta led to the theoretical Planck’s radiation law, which agrees with the experimental blackbody radiation curve; it also explains Wien’s and Stefan’s laws.

Conceptual Questions

Exercise:

Problem: Which surface has a higher temperature – the surface of a yellow star or that of a red star?

Solution:

yellow

Exercise:

Problem:

Describe what you would see when looking at a body whose temperature is increased from 1000 K to 1,000,000 K.

Exercise:

Problem: Explain the color changes in a hot body as its temperature is increased.

Solution:

goes from red to violet through the rainbow of colors

Exercise:

Problem: Speculate as to why UV light causes sunburn, whereas visible light does not.

Exercise:

Problem:

Two cavity radiators are constructed with walls made of different metals. At the same temperature, how would their radiation spectra differ?

Solution:

would not differ

Exercise:**Problem:**

Discuss why some bodies appear black, other bodies appear red, and still other bodies appear white.

Exercise:**Problem:**

If everything radiates electromagnetic energy, why can we not see objects at room temperature in a dark room?

Solution:

human eye does not see IR radiation

Exercise:**Problem:**

How much does the power radiated by a blackbody increase when its temperature (in K) is tripled?

Problems**Exercise:****Problem:**

A 200-W heater emits a 1.5- μm radiation. (a) What value of the energy quantum does it emit? (b) Assuming that the specific heat of a 4.0-kg body is $0.83\text{kcal}/\text{kg} \cdot \text{K}$, how many of these photons must be absorbed by the body to increase its temperature by 2 K? (c) How long does the heating process in (b) take, assuming that all radiation emitted by the heater gets absorbed by the body?

Solution:

a. 0.81 eV; b. 2.1×10^{23} ; c. 2 min 20 sec

Exercise:**Problem:**

A 900-W microwave generator in an oven generates energy quanta of frequency 2560 MHz. (a) How many energy quanta does it emit per second? (b) How many energy quanta must be absorbed by a pasta dish placed in the radiation cavity to increase its temperature by 45.0 K? Assume that the dish has a mass of 0.5 kg and that its specific heat is $0.9\text{kcal}/\text{kg} \cdot \text{K}$. (c) Assume that all energy quanta emitted by the generator are absorbed by the pasta dish. How long must we wait until the dish in (b) is ready?

Exercise:

Problem:

(a) For what temperature is the peak of blackbody radiation spectrum at 400 nm? (b) If the temperature of a blackbody is 800 K, at what wavelength does it radiate the most energy?

Solution:

a. 7245 K; b. 3.62 μm

Exercise:**Problem:**

The tungsten elements of incandescent light bulbs operate at 3200 K. At what wavelength does the filament radiate maximum energy?

Exercise:**Problem:**

Interstellar space is filled with radiation of wavelength 970 μm . This radiation is considered to be a remnant of the “big bang.” What is the corresponding blackbody temperature of this radiation?

Solution:

about 3 K

Exercise:**Problem:**

The radiant energy from the sun reaches its maximum at a wavelength of about 500.0 nm. What is the approximate temperature of the sun’s surface?

Glossary

absorber

any object that absorbs radiation

blackbody

perfect absorber/emitter

blackbody radiation

radiation emitted by a blackbody

emitter

any object that emits radiation

Planck’s hypothesis of energy quanta

energy exchanges between the radiation and the walls take place only in the form of discrete energy quanta

power intensity

energy that passes through a unit surface per unit time

quantized energies

discrete energies; not continuous

quantum state of a Planck's oscillator

any mode of vibration of Planck's oscillator, enumerated by quantum number

Stefan-Boltzmann constant

physical constant in Stefan's law

Photoelectric Effect

By the end of this section you will be able to:

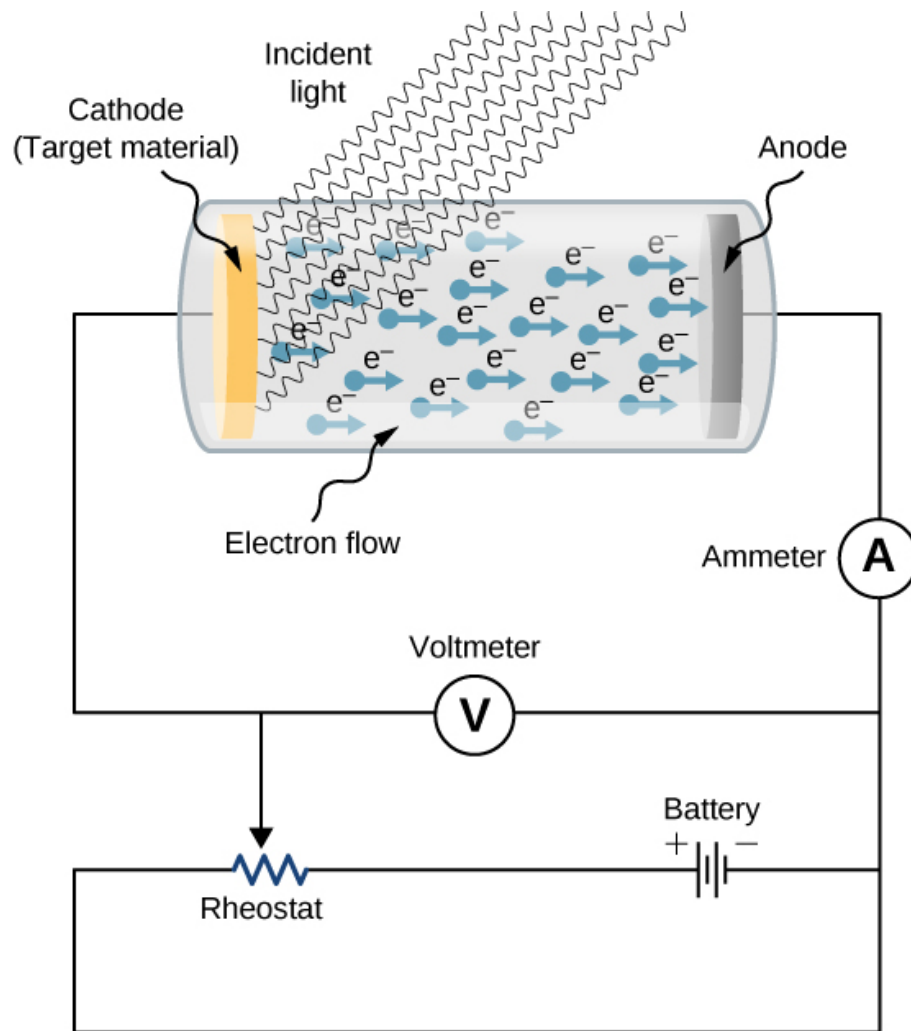
- Describe physical characteristics of the photoelectric effect
- Explain why the photoelectric effect cannot be explained by classical physics
- Describe how Einstein's idea of a particle of radiation explains the photoelectric effect

When a metal surface is exposed to a monochromatic electromagnetic wave of sufficiently short wavelength (or equivalently, above a threshold frequency), the incident radiation is absorbed and the exposed surface emits electrons. This phenomenon is known as the **photoelectric effect**. Electrons that are emitted in this process are called **photoelectrons**.

The experimental setup to study the photoelectric effect is shown schematically in [\[link\]](#). The target material serves as the cathode, which becomes the emitter of photoelectrons when it is illuminated by monochromatic radiation. We call this electrode the **photoelectrode**.

Photoelectrons are collected at the anode, which is kept at a higher potential with respect to the cathode. The potential difference between the electrodes can be increased or decreased, or its polarity can be reversed. The electrodes are enclosed in an evacuated glass tube so that photoelectrons do not lose their kinetic energy on collisions with air molecules in the space between electrodes.

When the target material is not exposed to radiation, no current is registered in this circuit because the circuit is broken (note, there is a gap between the electrodes). But when the target material is connected to the negative terminal of a battery and exposed to radiation, a current is registered in this circuit; this current is called the **photocurrent**. Suppose that we now reverse the potential difference between the electrodes so that the target material now connects with the positive terminal of a battery, and then we slowly increase the voltage. The photocurrent gradually dies out and eventually stops flowing completely at some value of this reversed voltage. The potential difference at which the photocurrent stops flowing is called the **stopping potential**.



An experimental setup to study the photoelectric effect. The anode and cathode are enclosed in an evacuated glass tube. The voltmeter measures the electric potential difference between the electrodes, and the ammeter measures the photocurrent. The incident radiation is monochromatic.

Characteristics of the Photoelectric Effect

The photoelectric effect has three important characteristics that cannot be explained by classical physics: (1) the absence of a lag time, (2) the independence of the kinetic energy of photoelectrons on the intensity of incident radiation, and (3) the presence of a cut-off frequency. Let's examine each of these characteristics.

The absence of lag time

When radiation strikes the target material in the electrode, electrons are emitted almost instantaneously, even at very low intensities of incident radiation. This absence of lag time contradicts our understanding based on classical physics. Classical physics predicts that for low-energy radiation, it would take significant time before irradiated electrons could gain sufficient energy to leave the electrode surface; however, such an energy buildup is not observed.

The intensity of incident radiation and the kinetic energy of photoelectrons

Typical experimental curves are shown in [\[link\]](#), in which the photocurrent is plotted versus the applied potential difference between the electrodes. For the positive potential difference, the current steadily grows until it reaches a plateau. Furthering the potential increase beyond this point does not increase the photocurrent at all. A higher intensity of radiation produces a higher value of photocurrent. For the negative potential difference, as the absolute value of the potential difference increases, the value of the photocurrent decreases and becomes zero at the stopping potential. For any intensity of incident radiation, whether the intensity is high or low, the value of the stopping potential always stays at one value.

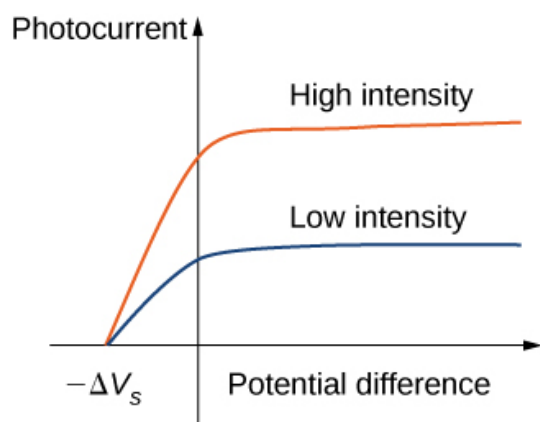
To understand why this result is unusual from the point of view of classical physics, we first have to analyze the energy of photoelectrons. A photoelectron that leaves the surface has kinetic energy K . It gained this energy from the incident electromagnetic wave. In the space between the electrodes, a photoelectron moves in the electric potential and its energy changes by the amount $q\Delta V$, where ΔV is the potential difference and $q = -e$. Because no forces are present but electric force, by applying the work-energy theorem, we obtain the energy balance $\Delta K - e\Delta V = 0$ for the photoelectron, where ΔK is the change in the photoelectron's kinetic energy. When the stopping potential $-\Delta V_s$ is applied, the photoelectron loses its initial kinetic energy K_i and comes to rest. Thus, its energy balance becomes $(0 - K_i) - e(-\Delta V_s) = 0$, so that $K_i = e\Delta V_s$. In the presence of the stopping potential, the largest kinetic energy K_{\max} that a photoelectron can have is its initial kinetic energy, which it has at the surface of the photoelectrode. Therefore, the largest kinetic energy of photoelectrons can be directly measured by measuring the stopping potential:

Note:

Equation:

$$K_{\max} = e\Delta V_s.$$

At this point we can see where the classical theory is at odds with the experimental results. In classical theory, the photoelectron absorbs electromagnetic energy in a continuous way; this means that when the incident radiation has a high intensity, the kinetic energy in [\[link\]](#) is expected to be high. Similarly, when the radiation has a low intensity, the kinetic energy is expected to be low. But the experiment shows that the maximum kinetic energy of photoelectrons is independent of the light intensity.

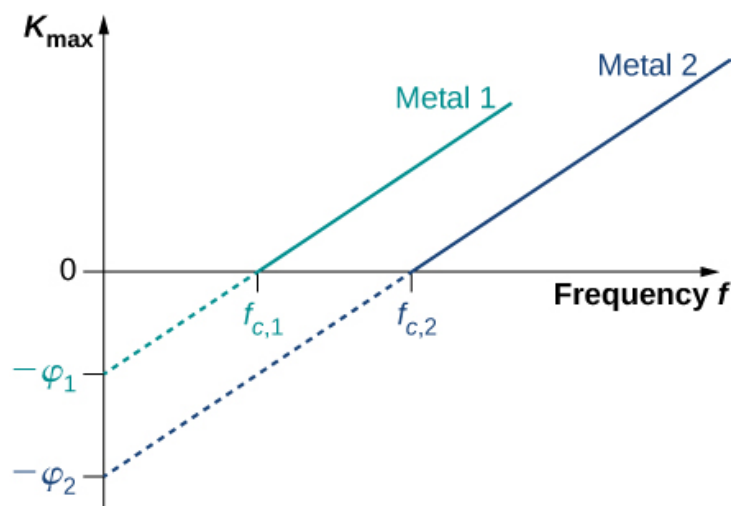


The detected photocurrent plotted versus the applied potential difference shows that for any intensity of incident radiation, whether the intensity is high (upper curve) or low (lower curve), the value of the stopping potential is always the same.

The presence of a cut-off frequency

For any metal surface, there is a minimum frequency of incident radiation below which photocurrent does not occur. The value of this **cut-off frequency** for the photoelectric effect is a physical property of the metal: Different materials have different values of cut-off frequency. Experimental data show a typical linear trend (see [\[link\]](#)). The kinetic energy of photoelectrons at the surface grows linearly with the increasing frequency of incident radiation. Measurements for all metal surfaces give linear plots with one slope. None of these observed phenomena is in accord with the classical understanding of nature. According to the classical description, the kinetic energy of photoelectrons should not depend on the frequency of incident radiation at all, and there should be no cut-off frequency. Instead, in the classical picture, electrons receive energy from the incident electromagnetic wave in a

continuous way, and the amount of energy they receive depends only on the intensity of the incident light and nothing else. So in the classical understanding, as long as the light is shining, the photoelectric effect is expected to continue.



Kinetic energy of photoelectrons at the surface versus the frequency of incident radiation. The photoelectric effect can only occur above the cut-off frequency f_c . Measurements for all metal surfaces give linear plots with one slope. Each metal surface has its own cut-off frequency.

The Work Function

The photoelectric effect was explained in 1905 by A. Einstein. Einstein reasoned that if Planck's hypothesis about energy quanta was correct for describing the energy exchange between electromagnetic radiation and cavity walls, it should also work to describe energy absorption from electromagnetic radiation by the surface of a photoelectrode. He postulated that an electromagnetic wave carries its energy in discrete packets. Einstein's postulate goes beyond Planck's hypothesis because it states that the light itself consists of energy quanta. In other words, it states that electromagnetic waves are quantized.

In Einstein's approach, a beam of monochromatic light of frequency f is made of photons. A **photon** is a particle of light. Each photon moves at the speed of light and carries an energy quantum E_f . A photon's energy depends only on its frequency f . Explicitly, the **energy of a photon** is

Note:
Equation:

$$E_f = hf$$

where h is Planck's constant. In the photoelectric effect, photons arrive at the metal surface and each photon gives away *all* of its energy to only *one* electron on the metal surface. This transfer of energy from photon to electron is of the “all or nothing” type, and there are no fractional transfers in which a photon would lose only part of its energy and survive. The essence of a **quantum phenomenon** is either a photon transfers its entire energy and ceases to exist or there is no transfer at all. This is in contrast with the classical picture, where fractional energy transfers are permitted. Having this quantum understanding, the energy balance for an electron on the surface that receives the energy E_f from a photon is

Equation:

$$E_f = K_{\max} + \phi$$

where K_{\max} is the kinetic energy, given by [\[link\]](#), that an electron has at the very instant it gets detached from the surface. In this energy balance equation, ϕ is the energy needed to detach a photoelectron from the surface. This energy ϕ is called the **work function** of the metal. Each metal has its characteristic work function, as illustrated in [\[link\]](#). To obtain the kinetic energy of photoelectrons at the surface, we simply invert the energy balance equation and use [\[link\]](#) to express the energy of the absorbed photon. This gives us the expression for the kinetic energy of photoelectrons, which explicitly depends on the frequency of incident radiation:

Note:
Equation:

$$K_{\max} = hf - \phi.$$

This equation has a simple mathematical form but its physics is profound. We can now elaborate on the physical meaning behind [\[link\]](#).

Typical Values of the Work Function for Some Common Metals	
Metal	ϕ (eV)
Na	2.46
Al	4.08
Pb	4.14
Zn	4.31
Fe	4.50
Cu	4.70
Ag	4.73
Pt	6.35

In Einstein's interpretation, interactions take place between individual electrons and individual photons. The absence of a lag time means that these one-on-one interactions occur instantaneously. This interaction time cannot be increased by lowering the light intensity. The light intensity corresponds to the number of photons arriving at the metal surface per unit time. Even at very low light intensities, the photoelectric effect still occurs because the interaction is between one electron and one photon. As long as there is at least one photon with enough energy to transfer it to a bound electron, a photoelectron will appear on the surface of the photoelectrode.

The existence of the cut-off frequency f_c for the photoelectric effect follows from [\[link\]](#) because the kinetic energy K_{\max} of the photoelectron can take only positive values. This means that there must be some threshold frequency for which the kinetic energy is zero, $0 = hf_c - \phi$. In this way, we obtain the explicit formula for cut-off frequency:

Note:
Equation:

$$f_c = \frac{\phi}{h}.$$

Cut-off frequency depends only on the work function of the metal and is in direct proportion to it. When the work function is large (when electrons are bound fast to the metal surface), the energy of the threshold photon must be large to produce a photoelectron, and then the corresponding threshold frequency is large. Photons with frequencies larger than the threshold frequency f_c always produce photoelectrons because they have $K_{\max} > 0$. Photons with frequencies smaller than f_c do not have enough energy to produce photoelectrons. Therefore, when incident radiation has a frequency below the cut-off frequency, the photoelectric effect is not observed. Because frequency f and wavelength λ of electromagnetic waves are related by the fundamental relation $\lambda f = c$ (where c is the speed of light in vacuum), the cut-off frequency has its corresponding **cut-off wavelength** λ_c :

Equation:

$$\lambda_c = \frac{c}{f_c} = \frac{c}{\phi/h} = \frac{hc}{\phi}.$$

In this equation, $hc = 1240 \text{ eV} \cdot \text{nm}$. Our observations can be restated in the following equivalent way: When the incident radiation has wavelengths longer than the cut-off wavelength, the photoelectric effect does not occur.

Example:

Photoelectric Effect for Silver

Radiation with wavelength 300 nm is incident on a silver surface. Will photoelectrons be observed?

Strategy

Photoelectrons can be ejected from the metal surface only when the incident radiation has a shorter wavelength than the cut-off wavelength. The work function of silver is $\phi = 4.73 \text{ eV}$ ([\[link\]](#)). To make the estimate, we use [\[link\]](#).

Solution

The threshold wavelength for observing the photoelectric effect in silver is

Equation:

$$\lambda_c = \frac{hc}{\phi} = \frac{1240 \text{ eV} \cdot \text{nm}}{4.73 \text{ eV}} = 262 \text{ nm}.$$

The incident radiation has wavelength 300 nm, which is longer than the cut-off wavelength; therefore, photoelectrons are not observed.

Significance

If the photoelectrode were made of sodium instead of silver, the cut-off wavelength would be 504 nm and photoelectrons would be observed.

[\[link\]](#) in Einstein's model tells us that the maximum kinetic energy of photoelectrons is a linear function of the frequency of incident radiation, which is illustrated in [\[link\]](#). For any metal, the slope of this plot has a value of Planck's constant. The intercept with the K_{\max} -axis gives us a value of the work function that is characteristic for the metal. On the other hand, K_{\max} can be directly measured in the experiment by measuring the value of the stopping potential ΔV_s (see [\[link\]](#)) at which the photocurrent stops. These direct measurements allow us to determine experimentally the value of Planck's constant, as well as work functions of materials.

Einstein's model also gives a straightforward explanation for the photocurrent values shown in [\[link\]](#). For example, doubling the intensity of radiation translates to doubling the number of photons that strike the surface per unit time. The larger the number of photons, the larger is the number of photoelectrons, which leads to a larger photocurrent in the circuit. This is how radiation intensity affects the photocurrent. The photocurrent must reach a plateau at some value of potential difference because, in unit time, the number of photoelectrons is equal to the number of incident photons and the number of incident photons does not depend on the applied potential difference at all, but only on the intensity of incident radiation. The stopping potential does not change with the radiation intensity because the kinetic energy of photoelectrons (see [\[link\]](#)) does not depend on the radiation intensity.

Example:

Work Function and Cut-Off Frequency

When a 180-nm light is used in an experiment with an unknown metal, the measured photocurrent drops to zero at potential -0.80 V. Determine the work function of the metal and its cut-off frequency for the photoelectric effect.

Strategy

To find the cut-off frequency f_c , we use [\[link\]](#), but first we must find the work function ϕ . To find ϕ , we use [\[link\]](#) and [\[link\]](#). Photocurrent drops to zero at the stopping value of potential, so we identify $\Delta V_s = 0.8$ V.

Solution

We use [\[link\]](#) to find the kinetic energy of the photoelectrons:

Equation:

$$K_{\max} = e\Delta V_s = e(0.80\text{V}) = 0.80\text{ eV}.$$

Now we solve [\[link\]](#) for ϕ :

Equation:

$$\phi = hf - K_{\max} = \frac{hc}{\lambda} - K_{\max} = \frac{1240\text{ eV} \cdot \text{nm}}{180\text{ nm}} - 0.80\text{ eV} = 6.09\text{ eV}.$$

Finally, we use [\[link\]](#) to find the cut-off frequency:

Equation:

$$f_c = \frac{\phi}{h} = \frac{6.09 \text{ eV}}{4.136 \times 10^{-15} \text{ eV} \cdot \text{s}} = 1.47 \times 10^{15} \text{ Hz}.$$

Significance

In calculations like the one shown in this example, it is convenient to use Planck's constant in the units of $\text{eV} \cdot \text{s}$ and express all energies in eV instead of joules.

Example:

The Photon Energy and Kinetic Energy of Photoelectrons

A 430-nm violet light is incident on a calcium photoelectrode with a work function of 2.71 eV.

Find the energy of the incident photons and the maximum kinetic energy of ejected electrons.

Strategy

The energy of the incident photon is $E_f = hf = hc/\lambda$, where we use $f\lambda = c$. To obtain the maximum energy of the ejected electrons, we use [\[link\]](#).

Solution

Equation:

$$E_f = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{430 \text{ nm}} = 2.88 \text{ eV}, \quad K_{\text{max}} = E_f - \phi = 2.88 \text{ eV} - 2.71 \text{ eV} = 0.17 \text{ eV}$$

Significance

In this experimental setup, photoelectrons stop flowing at the stopping potential of 0.17 V.

Note:

Exercise:

Problem:

Check Your Understanding A yellow 589-nm light is incident on a surface whose work function is 1.20 eV. What is the stopping potential? What is the cut-off wavelength?

Solution:

−0.91 V; 1040 nm

Note:

Exercise:

Problem:

Check Your Understanding Cut-off frequency for the photoelectric effect in some materials is $8.0 \times 10^{13} \text{ Hz}$. When the incident light has a frequency of $1.2 \times 10^{14} \text{ Hz}$, the stopping potential is measured as -0.16 V . Estimate a value of Planck's constant from these data (in units $\text{J} \cdot \text{s}$ and $\text{eV} \cdot \text{s}$) and determine the percentage error of your estimation.

Solution:

$$h = 6.40 \times 10^{-34} \text{ J} \cdot \text{s} = 4.0 \times 10^{-15} \text{ eV} \cdot \text{s}; -3.5\%$$

Summary

- The photoelectric effect occurs when photoelectrons are ejected from a metal surface in response to monochromatic radiation incident on the surface. It has three characteristics: (1) it is instantaneous, (2) it occurs only when the radiation is above a cut-off frequency, and (3) kinetic energies of photoelectrons at the surface do not depend of the intensity of radiation. The photoelectric effect cannot be explained by classical theory.
- We can explain the photoelectric effect by assuming that radiation consists of photons (particles of light). Each photon carries a quantum of energy. The energy of a photon depends only on its frequency, which is the frequency of the radiation. At the surface, the entire energy of a photon is transferred to one photoelectron.
- The maximum kinetic energy of a photoelectron at the metal surface is the difference between the energy of the incident photon and the work function of the metal. The work function is the binding energy of electrons to the metal surface. Each metal has its own characteristic work function.

Conceptual Questions

Exercise:**Problem:**

For the same monochromatic light source, would the photoelectric effect occur for all metals?

Solution:

No

Exercise:

Problem:

In the interpretation of the photoelectric effect, how is it known that an electron does not absorb more than one photon?

Exercise:**Problem:**

Explain how you can determine the work function from a plot of the stopping potential versus the frequency of the incident radiation in a photoelectric effect experiment. Can you determine the value of Planck's constant from this plot?

Solution:

from the slope

Exercise:**Problem:**

Suppose that in the photoelectric-effect experiment we make a plot of the detected current versus the applied potential difference. What information do we obtain from such a plot? Can we determine from it the value of Planck's constant? Can we determine the work function of the metal?

Exercise:**Problem:**

Speculate how increasing the temperature of a photoelectrode affects the outcomes of the photoelectric effect experiment.

Solution:

Answers may vary

Exercise:**Problem:**

Which aspects of the photoelectric effect cannot be explained by classical physics?

Exercise:**Problem:**

Is the photoelectric effect a consequence of the wave character of radiation or is it a consequence of the particle character of radiation? Explain briefly.

Solution:

the particle character

Exercise:

Problem:

The metals sodium, iron, and molybdenum have work functions 2.5 eV, 3.9 eV, and 4.2 eV, respectively. Which of these metals will emit photoelectrons when illuminated with 400 nm light?

Problems

Exercise:

Problem: A photon has energy 20 keV. What are its frequency and wavelength?

Solution:

$$4.835 \times 10^{18} \text{ Hz}; 0.620 \text{ \AA}$$

Exercise:

Problem:

The wavelengths of visible light range from approximately 400 to 750 nm. What is the corresponding range of photon energies for visible light?

Exercise:

Problem:

What is the longest wavelength of radiation that can eject a photoelectron from silver? Is it in the visible range?

Solution:

263 nm; no

Exercise:

Problem:

What is the longest wavelength of radiation that can eject a photoelectron from potassium, given the work function of potassium 2.24 eV? Is it in the visible range?

Exercise:

Problem:

Estimate the binding energy of electrons in magnesium, given that the wavelength of 337 nm is the longest wavelength that a photon may have to eject a photoelectron from magnesium photoelectrode.

Solution:

3.68 eV

Exercise:**Problem:**

The work function for potassium is 2.26 eV. What is the cutoff frequency when this metal is used as photoelectrode? What is the stopping potential when for the emitted electrons when this photoelectrode is exposed to radiation of frequency 1200 THz?

Exercise:**Problem:**

Estimate the work function of aluminum, given that the wavelength of 304 nm is the longest wavelength that a photon may have to eject a photoelectron from aluminum photoelectrode.

Solution:

4.09 eV

Exercise:**Problem:**

What is the maximum kinetic energy of photoelectrons ejected from sodium by the incident radiation of wavelength 450 nm?

Exercise:**Problem:**

A 120-nm UV radiation illuminates a silver-plated electrode. What is the maximum kinetic energy of the ejected photoelectrons?

Solution:

5.60 eV

Exercise:

Problem:

A 400-nm violet light ejects photoelectrons with a maximum kinetic energy of 0.860 eV from sodium photoelectrode. What is the work function of sodium?

Exercise:**Problem:**

A 600-nm light falls on a photoelectric surface and electrons with the maximum kinetic energy of 0.17 eV are emitted. Determine (a) the work function and (b) the cutoff frequency of the surface. (c) What is the stopping potential when the surface is illuminated with light of wavelength 400 nm?

Solution:

a. 1.89 eV; b. 459 THz; c. 1.21 V

Exercise:**Problem:**

The cutoff wavelength for the emission of photoelectrons from a particular surface is 500 nm. Find the maximum kinetic energy of the ejected photoelectrons when the surface is illuminated with light of wavelength 600 nm.

Exercise:**Problem:**

Find the wavelength of radiation that can eject 2.00-eV electrons from calcium electrode. The work function for calcium is 2.71 eV. In what range is this radiation?

Solution:

264 nm; UV

Exercise:**Problem:**

Find the wavelength of radiation that can eject 0.10-eV electrons from potassium electrode. The work function for potassium is 2.24 eV. In what range is this radiation?

Exercise:**Problem:**

Find the maximum velocity of photoelectrons ejected by an 80-nm radiation, if the work function of photoelectrode is 4.73 eV.

Solution:

$$1.95 \times 10^6 \text{ m/s}$$

Glossary

cut-off frequency

frequency of incident light below which the photoelectric effect does not occur

cut-off wavelength

wavelength of incident light that corresponds to cut-off frequency

energy of a photon

quantum of radiant energy, depends only on a photon's frequency

photocurrent

in a circuit, current that flows when a photoelectrode is illuminated

photoelectric effect

emission of electrons from a metal surface exposed to electromagnetic radiation of the proper frequency

photoelectrode

in a circuit, an electrode that emits photoelectrons

photoelectron

electron emitted from a metal surface in the presence of incident radiation

photon

particle of light

quantum phenomenon

in interaction with matter, photon transfers either all its energy or nothing

stopping potential

in a circuit, potential difference that stops photocurrent

work function

energy needed to detach photoelectron from the metal surface

The Compton Effect

By the end of this section, you will be able to:

- Describe Compton's experiment
- Explain the Compton wavelength shift
- Describe how experiments with X-rays confirm the particle nature of radiation

Two of Einstein's influential ideas introduced in 1905 were the theory of special relativity and the concept of a light quantum, which we now call a photon. Beyond 1905, Einstein went further to suggest that freely propagating electromagnetic waves consisted of photons that are particles of light in the same sense that electrons or other massive particles are particles of matter. A beam of monochromatic light of wavelength λ (or equivalently, of frequency f) can be seen either as a classical wave or as a collection of photons that travel in a vacuum with one speed, c (the speed of light), and all carrying the same energy, $E_f = hf$. This idea proved useful for explaining the interactions of light with particles of matter.

Momentum of a Photon

Unlike a particle of matter that is characterized by its rest mass m_0 , a photon is massless. In a vacuum, unlike a particle of matter that may vary its speed but cannot reach the speed of light, a photon travels at only one speed, which is exactly the speed of light. From the point of view of Newtonian classical mechanics, these two characteristics imply that a photon should not exist at all. For example, how can we find the linear momentum or kinetic energy of a body whose mass is zero? This apparent paradox vanishes if we describe a photon as a relativistic particle. According to the theory of special relativity, any particle in nature obeys the relativistic energy equation

Note:

Equation:

$$E^2 = p^2 c^2 + m_0^2 c^4.$$

This relation can also be applied to a photon. In [\[link\]](#), E is the total energy of a particle, p is its linear momentum, and m_0 is its rest mass. For a photon, we simply set $m_0 = 0$ in this equation. This leads to the expression for the momentum p_f of a photon

Note:
Equation:

$$p_f = \frac{E_f}{c}.$$

Here the photon's energy E_f is the same as that of a light quantum of frequency f , which we introduced to explain the photoelectric effect:

Note:
Equation:

$$E_f = hf = \frac{hc}{\lambda}.$$

The wave relation that connects frequency f with wavelength λ and speed c also holds for photons:

Equation:

$$\lambda f = c$$

Therefore, a photon can be equivalently characterized by either its energy and wavelength, or its frequency and momentum. [\[link\]](#) and [\[link\]](#) can be combined into the explicit relation between a photon's momentum and its wavelength:

Note:

Equation:

$$p_f = \frac{h}{\lambda}.$$

Notice that this equation gives us only the magnitude of the photon's momentum and contains no information about the direction in which the photon is moving. To include the direction, it is customary to write the photon's momentum as a vector:

Note:

Equation:

$$\vec{p}_f = \hbar \vec{k}.$$

In [\[link\]](#), $\hbar = h / 2\pi$ is the **reduced Planck's constant** (pronounced “h-bar”), which is just Planck's constant divided by the factor 2π . Vector \vec{k} is called the “wave vector” or propagation vector (the direction in which a photon is moving). The **propagation vector** shows the direction of the

photon's linear momentum vector. The magnitude of the wave vector is $k = \left| \vec{\mathbf{k}} \right| = 2\pi / \lambda$ and is called the **wave number**. Notice that this equation does not introduce any new physics. We can verify that the magnitude of the vector in [\[link\]](#) is the same as that given by [\[link\]](#).

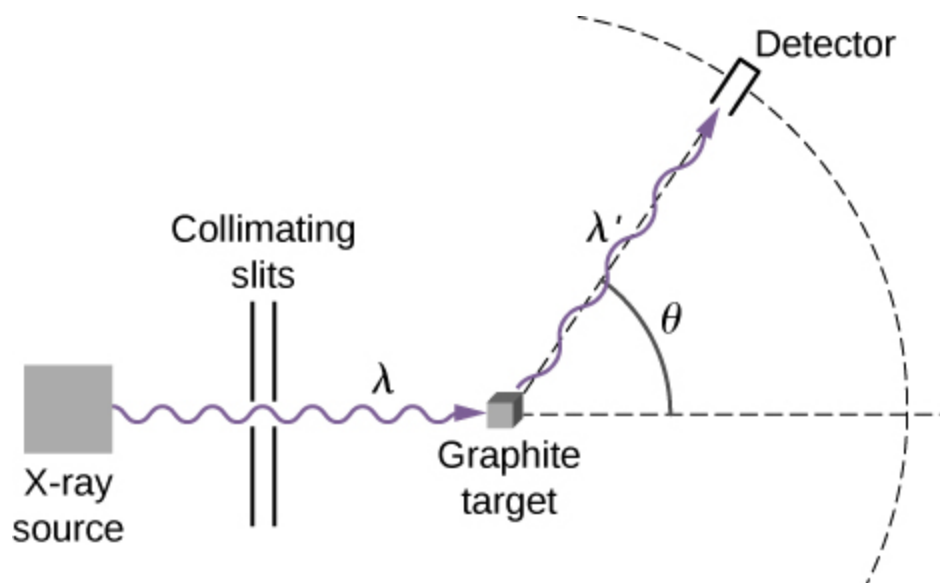
The Compton Effect

The **Compton effect** is the term used for an unusual result observed when X-rays are scattered on some materials. By classical theory, when an electromagnetic wave is scattered off atoms, the wavelength of the scattered radiation is expected to be the same as the wavelength of the incident radiation. Contrary to this prediction of classical physics, observations show that when X-rays are scattered off some materials, such as graphite, the scattered X-rays have different wavelengths from the wavelength of the incident X-rays. This classically unexplainable phenomenon was studied experimentally by Arthur H. Compton and his collaborators, and Compton gave its explanation in 1923.

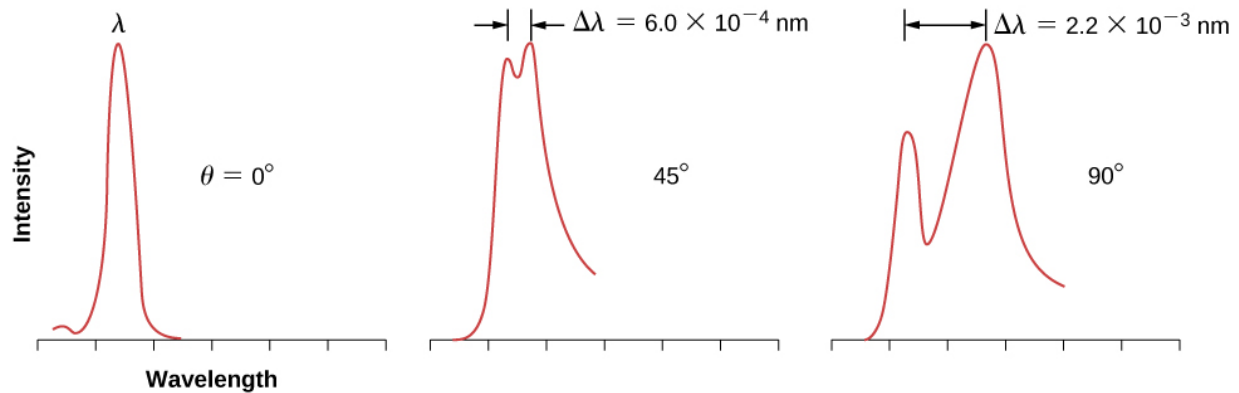
To explain the shift in wavelengths measured in the experiment, Compton used Einstein's idea of light as a particle. The Compton effect has a very important place in the history of physics because it shows that electromagnetic radiation cannot be explained as a purely wave phenomenon. The explanation of the Compton effect gave a convincing argument to the physics community that electromagnetic waves can indeed behave like a stream of photons, which placed the concept of a photon on firm ground.

The schematics of Compton's experimental setup are shown in [\[link\]](#). The idea of the experiment is straightforward: Monochromatic X-rays with wavelength λ are incident on a sample of graphite (the "target"), where they interact with atoms inside the sample; they later emerge as scattered X-rays with wavelength λ' . A detector placed behind the target can measure the intensity of radiation scattered in any direction θ with respect to the direction of the incident X-ray beam. This **scattering angle**, θ , is the angle between the direction of the scattered beam and the direction of the incident beam. In this experiment, we know the intensity and the wavelength λ of

the incoming (incident) beam; and for a given scattering angle θ , we measure the intensity and the wavelength λ' of the outgoing (scattered) beam. Typical results of these measurements are shown in [\[link\]](#), where the x-axis is the wavelength of the scattered X-rays and the y-axis is the intensity of the scattered X-rays, measured for different scattering angles (indicated on the graphs). For all scattering angles (except for $\theta = 0^\circ$), we measure two intensity peaks. One peak is located at the wavelength λ , which is the wavelength of the incident beam. The other peak is located at some other wavelength, λ' . The two peaks are separated by $\Delta\lambda$, which depends on the scattering angle θ of the outgoing beam (in the direction of observation). The separation $\Delta\lambda$ is called the **Compton shift**.



Experimental setup for studying Compton scattering.



Experimental data show the Compton effect for X-rays scattering off graphite at various angles: The intensity of the scattered beam has two peaks. One peak appears at the wavelength λ of the incident radiation and the second peak appears at wavelength λ' . The separation $\Delta\lambda$ between the peaks depends on the scattering angle θ , which is the angular position of the detector in [\[link\]](#). The experimental data in this figure are plotted in arbitrary units so that the height of the profile reflects the intensity of the scattered beam above background noise.

Compton Shift

As given by Compton, the explanation of the Compton shift is that in the target material, graphite, valence electrons are loosely bound in the atoms and behave like free electrons. Compton assumed that the incident X-ray radiation is a stream of photons. An incoming photon in this stream collides with a valence electron in the graphite target. In the course of this collision, the incoming photon transfers some part of its energy and momentum to the target electron and leaves the scene as a scattered photon. This model explains in qualitative terms why the scattered radiation has a longer wavelength than the incident radiation. Put simply, a photon that has lost some of its energy emerges as a photon with a lower frequency, or equivalently, with a longer wavelength. To show that his model was correct, Compton used it to derive the expression for the Compton shift. In his derivation, he assumed that both photon and electron are relativistic particles and that the collision obeys two commonsense principles: (1) the

conservation of linear momentum and (2) the conservation of total relativistic energy.

In the following derivation of the Compton shift, E_f and $\vec{\mathbf{p}}_f$ denote the energy and momentum, respectively, of an incident photon with frequency f . The photon collides with a relativistic electron at rest, which means that immediately before the collision, the electron's energy is entirely its rest mass energy, m_0c^2 . Immediately after the collision, the electron has energy E and momentum $\vec{\mathbf{p}}$, both of which satisfy [\[link\]](#). Immediately after the collision, the outgoing photon has energy \tilde{E}_f , momentum $\vec{\tilde{\mathbf{p}}}_f$, and frequency f' . The direction of the incident photon is horizontal from left to right, and the direction of the outgoing photon is at the angle θ , as illustrated in [\[link\]](#). The scattering angle θ is the angle between the momentum vectors $\vec{\mathbf{p}}_f$ and $\vec{\tilde{\mathbf{p}}}_f$, and we can write their scalar product:

Equation:

$$\vec{\mathbf{p}}_f \cdot \vec{\tilde{\mathbf{p}}}_f = p_f \tilde{p}_f \cos\theta.$$

Following Compton's argument, we assume that the colliding photon and electron form an isolated system. This assumption is valid for weakly bound electrons that, to a good approximation, can be treated as free particles. Our first equation is the conservation of energy for the photon-electron system:

Equation:

$$E_f + m_0c^2 = \tilde{E}_f + E.$$

The left side of this equation is the energy of the system at the instant immediately before the collision, and the right side of the equation is the energy of the system at the instant immediately after the collision. Our second equation is the conservation of linear momentum for the photon-electron system where the electron is at rest at the instant immediately before the collision:

Equation:

$$\vec{p}_f = \vec{\tilde{p}}_f + \vec{p}.$$

The left side of this equation is the momentum of the system right before the collision, and the right side of the equation is the momentum of the system right after collision. The entire physics of Compton scattering is contained in these three preceding equations—the remaining part is algebra. At this point, we could jump to the concluding formula for the Compton shift, but it is beneficial to highlight the main algebraic steps that lead to Compton’s formula, which we give here as follows.

We start with rearranging the terms in [\[link\]](#) and squaring it:

Equation:

$$\left[\left(E_f - \tilde{E}_f \right) + m_0 c^2 \right]^2 = E^2.$$

In the next step, we substitute [\[link\]](#) for E^2 , simplify, and divide both sides by c^2 to obtain

Equation:

$$\left(E_f / c - \tilde{E}_f / c \right)^2 + 2m_0 c \left(E_f / c - \tilde{E}_f / c \right) = p^2.$$

Now we can use [\[link\]](#) to express this form of the energy equation in terms of momenta. The result is

Equation:

$$(p_f - \tilde{p}_f)^2 + 2m_0 c (p_f - \tilde{p}_f) = p^2.$$

To eliminate p^2 , we turn to the momentum equation [\[link\]](#), rearrange its terms, and square it to obtain

Equation:

$$\left(\vec{p}_f - \vec{\tilde{p}}_f\right)^2 = p^2 \text{ and } \left(\vec{p}_f - \vec{\tilde{p}}_f\right)^2 = p_f^2 + \tilde{p}_f^2 - 2\vec{p}_f \cdot \vec{\tilde{p}}_f.$$

The product of the momentum vectors is given by [\[link\]](#). When we substitute this result for p^2 in [\[link\]](#), we obtain the energy equation that contains the scattering angle θ :

Equation:

$$(p_f - \tilde{p}_f)^2 + 2m_0c(p_f - \tilde{p}_f) = p_f^2 + \tilde{p}_f^2 - 2p_f\tilde{p}_f\cos\theta.$$

With further algebra, this result can be simplified to

Equation:

$$\frac{1}{\tilde{p}_f} - \frac{1}{p_f} = \frac{1}{m_0c}(1 - \cos\theta).$$

Now recall [\[link\]](#) and write: $1/\tilde{p}_f = \lambda'/h$ and $1/p_f = \lambda/h$. When these relations are substituted into [\[link\]](#), we obtain the relation for the Compton shift:

Equation:

$$\lambda' - \lambda = \frac{h}{m_0c}(1 - \cos\theta).$$

The factor h/m_0c is called the **Compton wavelength** of the electron:

Note:

Equation:

$$\lambda_c = \frac{h}{m_0c} = 0.00243 \text{ nm} = 2.43 \text{ pm}.$$

Denoting the shift as $\Delta\lambda = \lambda' - \lambda$, the concluding result can be rewritten as

Note:

Equation:

$$\Delta\lambda = \lambda_c(1 - \cos\theta).$$

This formula for the Compton shift describes outstandingly well the experimental results shown in [\[link\]](#). Scattering data measured for molybdenum, graphite, calcite, and many other target materials are in accord with this theoretical result. The nonshifted peak shown in [\[link\]](#) is due to photon collisions with tightly bound inner electrons in the target material. Photons that collide with the inner electrons of the target atoms in fact collide with the entire atom. In this extreme case, the rest mass in [\[link\]](#) must be changed to the rest mass of the atom. This type of shift is four orders of magnitude smaller than the shift caused by collisions with electrons and is so small that it can be neglected.

Compton scattering is an example of **inelastic scattering**, in which the scattered radiation has a longer wavelength than the wavelength of the incident radiation. In today's usage, the term "Compton scattering" is used for the inelastic scattering of photons by free, charged particles. In Compton scattering, treating photons as particles with momenta that can be transferred to charged particles provides the theoretical background to explain the wavelength shifts measured in experiments; this is the evidence that radiation consists of photons.

Example:

Compton Scattering

An incident 71-pm X-ray is incident on a calcite target. Find the wavelength of the X-ray scattered at a 30° angle. What is the largest shift

that can be expected in this experiment?

Strategy

To find the wavelength of the scattered X-ray, first we must find the Compton shift for the given scattering angle, $\theta = 30^\circ$. We use [\[link\]](#). Then we add this shift to the incident wavelength to obtain the scattered wavelength. The largest Compton shift occurs at the angle θ when $1 - \cos\theta$ has the largest value, which is for the angle $\theta = 180^\circ$.

Solution

The shift at $\theta = 30^\circ$ is

Equation:

$$\Delta\lambda = \lambda_c(1 - \cos 30^\circ) = 0.134\lambda_c = (0.134)(2.43) \text{ pm} = 0.325 \text{ pm}.$$

This gives the scattered wavelength:

Equation:

$$\lambda' = \lambda + \Delta\lambda = (71 + 0.325) \text{ pm} = 71.325 \text{ pm}.$$

The largest shift is

Equation:

$$(\Delta\lambda)_{\text{max}} = \lambda_c(1 - \cos 180^\circ) = 2(2.43 \text{ pm}) = 4.86 \text{ pm}.$$

Significance

The largest shift in wavelength is detected for the backscattered radiation; however, most of the photons from the incident beam pass through the target and only a small fraction of photons gets backscattered (typically, less than 5%). Therefore, these measurements require highly sensitive detectors.

Note:

Exercise:

Problem:

Check Your Understanding An incident 71-pm X-ray is incident on a calcite target. Find the wavelength of the X-ray scattered at a 60° angle. What is the smallest shift that can be expected in this experiment?

Solution:

$$(\Delta\lambda)_{\min} = 0\text{m at a } 0^\circ \text{ angle; } 71.0 \text{ pm} + 0.5\lambda_c = 72.215 \text{ pm}$$

Summary

- In the Compton effect, X-rays scattered off some materials have different wavelengths than the wavelength of the incident X-rays. This phenomenon does not have a classical explanation.
- The Compton effect is explained by assuming that radiation consists of photons that collide with weakly bound electrons in the target material. Both electron and photon are treated as relativistic particles. Conservation laws of the total energy and of momentum are obeyed in collisions.
- Treating the photon as a particle with momentum that can be transferred to an electron leads to a theoretical Compton shift that agrees with the wavelength shift measured in the experiment. This provides evidence that radiation consists of photons.
- Compton scattering is an inelastic scattering, in which scattered radiation has a longer wavelength than that of incident radiation.

Conceptual Questions**Exercise:**

Problem:

Discuss any similarities and differences between the photoelectric and the Compton effects.

Solution:

Answers may vary

Exercise:**Problem:**

Which has a greater momentum: an UV photon or an IR photon?

Exercise:**Problem:**

Does changing the intensity of a monochromatic light beam affect the momentum of the individual photons in the beam? Does such a change affect the net momentum of the beam?

Solution:

no; yes

Exercise:**Problem:**

Can the Compton effect occur with visible light? If so, will it be detectable?

Exercise:**Problem:**

Is it possible in the Compton experiment to observe scattered X-rays that have a shorter wavelength than the incident X-ray radiation?

Solution:

no

Exercise:

Problem:

Show that the Compton wavelength has the dimension of length.

Exercise:

Problem:

At what scattering angle is the wavelength shift in the Compton effect equal to the Compton wavelength?

Solution:

right angle

Problems

Exercise:

Problem: What is the momentum of a 589-nm yellow photon?

Exercise:

Problem: What is the momentum of a 4-cm microwave photon?

Solution:

$$1.66 \times 10^{-32} \text{ kg} \cdot \text{m/s}$$

Exercise:

Problem:

In a beam of white light (wavelengths from 400 to 750 nm), what range of momentum can the photons have?

Exercise:

Problem:

What is the energy of a photon whose momentum is $3.0 \times 10^{-24} \text{ kg} \cdot \text{m/s}$?

Solution:

56.21 eV

Exercise:**Problem:**

What is the wavelength of (a) a 12-keV X-ray photon; (b) a 2.0-MeV γ -ray photon?

Exercise:

Problem: Find the momentum and energy of a 1.0-Å photon.

Solution:

$6.63 \times 10^{-23} \text{ kg} \cdot \text{m/s}$; 124 keV

Exercise:**Problem:**

Find the wavelength and energy of a photon with momentum $5.00 \times 10^{-29} \text{ kg} \cdot \text{m/s}$.

Exercise:**Problem:**

A γ -ray photon has a momentum of $8.00 \times 10^{-21} \text{ kg} \cdot \text{m/s}$. Find its wavelength and energy.

Solution:

82.9 fm; 15 MeV

Exercise:**Problem:**

(a) Calculate the momentum of a $2.5\text{-}\mu\text{m}$ photon. (b) Find the velocity of an electron with the same momentum. (c) What is the kinetic energy of the electron, and how does it compare to that of the photon?

Exercise:**Problem:**

Show that $p = h / \lambda$ and $E_f = hf$ are consistent with the relativistic formula $E^2 = p^2c^2 + m_0^2c^2$.

Solution:

(Proof)

Exercise:**Problem:**

Show that the energy E in eV of a photon is given by $E = 1.241 \times 10^{-6} \text{eV} \cdot \text{m} / \lambda$, where λ is its wavelength in meters.

Exercise:**Problem:**

For collisions with free electrons, compare the Compton shift of a photon scattered as an angle of 30° to that of a photon scattered at 45° .

Solution:

$$\Delta\lambda_{30} / \Delta\lambda_{45} = 45.74\%$$

Exercise:

Problem:

X-rays of wavelength 12.5 pm are scattered from a block of carbon. What are the wavelengths of photons scattered at (a) 30° ; (b) 90° ; and, (c) 180° ?

Glossary

Compton effect

the change in wavelength when an X-ray is scattered by its interaction with some materials

Compton shift

difference between the wavelengths of the incident X-ray and the scattered X-ray

Compton wavelength

physical constant with the value $\lambda_c = 2.43 \text{ pm}$

inelastic scattering

scattering effect where kinetic energy is not conserved but the total energy is conserved

propagation vector

vector with magnitude $2\pi / \lambda$ that has the direction of the photon's linear momentum

reduced Planck's constant

Planck's constant divided by 2π

scattering angle

angle between the direction of the scattered beam and the direction of the incident beam

wave number

magnitude of the propagation vector

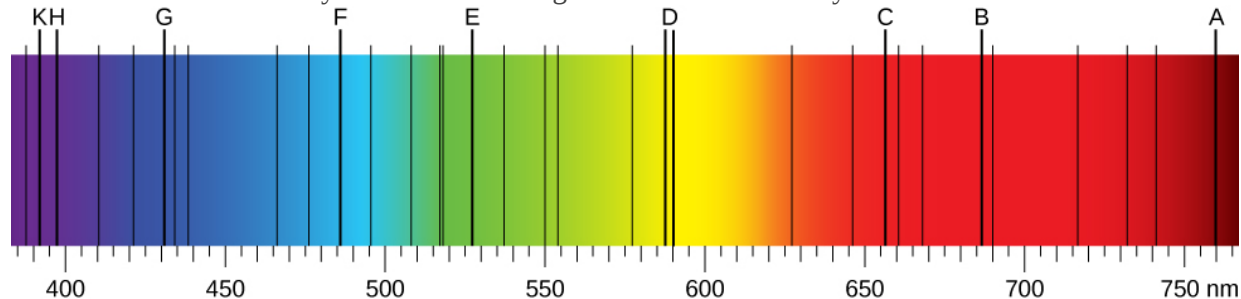
Bohr's Model of the Hydrogen Atom

By the end of this section, you will be able to:

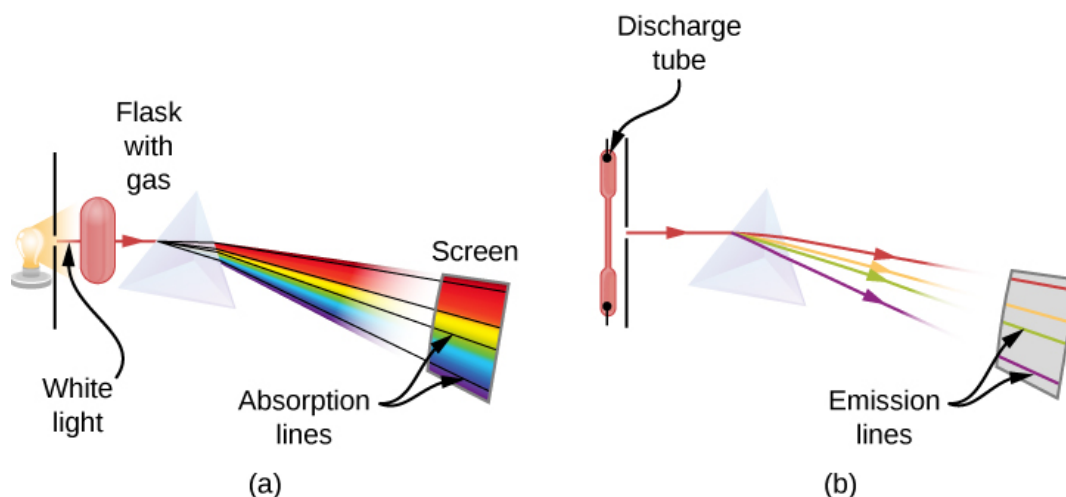
- Explain the difference between the absorption spectrum and the emission spectrum of radiation emitted by atoms
- Describe the Rutherford gold foil experiment and the discovery of the atomic nucleus
- Explain the atomic structure of hydrogen
- Describe the postulates of the early quantum theory for the hydrogen atom
- Summarize how Bohr's quantum model of the hydrogen atom explains the radiation spectrum of atomic hydrogen

Historically, Bohr's model of the hydrogen atom is the very first model of atomic structure that correctly explained the radiation spectra of atomic hydrogen. The model has a special place in the history of physics because it introduced an early quantum theory, which brought about new developments in scientific thought and later culminated in the development of quantum mechanics. To understand the specifics of Bohr's model, we must first review the nineteenth-century discoveries that prompted its formulation.

When we use a prism to analyze white light coming from the sun, several dark lines in the solar spectrum are observed ([link](#)). Solar absorption lines are called **Fraunhofer lines** after Joseph von Fraunhofer, who accurately measured their wavelengths. During 1854–1861, Gustav Kirchhoff and Robert Bunsen discovered that for the various chemical elements, the line **emission spectrum** of an element exactly matches its line **absorption spectrum**. The difference between the absorption spectrum and the emission spectrum is explained in [link](#). An absorption spectrum is observed when light passes through a gas. This spectrum appears as black lines that occur only at certain wavelengths on the background of the continuous spectrum of white light ([link](#)). The missing wavelengths tell us which wavelengths of the radiation are absorbed by the gas. The emission spectrum is observed when light is emitted by a gas. This spectrum is seen as colorful lines on the black background (see [link](#) and [link](#)). Positions of the emission lines tell us which wavelengths of the radiation are emitted by the gas. Each chemical element has its own characteristic emission spectrum. For each element, the positions of its emission lines are exactly the same as the positions of its absorption lines. This means that atoms of a specific element absorb radiation only at specific wavelengths and radiation that does not have these wavelengths is not absorbed by the element at all. This also means that the radiation emitted by atoms of each element has exactly the same wavelengths as the radiation they absorb.



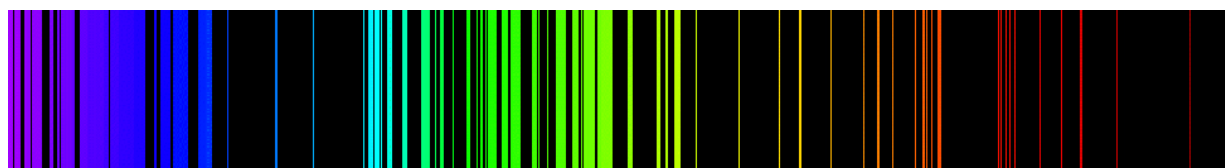
In the solar emission spectrum in the visible range from 380 nm to 710 nm, Fraunhofer lines are observed as vertical black lines at specific spectral positions in the continuous spectrum. Highly sensitive modern instruments observe thousands of such lines.



Observation of line spectra: (a) setup to observe absorption lines; (b) setup to observe emission lines. (a) White light passes through a cold gas that is contained in a glass flask. A prism is used to separate wavelengths of the passed light. In the spectrum of the passed light, some wavelengths are missing, which are seen as black absorption lines in the continuous spectrum on the viewing screen. (b) A gas is contained in a glass discharge tube that has electrodes at its ends. At a high potential difference between the electrodes, the gas glows and the light emitted from the gas passes through the prism that separates its wavelengths. In the spectrum of the emitted light, only specific wavelengths are present, which are seen as colorful emission lines on the screen.



The emission spectrum of atomic hydrogen: The spectral positions of emission lines are characteristic for hydrogen atoms. (credit: "Merikanto"/Wikimedia Commons)



The emission spectrum of atomic iron: The spectral positions of emission lines are characteristic for iron atoms.

Emission spectra of the elements have complex structures; they become even more complex for elements with higher atomic numbers. The simplest spectrum, shown in [\[link\]](#), belongs to the hydrogen

atom. Only four lines are visible to the human eye. As you read from right to left in [\[link\]](#), these lines are: red (656 nm), called the H- α line; aqua (486 nm), blue (434 nm), and violet (410 nm). The lines with wavelengths shorter than 400 nm appear in the ultraviolet part of the spectrum ([\[link\]](#), far left) and are invisible to the human eye. There are infinitely many invisible spectral lines in the series for hydrogen.

An empirical formula to describe the positions (wavelengths) λ of the hydrogen emission lines in this series was discovered in 1885 by Johann Balmer. It is known as the **Balmer formula**:

Note:

Equation:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{2^2} - \frac{1}{n^2} \right).$$

The constant $R_H = 1.09737 \times 10^7 \text{m}^{-1}$ is called the **Rydberg constant for hydrogen**. In [\[link\]](#), the positive integer n takes on values $n = 3, 4, 5, 6$ for the four visible lines in this series. The series of emission lines given by the Balmer formula is called the **Balmer series** for hydrogen. Other emission lines of hydrogen that were discovered in the twentieth century are described by the **Rydberg formula**, which summarizes all of the experimental data:

Note:

Equation:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right), \text{ where } n_i = n_f + 1, n_f + 2, n_f + 3, \dots$$

When $n_f = 1$, the series of spectral lines is called the **Lyman series**. When $n_f = 2$, the series is called the Balmer series, and in this case, the Rydberg formula coincides with the Balmer formula. When $n_f = 3$, the series is called the **Paschen series**. When $n_f = 4$, the series is called the **Brackett series**. When $n_f = 5$, the series is called the **Pfund series**. When $n_f = 6$, we have the **Humphreys series**. As you may guess, there are infinitely many such spectral bands in the spectrum of hydrogen because n_f can be any positive integer number.

The Rydberg formula for hydrogen gives the exact positions of the spectral lines as they are observed in a laboratory; however, at the beginning of the twentieth century, nobody could explain why it worked so well. The Rydberg formula remained unexplained until the first successful model of the hydrogen atom was proposed in 1913.

Example:**Limits of the Balmer Series**

Calculate the longest and the shortest wavelengths in the Balmer series.

Strategy

We can use either the Balmer formula or the Rydberg formula. The longest wavelength is obtained when $1/n_i$ is largest, which is when $n_i = n_f + 1 = 3$, because $n_f = 2$ for the Balmer series. The smallest wavelength is obtained when $1/n_i$ is smallest, which is $1/n_i \rightarrow 0$ when $n_i \rightarrow \infty$.

Solution

The long-wave limit:

Equation:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{2^2} - \frac{1}{3^2} \right) = (1.09737 \times 10^7) \frac{1}{\text{m}} \left(\frac{1}{4} - \frac{1}{9} \right) \Rightarrow \lambda = 656.3 \text{ nm}$$

The short-wave limit:

Equation:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{2^2} - 0 \right) = (1.09737 \times 10^7) \frac{1}{\text{m}} \left(\frac{1}{4} \right) \Rightarrow \lambda = 364.6 \text{ nm}$$

Significance

Note that there are infinitely many spectral lines lying between these two limits.

Note:**Exercise:****Problem:**

Check Your Understanding What are the limits of the Lyman series? Can you see these spectral lines?

Solution:

121.5 nm and 91.1 nm; no, these spectral bands are in the ultraviolet

The key to unlocking the mystery of atomic spectra is in understanding atomic structure. Scientists have long known that matter is made of atoms. According to nineteenth-century science, atoms are the smallest indivisible quantities of matter. This scientific belief was shattered by a series of groundbreaking experiments that proved the existence of subatomic particles, such as electrons, protons, and neutrons.

The electron was discovered and identified as the smallest quantity of electric charge by J.J. Thomson in 1897 in his cathode ray experiments, also known as β -ray experiments: A **β -ray** is a beam of electrons. In 1904, Thomson proposed the first model of atomic structure, known as the “plum pudding” model, in which an atom consisted of an unknown positively charged matter with negative electrons embedded in it like plums in a pudding. Around 1900, E. Rutherford, and independently, Paul Ulrich Villard, classified all radiation known at that time as **α -rays**, **β -rays**, and **γ -rays** (a γ -ray is a

beam of highly energetic photons). In 1907, Rutherford and Thomas Royds used spectroscopy methods to show that positively charged particles of α -radiation (called **α -particles**) are in fact doubly ionized atoms of helium. In 1909, Rutherford, Ernest Marsden, and Hans Geiger used α -particles in their famous scattering experiment that disproved Thomson's model (see [Linear Momentum and Collisions](#)).

In the **Rutherford gold foil experiment** (also known as the Geiger–Marsden experiment), α -particles were incident on a thin gold foil and were scattered by gold atoms inside the foil (see [Types of Collisions](#)). The outgoing particles were detected by a 360° scintillation screen surrounding the gold target (for a detailed description of the experimental setup, see [Linear Momentum and Collisions](#)). When a scattered particle struck the screen, a tiny flash of light (scintillation) was observed at that location. By counting the scintillations seen at various angles with respect to the direction of the incident beam, the scientists could determine what fraction of the incident particles were scattered and what fraction were not deflected at all. If the plum pudding model were correct, there would be no back-scattered α -particles. However, the results of the Rutherford experiment showed that, although a sizable fraction of α -particles emerged from the foil not scattered at all as though the foil were not in their way, a significant fraction of α -particles were back-scattered toward the source. This kind of result was possible only when most of the mass and the entire positive charge of the gold atom were concentrated in a tiny space inside the atom.

In 1911, Rutherford proposed a **nuclear model of the atom**. In Rutherford's model, an atom contained a positively charged nucleus of negligible size, almost like a point, but included almost the entire mass of the atom. The atom also contained negative electrons that were located within the atom but relatively far away from the nucleus. Ten years later, Rutherford coined the name *proton* for the nucleus of hydrogen and the name *neutron* for a hypothetical electrically neutral particle that would mediate the binding of positive protons in the nucleus (the neutron was discovered in 1932 by James Chadwick). Rutherford is credited with the discovery of the atomic nucleus; however, the Rutherford model of atomic structure does not explain the Rydberg formula for the hydrogen emission lines.

Bohr's model of the hydrogen atom, proposed by Niels Bohr in 1913, was the first quantum model that correctly explained the hydrogen emission spectrum. Bohr's model combines the classical mechanics of planetary motion with the quantum concept of photons. Once Rutherford had established the existence of the atomic nucleus, Bohr's intuition that the negative electron in the hydrogen atom must revolve around the positive nucleus became a logical consequence of the inverse-square-distance law of electrostatic attraction. Recall that Coulomb's law describing the attraction between two opposite charges has a similar form to Newton's universal law of gravitation in the sense that the gravitational force and the electrostatic force are both decreasing as $1/r^2$, where r is the separation distance between the bodies. In the same way as Earth revolves around the sun, the negative electron in the hydrogen atom can revolve around the positive nucleus. However, an accelerating charge radiates its energy. Classically, if the electron moved around the nucleus in a planetary fashion, it would be undergoing centripetal acceleration, and thus would be radiating energy that would cause it to spiral down into the nucleus. Such a planetary hydrogen atom would not be stable, which is contrary to what we know about ordinary hydrogen atoms that do not disintegrate. Moreover, the classical motion of the electron is not able to explain the discrete emission spectrum of hydrogen.

To circumvent these two difficulties, Bohr proposed the following three **postulates of Bohr's model**:

1. The negative electron moves around the positive nucleus (proton) in a circular orbit. All electron orbits are centered at the nucleus. Not all classically possible orbits are available to an electron bound to the nucleus.
2. The allowed electron orbits satisfy the *first quantization condition*: In the n th orbit, the angular momentum L_n of the electron can take only discrete values:

Note:

Equation:

$$L_n = n\hbar, \text{ where } n = 1, 2, 3, \dots$$

This postulate says that the electron's angular momentum is quantized. Denoted by r_n and v_n , respectively, the radius of the n th orbit and the electron's speed in it, the first quantization condition can be expressed explicitly as

Equation:

$$m_e v_n r_n = n\hbar.$$

3. An electron is allowed to make transitions from one orbit where its energy is E_n to another orbit where its energy is E_m . When an atom absorbs a photon, the electron makes a transition to a higher-energy orbit. When an atom emits a photon, the electron transits to a lower-energy orbit. Electron transitions with the simultaneous photon absorption or photon emission take place *instantaneously*. The allowed electron transitions satisfy the *second quantization condition*:

Note:

Equation:

$$hf = |E_n - E_m|$$

where hf is the energy of either an emitted or an absorbed photon with frequency f . The second quantization condition states that an electron's change in energy in the hydrogen atom is quantized.

These three postulates of the early quantum theory of the hydrogen atom allow us to derive not only the Rydberg formula, but also the value of the Rydberg constant and other important properties of the hydrogen atom such as its energy levels, its ionization energy, and the sizes of electron orbits. Note that in Bohr's model, along with two nonclassical quantization postulates, we also have the classical description of the electron as a particle that is subjected to the Coulomb force, and its motion must obey Newton's laws of motion. The hydrogen atom, as an isolated system, must obey the laws of conservation of energy and momentum in the way we know from classical physics. Having this theoretical framework in mind, we are ready to proceed with our analysis.

Electron Orbits

To obtain the size r_n of the electron's n th orbit and the electron's speed v_n in it, we turn to Newtonian mechanics. As a charged particle, the electron experiences an electrostatic pull toward the positively charged nucleus in the center of its circular orbit. This electrostatic pull is the centripetal force that causes the electron to move in a circle around the nucleus. Therefore, the magnitude of centripetal force is identified with the magnitude of the electrostatic force:

Equation:

$$\frac{m_e v_n^2}{r_n} = \frac{1}{4\pi\epsilon_0} \frac{e^2}{r_n^2}.$$

Here, e denotes the value of the elementary charge. The negative electron and positive proton have the same value of charge,

$$|q| = e.$$

When [\[link\]](#) is combined with the first quantization condition given by [\[link\]](#), we can solve for the speed, v_n , and for the radius, r_n :

Equation:

$$v_n = \frac{1}{4\pi\epsilon_0} \frac{e^2}{\hbar} \frac{1}{n}$$

Equation:

$$r_n = 4\pi\epsilon_0 \frac{\hbar^2}{m_e e^2} n^2.$$

Note that these results tell us that the electron's speed as well as the radius of its orbit depend only on the index n that enumerates the orbit because all other quantities in the preceding equations are fundamental constants. We see from [\[link\]](#) that the size of the orbit grows as the square of n . This means that the second orbit is four times as large as the first orbit, and the third orbit is nine times as large as the first orbit, and so on. We also see from [\[link\]](#) that the electron's speed in the orbit decreases as the orbit size increases. The electron's speed is largest in the first Bohr orbit, for $n = 1$, which is the orbit closest to the nucleus. The radius of the first Bohr orbit is called the **Bohr radius of hydrogen**, denoted as a_0 . Its value is obtained by setting $n = 1$ in [\[link\]](#):

Note:

Equation:

$$a_0 = 4\pi\epsilon_0 \frac{\hbar^2}{m_e e^2} = 5.29 \times 10^{-11} \text{m} = 0.529 \text{\AA}.$$

We can substitute a_0 in [\[link\]](#) to express the radius of the n th orbit in terms of a_0 :

Note:

Equation:

$$r_n = a_0 n^2.$$

This result means that the electron orbits in hydrogen atom are *quantized* because the orbital radius takes on only specific values of $a_0, 4a_0, 9a_0, 16a_0, \dots$ given by [\[link\]](#), and no other values are allowed.

Electron Energies

The total energy E_n of an electron in the n th orbit is the sum of its kinetic energy K_n and its electrostatic potential energy U_n . Utilizing [\[link\]](#), we find that

Equation:

$$K_n = \frac{1}{2} m_e v_n^2 = \frac{1}{32\pi^2 \epsilon_0^2} \frac{m_e e^4}{\hbar^2} \frac{1}{n^2}.$$

Recall that the electrostatic potential energy of interaction between two charges q_1 and q_2 that are separated by a distance r_{12} is $(1/4\pi\epsilon_0)q_1q_2/r_{12}$. Here, $q_1 = +e$ is the charge of the nucleus in the hydrogen atom (the charge of the proton), $q_2 = -e$ is the charge of the electron and $r_{12} = r_n$ is the radius of the n th orbit. Now we use [\[link\]](#) to find the potential energy of the electron:

Equation:

$$U_n = -\frac{1}{4\pi\epsilon_0} \frac{e^2}{r_n} = -\frac{1}{16\pi^2 \epsilon_0^2} \frac{m_e e^4}{\hbar^2} \frac{1}{n^2}.$$

The total energy of the electron is the sum of [\[link\]](#) and [\[link\]](#):

Equation:

$$E_n = K_n + U_n = -\frac{1}{32\pi^2 \epsilon_0^2} \frac{m_e e^4}{\hbar^2} \frac{1}{n^2}.$$

Note that the energy depends only on the index n because the remaining symbols in [\[link\]](#) are physical constants. The value of the constant factor in [\[link\]](#) is

Note:

Equation:

$$E_0 = \frac{1}{32\pi^2 \epsilon_0^2} \frac{m_e e^4}{\hbar^2} = \frac{1}{8\epsilon_0^2} \frac{m_e e^4}{h^2} = 2.17 \times 10^{-18} \text{ J} = 13.6 \text{ eV}.$$

It is convenient to express the electron's energy in the n th orbit in terms of this energy, as

Note:

Equation:

$$E_n = -E_0 \frac{1}{n^2}.$$

Now we can see that the electron energies in the hydrogen atom are *quantized* because they can have only discrete values of $-E_0, -E_0/4, -E_0/9, -E_0/16, \dots$ given by [\[link\]](#), and no other energy values are allowed. This set of allowed electron energies is called the **energy spectrum of hydrogen** ([\[link\]](#)). The index n that enumerates energy levels in Bohr's model is called the energy **quantum number**. We identify the energy of the electron inside the hydrogen atom with the energy of the hydrogen atom. Note that the smallest value of energy is obtained for $n = 1$, so the hydrogen atom cannot have energy smaller than that. This smallest value of the electron energy in the hydrogen atom is called the **ground state energy of the hydrogen atom** and its value is

Note:

Equation:

$$E_1 = -E_0 = -13.6 \text{ eV}.$$

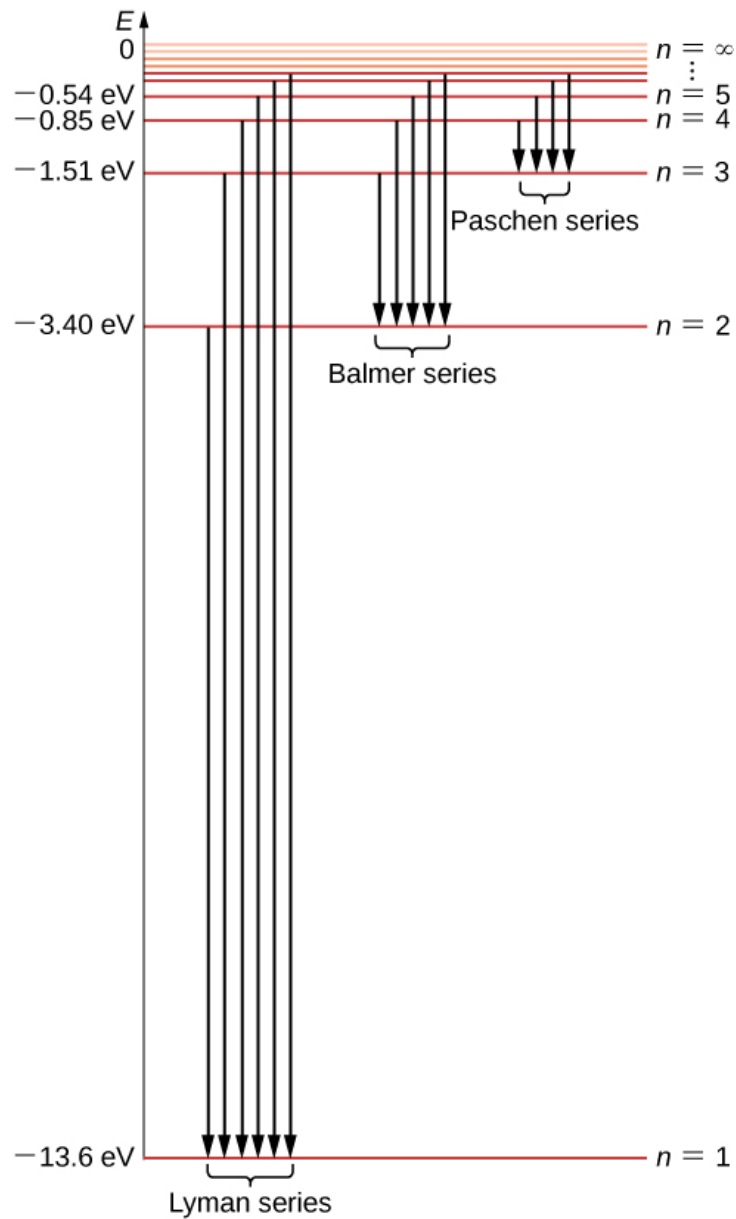
The hydrogen atom may have other energies that are higher than the ground state. These higher energy states are known as **excited energy states of a hydrogen atom**.

There is only one ground state, but there are infinitely many excited states because there are infinitely many values of n in [\[link\]](#). We say that the electron is in the "first excited state" when its energy is E_2 (when $n = 2$), the second excited state when its energy is E_3 (when $n = 3$) and, in general, in the n th excited state when its energy is E_{n+1} . There is no highest-of-all excited state; however, there is a limit to the sequence of excited states. If we keep increasing n in [\[link\]](#), we find that the limit is $-\lim_{n \rightarrow \infty} E_0/n^2 = 0$. In this limit, the electron is no longer bound to the nucleus but becomes a free electron. An electron remains bound in the hydrogen atom as long as its energy is negative. An electron that orbits the nucleus in the first Bohr orbit, closest to the nucleus, is in the ground state, where its energy has the smallest value. In the ground state, the electron is most strongly bound to the nucleus and its energy is given by [\[link\]](#). If we want to remove this electron from the atom, we must supply it with enough energy, E_∞ , to at least balance out its ground state energy E_1 :

Equation:

$$E_\infty + E_1 = 0 \Rightarrow E_\infty = -E_1 = -(-E_0) = E_0 = 13.6 \text{ eV}.$$

The energy that is needed to remove the electron from the atom is called the **ionization energy**. The ionization energy E_∞ that is needed to remove the electron from the first Bohr orbit is called the **ionization limit of the hydrogen atom**. The ionization limit in [\[link\]](#) that we obtain in Bohr's model agrees with experimental value.



The energy spectrum of the hydrogen atom. Energy levels (horizontal lines) represent the bound states of an electron in the atom. There is only one ground state, $n = 1$, and infinite quantized excited states. The states are enumerated by the quantum number $n = 1, 2, 3, 4, \dots$. Vertical lines illustrate the allowed electron transitions between the states. Downward arrows illustrate transitions with an emission of a photon with a wavelength in the indicated spectral band.

Spectral Emission Lines of Hydrogen

To obtain the wavelengths of the emitted radiation when an electron makes a transition from the n th orbit to the m th orbit, we use the second of Bohr's quantization conditions and [\[link\]](#) for energies. The emission of energy from the atom can occur only when an electron makes a transition from an excited state to a lower-energy state. In the course of such a transition, the emitted photon carries away the difference of energies between the states involved in the transition. The transition cannot go in the other direction because the energy of a photon cannot be negative, which means that for emission we must have $E_n > E_m$ and $n > m$. Therefore, the third of Bohr's postulates gives

Equation:

$$hf = |E_n - E_m| = E_n - E_m = -E_0 \frac{1}{n^2} + E_0 \frac{1}{m^2} = E_0 \left(\frac{1}{m^2} - \frac{1}{n^2} \right).$$

Now we express the photon's energy in terms of its wavelength, $hf = hc/\lambda$, and divide both sides of [\[link\]](#) by hc . The result is

Equation:

$$\frac{1}{\lambda} = \frac{E_0}{hc} \left(\frac{1}{m^2} - \frac{1}{n^2} \right).$$

The value of the constant in this equation is

Equation:

$$\frac{E_0}{hc} = \frac{13.6 \text{ eV}}{(4.136 \times 10^{-15} \text{ eV} \cdot \text{s})(2.997 \times 10^8 \text{ m/s})} = 1.097 \times 10^7 \frac{1}{\text{m}}.$$

This value is exactly the Rydberg constant R_H in the Rydberg heuristic formula [\[link\]](#). In fact, [\[link\]](#) is identical to the Rydberg formula, because for a given m , we have $n = m + 1, m + 2, \dots$. In this way, the Bohr quantum model of the hydrogen atom allows us to derive the experimental Rydberg constant from first principles and to express it in terms of fundamental constants. Transitions between the allowed electron orbits are illustrated in [\[link\]](#).

We can repeat the same steps that led to [\[link\]](#) to obtain the wavelength of the absorbed radiation; this again gives [\[link\]](#) but this time for the positions of absorption lines in the absorption spectrum of hydrogen. The only difference is that for absorption, the quantum number m is the index of the orbit occupied by the electron before the transition (lower-energy orbit) and the quantum number n is the index of the orbit to which the electron makes the transition (higher-energy orbit). The difference between the electron energies in these two orbits is the energy of the absorbed photon.

Example:

Size and Ionization Energy of the Hydrogen Atom in an Excited State

If a hydrogen atom in the ground state absorbs a 93.7-nm photon, corresponding to a transition line in the Lyman series, how does this affect the atom's energy and size? How much energy is needed to ionize the atom when it is in this excited state? Give your answers in absolute units, and relative to the ground state.

Strategy

Before the absorption, the atom is in its ground state. This means that the electron transition takes place from the orbit $m = 1$ to some higher n th orbit. First, we must determine n for the absorbed wavelength $\lambda = 93.7$ nm. Then, we can use [\[link\]](#) to find the energy E_n of the excited state and its ionization energy $E_{\infty,n}$, and use [\[link\]](#) to find the radius r_n of the atom in the excited state. To estimate n , we use [\[link\]](#).

Solution

Substitute $m = 1$ and $\lambda = 93.7$ nm in [\[link\]](#) and solve for n . You should not expect to obtain a perfect integer answer because of rounding errors, but your answer will be close to an integer, and you can estimate n by taking the integral part of your answer:

Equation:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{1^2} - \frac{1}{n^2} \right) \Rightarrow n = \frac{1}{\sqrt{1 - \frac{1}{\lambda R_H}}} = \frac{1}{\sqrt{1 - \frac{1}{(93.7 \times 10^{-9} \text{ m})(1.097 \times 10^7 \text{ m}^{-1})}}} = 6.07 \Rightarrow n = 6.$$

The radius of the $n = 6$ orbit is

Equation:

$$r_n = a_0 n^2 = a_0 6^2 = 36a_0 = 36(0.529 \times 10^{-10} \text{ m}) = 19.04 \times 10^{-10} \text{ m} \cong 19.0 \text{ \AA}.$$

Thus, after absorbing the 93.7-nm photon, the size of the hydrogen atom in the excited $n = 6$ state is 36 times larger than before the absorption, when the atom was in the ground state. The energy of the fifth excited state ($n = 6$) is:

Equation:

$$E_n = -\frac{E_0}{n^2} = -\frac{E_0}{6^2} = -\frac{E_0}{36} = -\frac{13.6 \text{ eV}}{36} \cong -0.378 \text{ eV}.$$

After absorbing the 93.7-nm photon, the energy of the hydrogen atom is larger than it was before the absorption. Ionization of the atom when it is in the fifth excited state ($n = 6$) requires 36 times less energy than is needed when the atom is in the ground state:

Equation:

$$E_{\infty,6} = -E_6 = -(-0.378 \text{ eV}) = 0.378 \text{ eV}.$$

Significance

We can analyze any spectral line in the spectrum of hydrogen in the same way. Thus, the experimental measurements of spectral lines provide us with information about the atomic structure of the hydrogen atom.

Note:**Exercise:****Problem:**

Check Your Understanding When an electron in a hydrogen atom is in the first excited state, what prediction does the Bohr model give about its orbital speed and kinetic energy? What is the magnitude of its orbital angular momentum?

Solution:

$$v_2 = 1.1 \times 10^6 \text{ m/s} \cong 0.0036c; L_2 = 2\hbar \quad K_2 = 3.4 \text{ eV}$$

Bohr's model of the hydrogen atom also correctly predicts the spectra of some hydrogen-like ions.

Hydrogen-like ions are atoms of elements with an atomic number Z larger than one ($Z = 1$ for hydrogen) but with all electrons removed except one. For example, an electrically neutral helium atom has an atomic number $Z = 2$. This means it has two electrons orbiting the nucleus with a charge of $q = +Ze$. When one of the orbiting electrons is removed from the helium atom (we say, when the helium atom is singly ionized), what remains is a hydrogen-like atomic structure where the remaining electron orbits the nucleus with a charge of $q = +Ze$. This type of situation is described by the Bohr model. Assuming that the charge of the nucleus is not $+e$ but $+Ze$, we can repeat all steps, beginning with [\[link\]](#), to obtain the results for a hydrogen-like ion:

Note:**Equation:**

$$r_n = \frac{a_0}{Z} n^2$$

where a_0 is the Bohr orbit of hydrogen, and

Note:**Equation:**

$$E_n = -Z^2 E_0 \frac{1}{n^2}$$

where E_0 is the ionization limit of a hydrogen atom. These equations are good approximations as long as the atomic number Z is not too large.

The Bohr model is important because it was the first model to postulate the quantization of electron orbits in atoms. Thus, it represents an early quantum theory that gave a start to developing modern quantum theory. It introduced the concept of a quantum number to describe atomic states. The limitation of the early quantum theory is that it cannot describe atoms in which the number of electrons orbiting the nucleus is larger than one. The Bohr model of hydrogen is a semi-classical model because it combines the classical concept of electron orbits with the new concept of quantization. The remarkable success of this model prompted many physicists to seek an explanation for why such a model should work at all, and to seek an understanding of the physics behind the postulates of early quantum theory. This search brought about the onset of an entirely new concept of "matter waves."

Summary

- Positions of absorption and emission lines in the spectrum of atomic hydrogen are given by the experimental Rydberg formula. Classical physics cannot explain the spectrum of atomic hydrogen.
- The Bohr model of hydrogen was the first model of atomic structure to correctly explain the radiation spectra of atomic hydrogen. It was preceded by the Rutherford nuclear model of the atom. In Rutherford's model, an atom consists of a positively charged point-like nucleus that contains almost the entire mass of the atom and of negative electrons that are located far away from the nucleus.
- Bohr's model of the hydrogen atom is based on three postulates: (1) an electron moves around the nucleus in a circular orbit, (2) an electron's angular momentum in the orbit is quantized, and (3) the change in an electron's energy as it makes a quantum jump from one orbit to another is always accompanied by the emission or absorption of a photon. Bohr's model is semi-classical because it combines the classical concept of electron orbit (postulate 1) with the new concept of quantization (postulates 2 and 3).
- Bohr's model of the hydrogen atom explains the emission and absorption spectra of atomic hydrogen and hydrogen-like ions with low atomic numbers. It was the first model to introduce the concept of a quantum number to describe atomic states and to postulate quantization of electron orbits in the atom. Bohr's model is an important step in the development of quantum mechanics, which deals with many-electron atoms.

Conceptual Questions

Exercise:

Problem:

Explain why the patterns of bright emission spectral lines have an identical spectral position to the pattern of dark absorption spectral lines for a given gaseous element.

Exercise:

Problem: Do the various spectral lines of the hydrogen atom overlap?

Solution:

no

Exercise:

Problem:

The Balmer series for hydrogen was discovered before either the Lyman or the Paschen series. Why?

Exercise:

Problem:

When the absorption spectrum of hydrogen at room temperature is analyzed, absorption lines for the Lyman series are found, but none are found for the Balmer series. What does this tell us about the energy state of most hydrogen atoms at room temperature?

Solution:

They are at ground state.

Exercise:

Problem:

Hydrogen accounts for about 75% by mass of the matter at the surfaces of most stars. However, the absorption lines of hydrogen are strongest (of highest intensity) in the spectra of stars with a surface temperature of about 9000 K. They are weaker in the sun spectrum and are essentially nonexistent in very hot (temperatures above 25,000 K) or rather cool (temperatures below 3500 K) stars. Speculate as to why surface temperature affects the hydrogen absorption lines that we observe.

Exercise:

Problem:

Discuss the similarities and differences between Thomson's model of the hydrogen atom and Bohr's model of the hydrogen atom.

Solution:

Answers may vary

Exercise:

Problem:

Discuss the way in which Thomson's model is nonphysical. Support your argument with experimental evidence.

Exercise:

Problem:

If, in a hydrogen atom, an electron moves to an orbit with a larger radius, does the energy of the hydrogen atom increase or decrease?

Solution:

increase

Exercise:

Problem:

How is the energy conserved when an atom makes a transition from a higher to a lower energy state?

Exercise:

Problem:

Suppose an electron in a hydrogen atom makes a transition from the $(n+1)$ th orbit to the n th orbit. Is the wavelength of the emitted photon longer for larger values of n , or for smaller values of n ?

Solution:

for larger n

Exercise:

Problem: Discuss why the allowed energies of the hydrogen atom are negative.

Exercise:

Problem: Can a hydrogen atom absorb a photon whose energy is greater than 13.6 eV?

Solution:

Yes, the excess of 13.6 eV will become kinetic energy of a free electron.

Exercise:

Problem: Why can you see through glass but not through wood?

Exercise:

Problem: Do gravitational forces have a significant effect on atomic energy levels?

Solution:

no

Exercise:

Problem: Show that Planck's constant has the dimensions of angular momentum.

Problems

Exercise:

Problem:

Calculate the wavelength of the first line in the Lyman series and show that this line lies in the ultraviolet part of the spectrum.

Solution:

121.5 nm

Exercise:

Problem:

Calculate the wavelength of the fifth line in the Lyman series and show that this line lies in the ultraviolet part of the spectrum.

Exercise:

Problem:

Calculate the energy changes corresponding to the transitions of the hydrogen atom: (a) from $n = 3$ to $n = 4$; (b) from $n = 2$ to $n = 1$; and (c) from $n = 3$ to $n = \infty$.

Solution:

a. 0.661 eV; b. -10.2 eV; c. 1.511 eV

Exercise:

Problem: Determine the wavelength of the third Balmer line (transition from $n = 5$ to $n = 2$).

Exercise:**Problem:**

What is the frequency of the photon absorbed when the hydrogen atom makes the transition from the ground state to the $n = 4$ state?

Solution:

3038 THz

Exercise:**Problem:**

When a hydrogen atom is in its ground state, what are the shortest and longest wavelengths of the photons it can absorb without being ionized?

Exercise:**Problem:**

When a hydrogen atom is in its third excited state, what are the shortest and longest wavelengths of the photons it can emit?

Solution:

97.33 nm

Exercise:**Problem:**

What is the longest wavelength that light can have if it is to be capable of ionizing the hydrogen atom in its ground state?

Exercise:**Problem:**

For an electron in a hydrogen atom in the $n = 2$ state, compute: (a) the angular momentum; (b) the kinetic energy; (c) the potential energy; and (d) the total energy.

Solution:

a. h/π ; b. 3.4 eV; c. - 6.8 eV; d. - 3.4 eV

Exercise:

Problem: Find the ionization energy of a hydrogen atom in the fourth energy state.

Exercise:

Problem:

It has been measured that it required 0.850 eV to remove an electron from the hydrogen atom. In what state was the atom before the ionization happened?

Solution:

$$n = 4$$

Exercise:

Problem: What is the radius of a hydrogen atom when the electron is in the first excited state?

Exercise:

Problem:

Find the shortest wavelength in the Balmer series. In what part of the spectrum does this line lie?

Solution:

365 nm; UV

Exercise:

Problem: Show that the entire Paschen series lies in the infrared part of the spectrum.

Exercise:

Problem:

Do the Balmer series and the Lyman series overlap? Why? Why not? (Hint: calculate the shortest Balmer line and the longest Lyman line.)

Solution:

no

Exercise:

Problem:

(a) Which line in the Balmer series is the first one in the UV part of the spectrum? (b) How many Balmer lines lie in the visible part of the spectrum? (c) How many Balmer lines lie in the UV?

Exercise:

Problem:

A 4.653- μm emission line of atomic hydrogen corresponds to transition between the states $n_f = 5$ and n_i . Find n_i .

Solution:

7

Glossary

absorption spectrum

wavelengths of absorbed radiation by atoms and molecules

α -particle

doubly ionized helium atom

α -ray

beam of α -particles (alpha-particles)

Balmer formula

describes the emission spectrum of a hydrogen atom in the visible-light range

Balmer series

spectral lines corresponding to electron transitions to/from the $n = 2$ state of the hydrogen atom, described by the Balmer formula

β -ray

beam of electrons

Bohr radius of hydrogen

radius of the first Bohr's orbit

Bohr's model of the hydrogen atom

first quantum model to explain emission spectra of hydrogen

Brackett series

spectral lines corresponding to electron transitions to/from the $n = 4$ state

emission spectrum

wavelengths of emitted radiation by atoms and molecules

energy spectrum of hydrogen

set of allowed discrete energies of an electron in a hydrogen atom

excited energy states of the H atom

energy state other than the ground state

Fraunhofer lines

dark absorption lines in the continuum solar emission spectrum

γ-ray

beam of highly energetic photons

ground state energy of the hydrogen atom

energy of an electron in the first Bohr orbit of the hydrogen atom

Humphreys series

spectral lines corresponding to electron transitions to/from the $n = 6$ state

hydrogen-like atom

ionized atom with one electron remaining and nucleus with charge $+Ze$

ionization energy

energy needed to remove an electron from an atom

ionization limit of the hydrogen atom

ionization energy needed to remove an electron from the first Bohr orbit

Lyman series

spectral lines corresponding to electron transitions to/from the ground state

nuclear model of the atom

heavy positively charged nucleus at the center is surrounded by electrons, proposed by Rutherford

Paschen series

spectral lines corresponding to electron transitions to/from the $n = 3$ state

Pfund series

spectral lines corresponding to electron transitions to/from the $n = 5$ state

postulates of Bohr's model

three assumptions that set a frame for Bohr's model

quantum number

index that enumerates energy levels

Rutherford's gold foil experiment

first experiment to demonstrate the existence of the atomic nucleus

Rydberg constant for hydrogen

physical constant in the Balmer formula

Rydberg formula

experimentally found positions of spectral lines of hydrogen atom

De Broglie's Matter Waves

By the end of this section, you will be able to:

- Describe de Broglie's hypothesis of matter waves
- Explain how the de Broglie's hypothesis gives the rationale for the quantization of angular momentum in Bohr's quantum theory of the hydrogen atom
- Describe the Davisson–Germer experiment
- Interpret de Broglie's idea of matter waves and how they account for electron diffraction phenomena

Compton's formula established that an electromagnetic wave can behave like a particle of light when interacting with matter. In 1924, Louis de Broglie proposed a new speculative hypothesis that electrons and other particles of matter can behave like waves. Today, this idea is known as **de Broglie's hypothesis of matter waves**. In 1926, De Broglie's hypothesis, together with Bohr's early quantum theory, led to the development of a new theory of **wave quantum mechanics** to describe the physics of atoms and subatomic particles. Quantum mechanics has paved the way for new engineering inventions and technologies, such as the laser and magnetic resonance imaging (MRI). These new technologies drive discoveries in other sciences such as biology and chemistry.

According to de Broglie's hypothesis, massless photons as well as massive particles must satisfy one common set of relations that connect the energy E with the frequency f , and the linear momentum p with the wavelength λ . We have discussed these relations for photons in the context of Compton's effect. We are recalling them now in a more general context. Any particle that has energy and momentum is a **de Broglie wave** of frequency f and wavelength λ :

Note:
Equation:

$$E = hf$$

Note:
Equation:

$$\lambda = \frac{h}{p}.$$

Here, E and p are, respectively, the relativistic energy and the momentum of a particle. De Broglie's relations are usually expressed in terms of the wave vector \vec{k} , $k = 2\pi / \lambda$, and the wave frequency $\omega = 2\pi f$, as we usually do for waves:

Equation:

$$E = \hbar\omega$$

Equation:

$$\vec{p} = \hbar\vec{k}.$$

Wave theory tells us that a wave carries its energy with the **group velocity**. For matter waves, this group velocity is the velocity u of the particle. Identifying the energy E and momentum p of a particle with its relativistic energy mc^2 and its relativistic momentum mu , respectively, it follows from de Broglie relations that matter waves satisfy the following relation:

Note:

Equation:

$$\lambda f = \frac{\omega}{k} = \frac{E/\hbar}{p/\hbar} = \frac{E}{p} = \frac{mc^2}{mu} = \frac{c^2}{u} = \frac{c}{\beta}$$

where $\beta = u/c$. When a particle is massless we have $u = c$ and [\[link\]](#) becomes $\lambda f = c$.

Example:

How Long Are de Broglie Matter Waves?

Calculate the de Broglie wavelength of: (a) a 0.65-kg basketball thrown at a speed of 10 m/s, (b) a nonrelativistic electron with a kinetic energy of 1.0 eV, and (c) a relativistic electron with a kinetic energy of 108 keV.

Strategy

We use [\[link\]](#) to find the de Broglie wavelength. When the problem involves a nonrelativistic object moving with a nonrelativistic speed u , such as in (a) when $\beta = u/c \ll 1$, we use nonrelativistic momentum p . When the nonrelativistic approximation cannot be used, such as in (c), we must use the relativistic momentum $p = mu = m_0\gamma u = E_0\gamma\beta/c$, where the rest mass energy of a particle is $E_0 = mc^2$ and γ is the Lorentz factor $\gamma = 1/\sqrt{1-\beta^2}$. The total energy E of a particle is given by [\[link\]](#) and the kinetic energy is $K = E - E_0 = (\gamma - 1)E_0$. When the kinetic energy is known, we can invert [\[link\]](#) to find the momentum

$p = \sqrt{(E^2 - E_0^2)/c^2} = \sqrt{K(K + 2E_0)}/c$ and substitute in [\[link\]](#) to obtain

Equation:

$$\lambda = \frac{h}{p} = \frac{hc}{\sqrt{K(K + 2E_0)}}.$$

Depending on the problem at hand, in this equation we can use the following values for hc :

$$hc = (6.626 \times 10^{-34} \text{ J} \cdot \text{s})(2.998 \times 10^8 \text{ m/s}) = 1.986 \times 10^{-25} \text{ J} \cdot \text{m} = 1.241 \text{ eV} \cdot \mu\text{m}$$

Solution

- a. For the basketball, the kinetic energy is

Equation:

$$K = mu^2 / 2 = (0.65 \text{ kg})(10 \text{ m/s})^2 / 2 = 32.5 \text{ J}$$

and the rest mass energy is

Equation:

$$E_0 = mc^2 = (0.65 \text{ kg})(2.998 \times 10^8 \text{ m/s})^2 = 5.84 \times 10^{16} \text{ J}.$$

We see that $K / (K + E_0) \ll 1$ and use $p = mu = (0.65 \text{ kg})(10 \text{ m/s}) = 6.5 \text{ J} \cdot \text{s/m}$:

Equation:

$$\lambda = \frac{h}{p} = \frac{6.626 \times 10^{-34} \text{ J} \cdot \text{s}}{6.5 \text{ J} \cdot \text{s/m}} = 1.02 \times 10^{-34} \text{ m}.$$

- b. For the nonrelativistic electron,

Equation:

$$E_0 = mc^2 = (9.109 \times 10^{-31} \text{ kg})(2.998 \times 10^8 \text{ m/s})^2 = 511 \text{ keV}$$

and when $K = 1.0 \text{ eV}$, we have $K / (K + E_0) = (1 / 512) \times 10^{-3} \ll 1$, so we can use the nonrelativistic formula. However, it is simpler here to use [\[link\]](#):

Equation:

$$\lambda = \frac{h}{p} = \frac{hc}{\sqrt{K(K + 2E_0)}} = \frac{1.241 \text{ eV} \cdot \mu\text{m}}{\sqrt{(1.0 \text{ eV})[1.0 \text{ eV} + 2(511 \text{ keV})]}} = 1.23 \text{ nm}.$$

If we use nonrelativistic momentum, we obtain the same result because 1 eV is much smaller than the rest mass of the electron.

- c. For a fast electron with $K = 108 \text{ keV}$, relativistic effects cannot be neglected because its total energy is $E = K + E_0 = 108 \text{ keV} + 511 \text{ keV} = 619 \text{ keV}$ and $K / E = 108 / 619$ is not negligible:

Equation:

$$\lambda = \frac{h}{p} = \frac{hc}{\sqrt{K(K + 2E_0)}} = \frac{1.241 \text{ eV} \cdot \mu\text{m}}{\sqrt{108 \text{ keV}[108 \text{ keV} + 2(511 \text{ keV})]}} = 3.55 \text{ pm}.$$

Significance

We see from these estimates that De Broglie's wavelengths of macroscopic objects such as a ball are immeasurably small. Therefore, even if they exist, they are not detectable and do not affect the motion of macroscopic objects.

Note:

Exercise:

Problem:

Check Your Understanding What is de Broglie's wavelength of a nonrelativistic proton with a kinetic energy of 1.0 eV?

Solution:

29 pm

Using the concept of the electron matter wave, de Broglie provided a rationale for the quantization of the electron's angular momentum in the hydrogen atom, which was postulated in Bohr's quantum theory. The physical explanation for the first Bohr quantization condition comes naturally when we assume that an electron in a hydrogen atom behaves not like a particle but like a wave. To see it clearly, imagine a stretched guitar string that is clamped at both ends and vibrates in one of its normal modes. If the length of the string is l ([link](#)), the wavelengths of these vibrations cannot be arbitrary but must be such that an integer k number of half-wavelengths $\lambda/2$ fit exactly on the distance l between the ends. This is the condition $l = k\lambda/2$ for a standing wave on a string. Now suppose that instead of having the string clamped at the walls, we bend its length into a circle and fasten its ends to each other. This produces a circular string that vibrates in normal modes, satisfying the same standing-wave condition, but the number of half-wavelengths must now be an even number k , $k = 2n$, and the length l is now connected to the radius r_n of the circle. This means that the radii are not arbitrary but must satisfy the following standing-wave condition:

Equation:

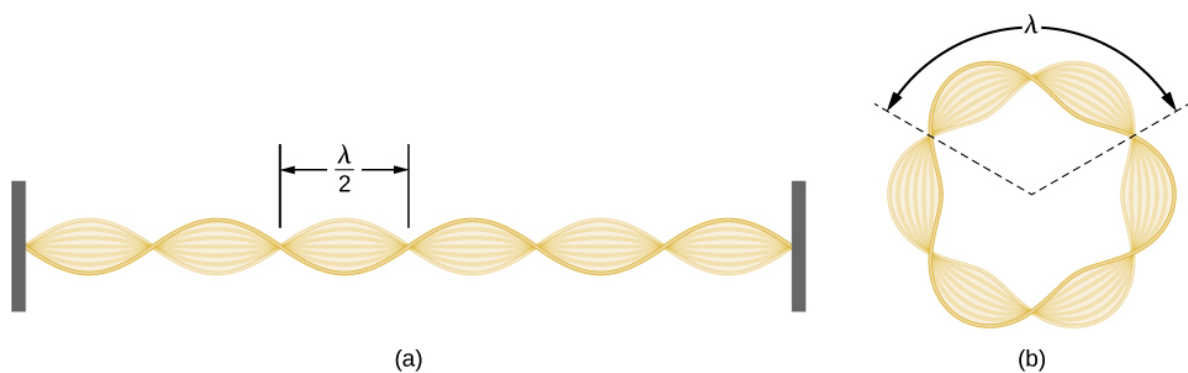
$$2\pi r_n = 2n \frac{\lambda}{2}.$$

If an electron in the n th Bohr orbit moves as a wave, by [link](#) its wavelength must be equal to $\lambda = 2\pi r_n / n$. Assuming that [link](#) is valid, the electron wave of this wavelength corresponds to the electron's linear momentum, $p = h / \lambda = nh / (2\pi r_n) = n\hbar / r_n$. In a circular orbit, therefore, the electron's angular momentum must be

Equation:

$$L_n = r_n p = r_n \frac{n\hbar}{r_n} = n\hbar.$$

This equation is the first of Bohr's quantization conditions, given by [link](#). Providing a physical explanation for Bohr's quantization condition is a convincing theoretical argument for the existence of matter waves.



Standing-wave pattern: (a) a stretched string clamped at the walls; (b) an electron wave trapped in the third Bohr orbit in the hydrogen atom.

Example:

The Electron Wave in the Ground State of Hydrogen

Find the de Broglie wavelength of an electron in the ground state of hydrogen.

Strategy

We combine the first quantization condition in [\[link\]](#) with [\[link\]](#) and use [\[link\]](#) for the first Bohr radius with $n = 1$.

Solution

When $n = 1$ and $r_n = a_0 = 0.529 \text{ \AA}$, the Bohr quantization condition gives $a_0 p = 1 \cdot \hbar \Rightarrow p = \hbar / a_0$. The electron wavelength is:

Equation:

$$\lambda = h / p = h / \hbar / a_0 = 2\pi a_0 = 2\pi(0.529 \text{ \AA}) = 3.324 \text{ \AA}.$$

Significance

We obtain the same result when we use [\[link\]](#) directly.

Note:

Exercise:

Problem:

Check Your Understanding Find the de Broglie wavelength of an electron in the third excited state of hydrogen.

Solution:

$$\lambda = 2\pi n a_0 = 2(3.324 \text{ \AA}) = 6.648 \text{ \AA}$$

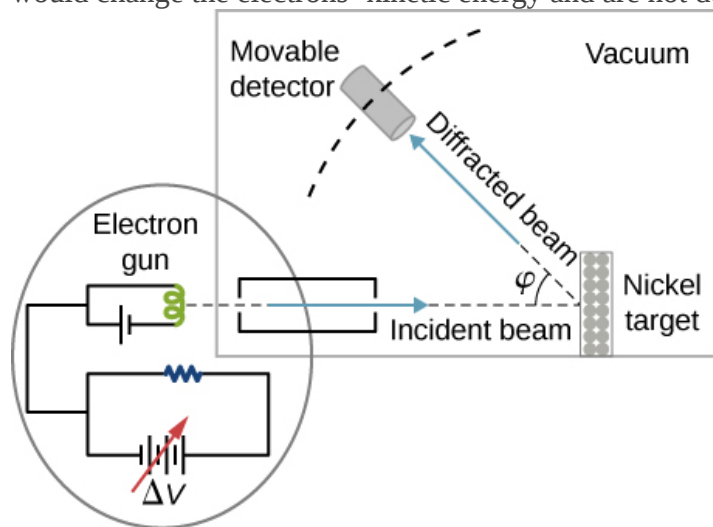
Experimental confirmation of matter waves came in 1927 when C. Davisson and L. Germer performed a series of electron-scattering experiments that clearly showed that electrons do behave like waves. Davisson and Germer did not set up their experiment to confirm de Broglie's hypothesis: The confirmation came as a byproduct of their routine experimental studies of metal surfaces under electron bombardment.

In the particular experiment that provided the very first evidence of electron waves (known today as the **Davisson–Germer experiment**), they studied a surface of nickel. Their nickel sample was specially prepared in a high-temperature oven to change its usual polycrystalline structure to a form in which large single-crystal domains occupy the volume. [\[link\]](#) shows the experimental setup. Thermal electrons are released from a heated element (usually made of tungsten) in the electron gun and accelerated through a potential difference ΔV , becoming a well-collimated beam of electrons produced by an electron gun. The kinetic energy K of the electrons is adjusted by selecting a value of the potential difference in the electron gun. This produces a beam of electrons with a set value of linear momentum, in accordance with the conservation of energy:

Equation:

$$e\Delta V = K = \frac{p^2}{2m} \Rightarrow p = \sqrt{2me\Delta V}.$$

The electron beam is incident on the nickel sample in the direction normal to its surface. At the surface, it scatters in various directions. The intensity of the beam scattered in a selected direction φ is measured by a highly sensitive detector. The detector's angular position with respect to the direction of the incident beam can be varied from $\varphi = 0^\circ$ to $\varphi = 90^\circ$. The entire setup is enclosed in a vacuum chamber to prevent electron collisions with air molecules, as such thermal collisions would change the electrons' kinetic energy and are not desirable.



Schematics of the experimental setup of the Davisson–Germer diffraction experiment. A well-collimated beam of electrons is scattered off the nickel target. The kinetic energy of electrons in the incident beam is selected by adjusting a variable potential, ΔV , in the electron gun. Intensity of the

scattered electron beam is measured for a range of scattering angles φ , whereas the distance between the detector and the target does not change.

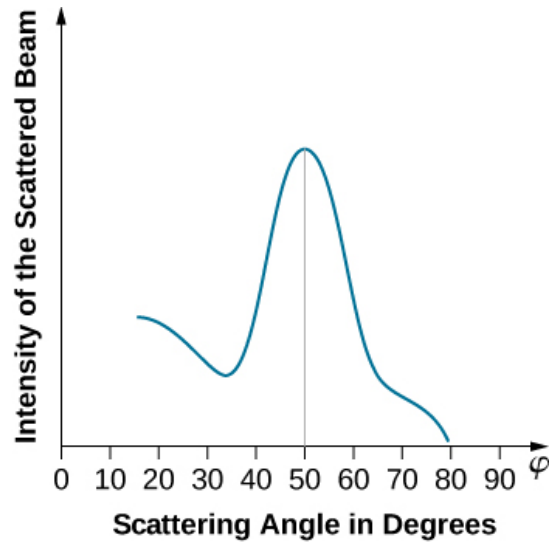
When the nickel target has a polycrystalline form with many randomly oriented microscopic crystals, the incident electrons scatter off its surface in various random directions. As a result, the intensity of the scattered electron beam is much the same in any direction, resembling a diffuse reflection of light from a porous surface. However, when the nickel target has a regular crystalline structure, the intensity of the scattered electron beam shows a clear maximum at a specific angle and the results show a clear diffraction pattern (see [\[link\]](#)). Similar diffraction patterns formed by X-rays scattered by various crystalline solids were studied in 1912 by father-and-son physicists William H. Bragg and William L. Bragg. The Bragg law in X-ray crystallography provides a connection between the wavelength λ of the radiation incident on a crystalline lattice, the lattice spacing, and the position of the interference maximum in the diffracted radiation (see [Diffraction](#)).

The lattice spacing of the Davisson–Germer target, determined with X-ray crystallography, was measured to be $a = 2.15 \text{ \AA}$. Unlike X-ray crystallography in which X-rays penetrate the sample, in the original Davisson–Germer experiment, only the surface atoms interact with the incident electron beam. For the surface diffraction, the maximum intensity of the reflected electron beam is observed for scattering angles that satisfy the condition $n\lambda = a\sin\varphi$ (see [\[link\]](#)). The first-order maximum (for $n = 1$) is measured at a scattering angle of $\varphi \approx 50^\circ$ at $\Delta V \approx 54\text{V}$, which gives the wavelength of the incident radiation as $\lambda = (2.15 \text{ \AA})\sin 50^\circ = 1.64 \text{ \AA}$. On the other hand, a 54-V potential accelerates the incident electrons to kinetic energies of $K = 54 \text{ eV}$. Their momentum, calculated from [\[link\]](#), is $p = 2.478 \times 10^{-5} \text{ eV} \cdot \text{s/m}$. When we substitute this result in [\[link\]](#), the de Broglie wavelength is obtained as

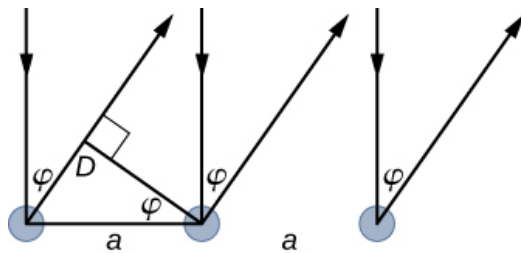
Equation:

$$\lambda = \frac{h}{p} = \frac{4.136 \times 10^{-15} \text{ eV} \cdot \text{s}}{2.478 \times 10^{-5} \text{ eV} \cdot \text{s/m}} = 1.67 \text{ \AA}.$$

The same result is obtained when we use $K = 54 \text{ eV}$ in [\[link\]](#). The proximity of this theoretical result to the Davisson–Germer experimental value of $\lambda = 1.64 \text{ \AA}$ is a convincing argument for the existence of de Broglie matter waves.



The experimental results of electron diffraction on a nickel target for the accelerating potential in the electron gun of about $\Delta V = 54\text{V}$: The intensity maximum is registered at the scattering angle of about $\varphi = 50^\circ$.



$$D = a \sin \varphi$$

$$D = n \lambda \quad n = 1, 2, 3, \dots$$

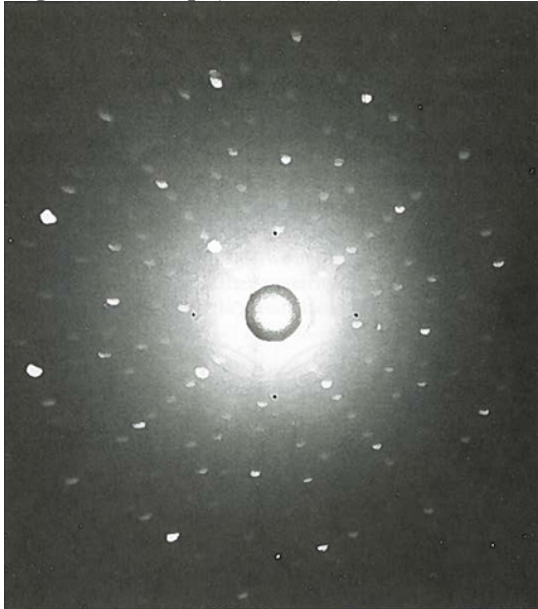
$$n\lambda = a \sin \varphi$$

In the surface diffraction of a monochromatic electromagnetic wave on a crystalline lattice structure, the in-phase incident beams are reflected from atoms on the surface. A ray reflected from the left atom travels an additional distance $D = a \sin \varphi$ to the detector, where a is the lattice spacing. The reflected beams remain in-phase when D is an integer multiple of their wavelength λ . The intensity of

the reflected waves has pronounced maxima for angles φ satisfying

$$n\lambda = a\sin\varphi.$$

Diffraction lines measured with low-energy electrons, such as those used in the Davisson–Germer experiment, are quite broad (see [\[link\]](#)) because the incident electrons are scattered only from the surface. The resolution of diffraction images greatly improves when a higher-energy electron beam passes through a thin metal foil. This occurs because the diffraction image is created by scattering off many crystalline planes inside the volume, and the maxima produced in scattering at Bragg angles are sharp (see [\[link\]](#)).



(a)



(b)

Diffraction patterns obtained in scattering on a crystalline solid: (a) with X-rays, and (b) with electrons. The observed pattern reflects the symmetry of the crystalline structure of the sample.

Since the work of Davisson and Germer, de Broglie's hypothesis has been extensively tested with various experimental techniques, and the existence of de Broglie waves has been confirmed for numerous elementary particles. Neutrons have been used in scattering experiments to determine crystalline structures of solids from interference patterns formed by neutron matter waves. The neutron has zero charge and its mass is comparable with the mass of a positively charged proton. Both neutrons and protons can be seen as matter waves. Therefore, the property of being a matter wave is not specific to electrically charged particles but is true of all particles in motion. Matter waves of molecules as large as carbon C_{60} have been measured. All physical objects, small or large, have an associated matter wave as long as they remain in motion. The universal character of de Broglie matter waves is firmly established.

Example:**Neutron Scattering**

Suppose that a neutron beam is used in a diffraction experiment on a typical crystalline solid. Estimate the kinetic energy of a neutron (in eV) in the neutron beam and compare it with kinetic energy of an ideal gas in equilibrium at room temperature.

Strategy

We assume that a typical crystal spacing a is of the order of 1.0 \AA . To observe a diffraction pattern on such a lattice, the neutron wavelength λ must be on the same order of magnitude as the lattice spacing. We use [\[link\]](#) to find the momentum p and kinetic energy K . To compare this energy with the energy E_T of ideal gas in equilibrium at room temperature $T = 300\text{K}$, we use the relation $K = \frac{3}{2}k_B T$, where $k_B = 8.62 \times 10^{-5} \text{ eV/K}$ is the Boltzmann constant.

Solution

We evaluate pc to compare it with the neutron's rest mass energy $E_0 = 940 \text{ MeV}$:

Equation:

$$p = \frac{h}{\lambda} \Rightarrow pc = \frac{hc}{\lambda} = \frac{1.241 \times 10^{-6} \text{ eV} \cdot \text{m}}{10^{-10} \text{ m}} = 12.41 \text{ keV}.$$

We see that $p^2 c^2 \ll E_0^2$ so $K \ll E_0$ and we can use the nonrelativistic kinetic energy:

Equation:

$$K = \frac{p^2}{2m_n} = \frac{h^2}{2\lambda^2 m_n} = \frac{(6.63 \times 10^{-34} \text{ J} \cdot \text{s})^2}{(2 \times 10^{-20} \text{ m}^2)(1.66 \times 10^{-27} \text{ kg})} = 1.32 \times 10^{-20} \text{ J} = 82.7 \text{ meV}.$$

Kinetic energy of ideal gas in equilibrium at 300 K is:

Equation:

$$K_T = \frac{3}{2} k_B T = \frac{3}{2} (8.62 \times 10^{-5} \text{ eV/K})(300\text{K}) = 38.8 \text{ MeV}.$$

We see that these energies are of the same order of magnitude.

Significance

Neutrons with energies in this range, which is typical for an ideal gas at room temperature, are called “thermal neutrons.”

Example:**Wavelength of a Relativistic Proton**

In a supercollider at CERN, protons can be accelerated to velocities of $0.75c$. What are their de Broglie wavelengths at this speed? What are their kinetic energies?

Strategy

The rest mass energy of a proton is

$E_0 = m_0 c^2 = (1.672 \times 10^{-27} \text{ kg})(2.998 \times 10^8 \text{ m/s})^2 = 938 \text{ MeV}$. When the proton's velocity is known, we have $\beta = 0.75$ and $\beta\gamma = 0.75 / \sqrt{1 - 0.75^2} = 1.134$. We obtain the wavelength λ and kinetic energy K from relativistic relations.

Solution

Equation:

$$\lambda = \frac{h}{p} = \frac{hc}{pc} = \frac{hc}{\beta\gamma E_0} = \frac{1.241 \text{ eV} \cdot \mu\text{m}}{1.134(938 \text{ MeV})} = 1.16 \text{ fm}$$

Equation:

$$K = E_0(\gamma - 1) = 938 \text{ MeV}(1 / \sqrt{1 - 0.75^2} - 1) = 480.1 \text{ MeV}$$

Significance

Notice that because a proton is 1835 times more massive than an electron, if this experiment were performed with electrons, a simple rescaling of these results would give us the electron's wavelength of $(1835)0.77\text{fm} = 1.4 \text{ pm}$ and its kinetic energy of $480.1 \text{ MeV} / 1835 = 261.6 \text{ keV}$.

Note:**Exercise:****Problem:**

Check Your Understanding Find the de Broglie wavelength and kinetic energy of a free electron that travels at a speed of $0.75c$.

Solution:

$$\lambda = 2.14 \text{ pm}; K = 261.56 \text{ keV}$$

Summary

- De Broglie's hypothesis of matter waves postulates that any particle of matter that has linear momentum is also a wave. The wavelength of a matter wave associated with a particle is inversely proportional to the magnitude of the particle's linear momentum. The speed of the matter wave is the speed of the particle.
- De Broglie's concept of the electron matter wave provides a rationale for the quantization of the electron's angular momentum in Bohr's model of the hydrogen atom.
- In the Davisson–Germer experiment, electrons are scattered off a crystalline nickel surface. Diffraction patterns of electron matter waves are observed. They are the evidence for the existence of matter waves. Matter waves are observed in diffraction experiments with various particles.

Conceptual Questions**Exercise:**

Problem:

Which type of radiation is most suitable for the observation of diffraction patterns on crystalline solids; radio waves, visible light, or X-rays? Explain.

Solution:

X-rays, best resolving power

Exercise:**Problem:**

Speculate as to how the diffraction patterns of a typical crystal would be affected if γ -rays were used instead of X-rays.

Exercise:**Problem:**

If an electron and a proton are traveling at the same speed, which one has the shorter de Broglie wavelength?

Solution:

proton

Exercise:

Problem: If a particle is accelerating, how does this affect its de Broglie wavelength?

Exercise:**Problem:**

Why is the wave-like nature of matter not observed every day for macroscopic objects?

Solution:

negligibly small de Broglie's wavelengths

Exercise:

Problem: What is the wavelength of a neutron at rest? Explain.

Exercise:**Problem:**

Why does the setup of Davisson–Germer experiment need to be enclosed in a vacuum chamber? Discuss what result you expect when the chamber is not evacuated.

Solution:

to avoid collisions with air molecules

Problems

Exercise:

Problem: At what velocity will an electron have a wavelength of 1.00 m?

Exercise:

Problem:

What is the de Broglie wavelength of an electron travelling at a speed of $5.0 \times 10^6 \text{ m/s}$?

Solution:

145.5 pm

Exercise:

Problem:

What is the de Broglie wavelength of an electron that is accelerated from rest through a potential difference of 20 kV?

Exercise:

Problem:

What is the de Broglie wavelength of a proton whose kinetic energy is 2.0 MeV? 10.0 MeV?

Solution:

20 fm; 9 fm

Exercise:

Problem:

What is the de Broglie wavelength of a 10-kg football player running at a speed of 8.0 m/s?

Exercise:

Problem:

(a) What is the energy of an electron whose de Broglie wavelength is that of a photon of yellow light with wavelength 590 nm? (b) What is the de Broglie wavelength of an electron whose energy is that of the photon of yellow light?

Solution:

a. 2.103 eV; b. 0.846 nm

Exercise:

Problem:

The de Broglie wavelength of a neutron is 0.01 nm. What is the speed and energy of this neutron?

Exercise:

Problem: What is the wavelength of an electron that is moving at a 3% of the speed of light?

Solution:

80.9 pm

Exercise:**Problem:**

At what velocity does a proton have a 6.0-fm wavelength (about the size of a nucleus)? Give your answer in units of c .

Exercise:

Problem: What is the velocity of a 0.400-kg billiard ball if its wavelength is 7.50 fm?

Solution:

$2.21 \times 10^{-19} \text{ m/s}$

Exercise:**Problem:**

Find the wavelength of a proton that is moving at 1.00% of the speed of light (when $\beta = 0.01$).

Glossary

Davisson–Germer experiment

historically first electron-diffraction experiment that revealed electron waves

de Broglie’s hypothesis of matter waves

particles of matter can behave like waves

de Broglie wave

matter wave associated with any object that has mass and momentum

group velocity

velocity of a wave, energy travels with the group velocity

wave quantum mechanics

theory that explains the physics of atoms and subatomic particles

Wave-Particle Duality

By the end of this section, you will be able to:

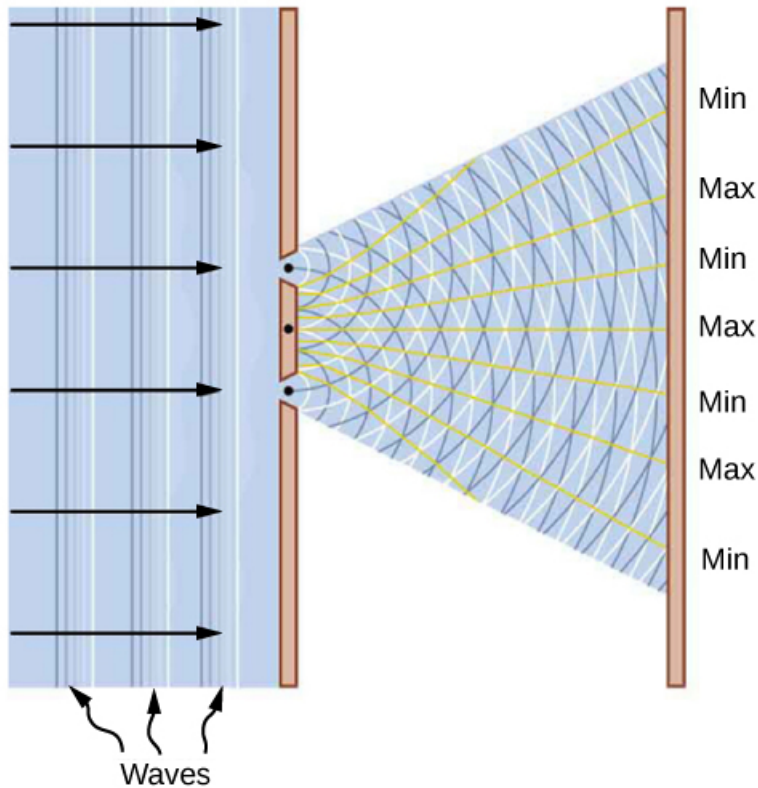
- Identify phenomena in which electromagnetic waves behave like a beam of photons and particles behave like waves
- Describe the physics principles behind electron microscopy
- Summarize the evolution of scientific thought that led to the development of quantum mechanics

The energy of radiation detected by a radio-signal receiving antenna comes as the energy of an electromagnetic wave. The same energy of radiation detected by a photocurrent in the photoelectric effect comes as the energy of individual photon particles. Therefore, the question arises about the nature of electromagnetic radiation: Is a photon a wave or is it a particle?

Similar questions may be asked about other known forms of energy. For example, an electron that forms part of an electric current in a circuit behaves like a particle moving in unison with other electrons inside the conductor. The same electron behaves as a wave when it passes through a solid crystalline structure and forms a diffraction image. Is an electron a wave or is it a particle? The same question can be extended to all particles of matter—elementary particles, as well as compound molecules—asking about their true physical nature. At our present state of knowledge, such questions about the true nature of things do not have conclusive answers.

All we can say is that **wave-particle duality** exists in nature: Under some experimental conditions, a particle appears to act as a particle, and under different experimental conditions, a particle appears to act a wave. Conversely, under some physical circumstances electromagnetic radiation acts as a wave, and under other physical circumstances, radiation acts as a beam of photons.

This dualistic interpretation is not a new physics concept brought about by specific discoveries in the twentieth century. It was already present in a debate between Isaac Newton and Christiaan Huygens about the nature of light, beginning in the year 1670. According to Newton, a beam of light is a collection of corpuscles of light. According to Huygens, light is a wave. The corpuscular hypothesis failed in 1803, when Thomas Young announced his **double-slit interference experiment** with light (see [\[link\]](#)), which firmly established light as a wave. In James Clerk Maxwell's theory of electromagnetism (completed by the year 1873), light is an electromagnetic wave. Maxwell's classical view of radiation as an electromagnetic wave is still valid today; however, it is unable to explain blackbody radiation and the photoelectric effect, where light acts as a beam of photons.



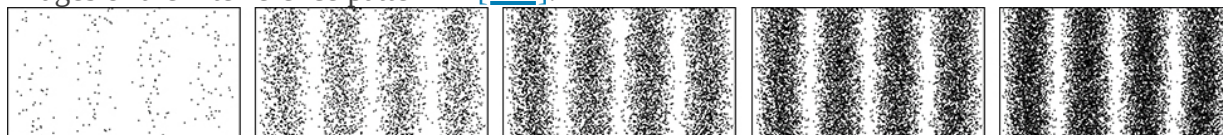
Young's double-slit experiment explains the interference of light by making an analogy with the interference of water waves. Two waves are generated at the positions of two slits in an opaque screen. The waves have the same wavelengths. They travel from their origins at the slits to the viewing screen placed to the right of the slits. The waves meet on the viewing screen. At the positions marked "Max" on the screen, the meeting waves are in-phase and the combined wave amplitude is enhanced. At positions marked "Min," the combined wave amplitude is zero. For light, this mechanism creates a bright-and-dark fringe pattern on the viewing screen.

A similar dichotomy existed in the interpretation of electricity. From Benjamin Franklin's observations of electricity in 1751 until J.J. Thomson's discovery of the electron in 1897, electric current was seen as a flow in a continuous electric medium. Within this theory of electric fluid, the present theory of electric circuits was developed, and electromagnetism and electromagnetic induction were discovered. Thomson's experiment showed that the unit of negative electric charge (an electron) can travel in a vacuum without any medium to carry the charge around, as in electric circuits. This discovery changed the way in which electricity is understood today and gave the electron its particle status. In Bohr's early quantum theory of the

hydrogen atom, both the electron and the proton are particles of matter. Likewise, in the Compton scattering of X-rays on electrons, the electron is a particle. On the other hand, in electron-scattering experiments on crystalline structures, the electron behaves as a wave.

A skeptic may raise a question that perhaps an electron might always be nothing more than a particle, and that the diffraction images obtained in electron-scattering experiments might be explained within some macroscopic model of a crystal and a macroscopic model of electrons coming at it like a rain of ping-pong balls. As a matter of fact, to investigate this question, we do not need a complex model of a crystal but just a couple of simple slits in a screen that is opaque to electrons. In other words, to gather convincing evidence about the nature of an electron, we need to repeat the Young double-slit experiment with electrons. If the electron is a wave, we should observe the formation of interference patterns typical for waves, such as those described in [\[link\]](#), even when electrons come through the slits one by one. However, if the electron is a not a wave but a particle, the interference fringes will not be formed.

The very first double-slit experiment with a beam of electrons, performed by Claus Jönsson in Germany in 1961, demonstrated that a beam of electrons indeed forms an interference pattern, which means that electrons collectively behave as a wave. The first double-slit experiments with *single* electrons passing through the slits one-by-one were performed by Giulio Pozzi in 1974 in Italy and by Akira Tonomura in 1989 in Japan. They show that interference fringes are formed gradually, even when electrons pass through the slits individually. This demonstrates conclusively that electron-diffraction images are formed because of the wave nature of electrons. The results seen in double-slit experiments with electrons are illustrated by the images of the interference pattern in [\[link\]](#).



Computer-simulated interference fringes seen in the Young double-slit experiment with electrons. One pattern is gradually formed on the screen, regardless of whether the electrons come through the slits as a beam or individually one-by-one.

Example:

Double-Slit Experiment with Electrons

In one experimental setup for studying interference patterns of electron waves, two slits are created in a gold-coated silicon membrane. Each slit is 62-nm wide and 4- μm long, and the separation between the slits is 272 nm. The electron beam is created in an electron gun by heating a tungsten element and by accelerating the electrons across a 600-V potential. The beam is subsequently collimated using electromagnetic lenses, and the collimated beam of electrons is sent through the slits. Find the angular position of the first-order bright fringe on the viewing screen.

Strategy

Recall that the angular position θ of the n th order bright fringe that is formed in Young's two-slit interference pattern (discussed in a previous chapter) is related to the separation, d , between the slits and to the wavelength, λ , of the incident light by the equation $d\sin\theta = n\lambda$, where $n = 0, \pm 1, \pm 2, \dots$. The separation is given and is equal to $d = 272 \text{ nm}$. For the first-order fringe, we take $n = 1$. The only thing we now need is the wavelength of the incident electron wave.

Since the electron has been accelerated from rest across a potential difference of $\Delta V = 600 \text{ V}$, its kinetic energy is $K = e\Delta V = 600 \text{ eV}$. The rest-mass energy of the electron is $E_0 = 511 \text{ keV}$.

We compute its de Broglie wavelength as that of a nonrelativistic electron because its kinetic energy K is much smaller than its rest energy E_0 , $K \ll E_0$.

Solution

The electron's wavelength is

Equation:

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m_e K}} = \frac{h}{\sqrt{2E_0/c^2 K}} = \frac{hc}{\sqrt{2E_0 K}} = \frac{1.241 \times 10^{-6} \text{ eV} \cdot \text{m}}{\sqrt{2(511 \text{ keV})(600 \text{ eV})}} = 0.050 \text{ nm}.$$

This λ is used to obtain the position of the first bright fringe:

Equation:

$$\sin\theta = \frac{1 \cdot \lambda}{d} = \frac{0.050 \text{ nm}}{272 \text{ nm}} = 0.000184 \Rightarrow \theta = 0.010^\circ.$$

Significance

Notice that this is also the angular resolution between two consecutive bright fringes up to about $n = 1000$. For example, between the zero-order fringe and the first-order fringe, between the first-order fringe and the second-order fringe, and so on.

Note:

Exercise:

Problem:

Check Your Understanding For the situation described in [\[link\]](#), find the angular position of the fifth-order bright fringe on the viewing screen.

Solution:

0.052°

The wave-particle dual nature of matter particles and of radiation is a declaration of our inability to describe physical reality within one unified classical theory because separately

neither a classical particle approach nor a classical wave approach can fully explain the observed phenomena. This limitation of the classical approach was realized by the year 1928, and a foundation for a new statistical theory, called quantum mechanics, was put in place by Bohr, Edwin Schrödinger, Werner Heisenberg, and Paul Dirac. Quantum mechanics takes de Broglie's idea of matter waves to be the fundamental property of all particles and gives it a statistical interpretation. According to this interpretation, a wave that is associated with a particle carries information about the probable positions of the particle and about its other properties. A single particle is seen as a moving *wave packet* such as the one shown in [\[link\]](#). We can intuitively sense from this example that if a particle is a wave packet, we will not be able to measure its exact position in the same sense as we cannot pinpoint a location of a wave packet in a vibrating guitar string. The uncertainty, Δx , in measuring the particle's position is connected to the uncertainty, Δp , in the simultaneous measuring of its linear momentum by Heisenberg's uncertainty principle:

Note:
Equation:

$$\Delta x \Delta p \geq \frac{1}{2} \hbar.$$

Heisenberg's principle expresses the law of nature that, at the quantum level, our perception is limited. For example, if we know the exact position of a body (which means that $\Delta x = 0$ in [\[link\]](#)) at the same time we cannot know its momentum, because then the uncertainty in its momentum becomes infinite (because $\Delta p \geq 0.5\hbar / \Delta x$ in [\[link\]](#)). The **Heisenberg uncertainty principle** sets the limit on the precision of *simultaneous* measurements of position and momentum of a particle; it shows that the best precision we can obtain is when we have an equals sign (=) in [\[link\]](#), and we cannot do better than that, even with the best instruments of the future. Heisenberg's principle is a consequence of the wave nature of particles.

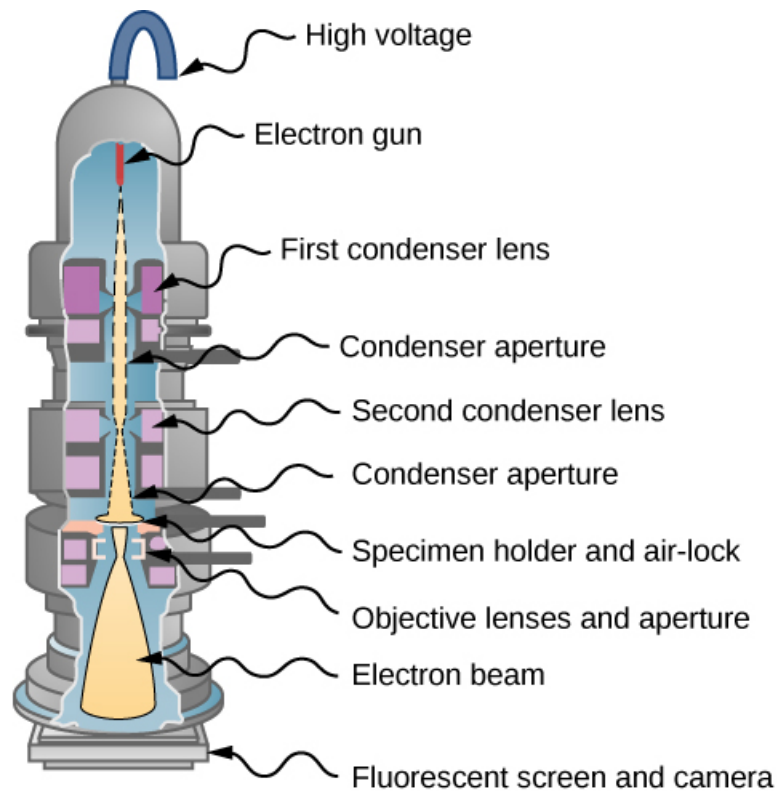


In this graphic, a particle is shown as a wave packet

and its position does not have an exact value.

We routinely use many electronic devices that exploit wave-particle duality without even realizing the sophistication of the physics underlying their operation. One example of a technology based on the particle properties of photons and electrons is a charge-coupled device, which is used for light detection in any instrumentation where high-quality digital data are required, such as in digital cameras or in medical sensors. An example in which the wave properties of electrons is exploited is an electron microscope.

In 1931, physicist Ernst Ruska—building on the idea that magnetic fields can direct an electron beam just as lenses can direct a beam of light in an optical microscope—developed the first prototype of the electron microscope. This development originated the field of **electron microscopy**. In the transmission electron microscope (TEM), shown in [\[link\]](#), electrons are produced by a hot tungsten element and accelerated by a potential difference in an electron gun, which gives them up to 400 keV in kinetic energy. After leaving the electron gun, the electron beam is focused by electromagnetic lenses (a system of condensing lenses) and transmitted through a specimen sample to be viewed. The image of the sample is reconstructed from the transmitted electron beam. The magnified image may be viewed either directly on a fluorescent screen or indirectly by sending it, for example, to a digital camera or a computer monitor. The entire setup consisting of the electron gun, the lenses, the specimen, and the fluorescent screen are enclosed in a vacuum chamber to prevent the energy loss from the beam. Resolution of the TEM is limited only by spherical aberration (discussed in a previous chapter). Modern high-resolution models of a TEM can have resolving power greater than 0.5 Å and magnifications higher than 50 million times. For comparison, the best resolving power obtained with light microscopy is currently about 97 nm. A limitation of the TEM is that the samples must be about 100-nm thick and biological samples require a special preparation involving chemical “fixing” to stabilize them for ultrathin slicing.



TEM: An electron beam produced by an electron gun is collimated by condenser lenses and passes through a specimen. The transmitted electrons are projected on a screen and the image is sent to a camera. (credit: modification of work by Dr. Graham Beards)

Such limitations do not appear in the scanning electron microscope (SEM), which was invented by Manfred von Ardenne in 1937. In an SEM, a typical energy of the electron beam is up to 40 keV and the beam is not transmitted through a sample but is scattered off its surface. Surface topography of the sample is reconstructed by analyzing back-scattered electrons, transmitted electrons, and the emitted radiation produced by electrons interacting with atoms in the sample. The resolving power of an SEM is better than 1 nm, and the magnification can be more than 250 times better than that obtained with a light microscope. The samples scanned by an SEM can be as large as several centimeters but they must be specially prepared, depending on electrical properties of the sample.

High magnifications of the TEM and SEM allow us to see individual molecules. High resolving powers of the TEM and SEM allow us to see fine details, such as those shown in the SEM micrograph of pollen at the beginning of this chapter ([link](#)).

Example:

Resolving Power of an Electron Microscope

If a 1.0-pm electron beam of a TEM passes through a 2.0- μm circular opening, what is the angle between the two just-resolvable point sources for this microscope?

Solution

We can directly use a formula for the resolving power, $\Delta\theta$, of a microscope (discussed in a previous chapter) when the wavelength of the incident radiation is $\lambda = 1.0 \text{ pm}$ and the diameter of the aperture is $D = 2.0\mu\text{m}$:

Equation:

$$\Delta\theta = 1.22 \frac{\lambda}{D} = 1.22 \frac{1.0 \text{ pm}}{2.0\mu\text{m}} = 6.10 \times 10^{-7} \text{ rad} = 3.50 \times 10^{-5} \text{ degree.}$$

Significance

Note that if we used a conventional microscope with a 400-nm light, the resolving power would be only 14° , which means that all of the fine details in the image would be blurred.

Note:

Exercise:

Problem:

Check Your Understanding Suppose that the diameter of the aperture in [\[link\]](#) is halved. How does it affect the resolving power?

Solution:

doubles it

Summary

- Wave-particle duality exists in nature: Under some experimental conditions, a particle acts as a particle; under other experimental conditions, a particle acts as a wave. Conversely, under some physical circumstances, electromagnetic radiation acts as a wave, and under other physical circumstances, radiation acts as a beam of photons.
- Modern-era double-slit experiments with electrons demonstrated conclusively that electron-diffraction images are formed because of the wave nature of electrons.
- The wave-particle dual nature of particles and of radiation has no classical explanation.
- Quantum theory takes the wave property to be the fundamental property of all particles. A particle is seen as a moving wave packet. The wave nature of particles imposes a limitation on the simultaneous measurement of the particle's position and momentum. Heisenberg's uncertainty principle sets the limits on precision in such simultaneous measurements.

- Wave-particle duality is exploited in many devices, such as charge-couple devices (used in digital cameras) or in the electron microscopy of the scanning electron microscope (SEM) and the transmission electron microscope (TEM).

Key Equations

Wien's displacement law	$\lambda_{\max} T = 2.898 \times 10^{-3} \text{m} \cdot \text{K}$
Stefan's law	$P(T) = \sigma AT^4$
Planck's constant	$h = 6.626 \times 10^{-34} \text{J} \cdot \text{s} = 4.136 \times 10^{-15} \text{eV} \cdot \text{s}$
Energy quantum of radiation	$\Delta E = hf$
Planck's blackbody radiation law	$I(\lambda, T) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1}$
Maximum kinetic energy of a photoelectron	$K_{\max} = e\Delta V_s$
Energy of a photon	$E_f = hf$
Energy balance for photoelectron	$K_{\max} = hf - \phi$
Cut-off frequency	$f_c = \frac{\phi}{h}$
Relativistic invariant energy equation	$E^2 = p^2 c^2 + m_0^2 c^4$
Energy-momentum relation for photon	$p_f = \frac{E_f}{c}$
Energy of a photon	$E_f = hf = \frac{hc}{\lambda}$
Magnitude of photon's momentum	$p_f = \frac{h}{\lambda}$
Photon's linear momentum vector	$\vec{p}_f = \hbar \vec{k}$
The Compton wavelength	$\lambda_c = \frac{h}{m_0 c} = 0.00243 \text{ nm}$

of an electron	
The Compton shift	$\Delta\lambda = \lambda_c(1 - \cos \theta)$
The Balmer formula	$\frac{1}{\lambda} = R_H \left(\frac{1}{2^2} - \frac{1}{n^2} \right)$
The Rydberg formula	$\frac{1}{\lambda} = R_H \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right), n_i = n_f + 1, n_f + 2, \dots$
Bohr's first quantization condition	$L_n = n\hbar, n = 1, 2, \dots$
Bohr's second quantization condition	$hf = E_n - E_m $
Bohr's radius of hydrogen	$a_0 = 4\pi\epsilon_0 \frac{\hbar^2}{m_e e^2} = 0.529\text{\AA}$
Bohr's radius of the n th orbit	$r_n = a_0 n^2$
Ground-state energy value, ionization limit	$E_0 = \frac{1}{8\epsilon_0^2} \frac{m_e e^4}{\hbar^2} = 13.6 \text{ eV}$
Electron's energy in the n th orbit	$E_n = -E_0 \frac{1}{n^2}$
Ground state energy of hydrogen	$E_1 = -E_0 = -13.6 \text{ eV}$
The n th orbit of hydrogen-like ion	$r_n = \frac{a_0}{Z} n^2$
The n th energy of hydrogen-like ion	$E_n = -Z^2 E_0 \frac{1}{n^2}$
Energy of a matter wave	$E = hf$
The de Broglie wavelength	$\lambda = \frac{h}{p}$
The frequency-wavelength relation for matter waves	$\lambda f = \frac{c}{\beta}$
Heisenberg's uncertainty principle	$\Delta x \Delta p \geq \frac{1}{2} \hbar$

Conceptual Questions

Exercise:

Problem:

Give an example of an experiment in which light behaves as waves. Give an example of an experiment in which light behaves as a stream of photons.

Exercise:

Problem:

Discuss: How does the interference of water waves differ from the interference of electrons? How are they analogous?

Solution:

Answers may vary

Exercise:

Problem: Give at least one argument in support of the matter-wave hypothesis.

Exercise:

Problem: Give at least one argument in support of the particle-nature of radiation.

Solution:

Answers may vary

Exercise:

Problem: Explain the importance of the Young double-slit experiment.

Exercise:

Problem:

Does the Heisenberg uncertainty principle allow a particle to be at rest in a designated region in space?

Solution:

yes

Exercise:

Problem: Can the de Broglie wavelength of a particle be known exactly?

Exercise:

Problem:

Do the photons of red light produce better resolution in a microscope than blue light photons? Explain.

Solution:

yes

Exercise:

Problem: Discuss the main difference between an SEM and a TEM.

Problems**Exercise:****Problem:**

An AM radio transmitter radiates 500 kW at a frequency of 760 kHz. How many photons per second does the emitter emit?

Solution:

$$9.929 \times 10^{32}$$

Exercise:**Problem:**

Find the Lorentz factor γ and de Broglie's wavelength for a 50-GeV electron in a particle accelerator.

Exercise:**Problem:**

Find the Lorentz factor γ and de Broglie's wavelength for a 1.0-TeV proton in a particle accelerator.

Solution:

$$\gamma = 1060; 0.00124 \text{ fm}$$

Exercise:

Problem: What is the kinetic energy of a 0.01-nm electron in a TEM?

Exercise:

Problem:

If electron is to be diffracted significantly by a crystal, its wavelength must be about equal to the spacing, d , of crystalline planes. Assuming $d = 0.250 \text{ nm}$, estimate the potential difference through which an electron must be accelerated from rest if it is to be diffracted by these planes.

Solution:

24.11 V

Exercise:**Problem:**

X-rays form ionizing radiation that is dangerous to living tissue and undetectable to the human eye. Suppose that a student researcher working in an X-ray diffraction laboratory is accidentally exposed to a fatal dose of radiation. Calculate the temperature increase of the researcher under the following conditions: the energy of X-ray photons is 200 keV and the researcher absorbs 4×10^{13} photons per each kilogram of body weight during the exposure. Assume that the specific heat of the student's body is $0.83 \text{ kcal/kg} \cdot \text{K}$.

Exercise:**Problem:**

Solar wind (radiation) that is incident on the top of Earth's atmosphere has an average intensity of 1.3 kW/m^2 . Suppose that you are building a solar sail that is to propel a small toy spaceship with a mass of 0.1 kg in the space between the International Space Station and the moon. The sail is made from a very light material, which perfectly reflects the incident radiation. To assess whether such a project is feasible, answer the following questions, assuming that radiation photons are incident only in normal direction to the sail reflecting surface. (a) What is the radiation pressure (force per m^2) of the radiation falling on the mirror-like sail? (b) Given the radiation pressure computed in (a), what will be the acceleration of the spaceship when the sail has of an area of 10.0 m^2 ? (c) Given the acceleration estimate in (b), how fast will the spaceship be moving after 24 hours when it starts from rest?

Solution:

a. $P = 2I/c = 8.67 \times 10^{-6} \text{ N/m}^2$; b. $a = PA/m = 8.67 \times 10^{-4} \text{ m/s}^2$; c. 74.91 m/s

Exercise:**Problem:**

Treat the human body as a blackbody and determine the percentage increase in the total power of its radiation when its temperature increases from 98.6° F to 103° F .

Exercise:**Problem:**

Show that Wien's displacement law results from Planck's radiation law. (*Hint:* substitute $x = hc / \lambda kT$ and write Planck's law in the form $I(x, T) = Ax^5 / (e^x - 1)$, where $A = 2\pi(kT)^5 / (h^4 c^3)$. Now, for fixed T , find the position of the maximum in $I(x, T)$ by solving for x in the equation $dI(x, T) / dx = 0$.)

Solution:

$$x = 4.965$$

Exercise:**Problem:**

Show that Stefan's law results from Planck's radiation law. *Hint:* To compute the total power of blackbody radiation emitted across the entire spectrum of wavelengths at a given temperature, integrate Planck's law over the entire spectrum $P(T) = \int_0^\infty I(\lambda, T) d\lambda$.

Use the substitution $x = hc / \lambda kT$ and the tabulated value of the integral

$$\int_0^\infty dx x^3 / (e^x - 1) = \pi^4 / 15.$$

Additional Problems**Exercise:****Problem:**

Determine the power intensity of radiation per unit wavelength emitted at a wavelength of 500.0 nm by a blackbody at a temperature of 10,000 K.

Solution:

$$7.124 \times 10^{16} \text{ W/m}^3$$

Exercise:**Problem:**

The HCl molecule oscillates at a frequency of 87.0 THz. What is the difference (in eV) between its adjacent energy levels?

Exercise:

Problem:

A quantum mechanical oscillator vibrates at a frequency of 250.0 THz. What is the minimum energy of radiation it can emit?

Solution:

1.034 eV

Exercise:**Problem:**

In about 5 billion years, the sun will evolve to a red giant. Assume that its surface temperature will decrease to about half its present value of 6000 K, while its present radius of $7.0 \times 10^8 \text{ m}$ will increase to $1.5 \times 10^{11} \text{ m}$ (which is the current Earth-sun distance). Calculate the ratio of the total power emitted by the sun in its red giant stage to its present power.

Exercise:**Problem:**

A sodium lamp emits 2.0 W of radiant energy, most of which has a wavelength of about 589 nm. Estimate the number of photons emitted per second by the lamp.

Solution:

5.93×10^{18}

Exercise:**Problem:**

Photoelectrons are ejected from a photoelectrode and are detected at a distance of 2.50 cm away from the photoelectrode. The work function of the photoelectrode is 2.71 eV and the incident radiation has a wavelength of 420 nm. How long does it take a photoelectron to travel to the detector?

Exercise:**Problem:**

If the work function of a metal is 3.2 eV, what is the maximum wavelength that a photon can have to eject a photoelectron from this metal surface?

Solution:

387.8 nm

Exercise:

Problem:

The work function of a photoelectric surface is 2.00 eV. What is the maximum speed of the photoelectrons emitted from this surface when a 450-nm light falls on it?

Exercise:**Problem:**

A 400-nm laser beam is projected onto a calcium electrode. The power of the laser beam is 2.00 mW and the work function of calcium is 2.31 eV. (a) How many photoelectrons per second are ejected? (b) What net power is carried away by photoelectrons?

Solution:

a. 4.02×10^{15} ; b. 0.533 mW

Exercise:**Problem:**

(a) Calculate the number of photoelectrons per second that are ejected from a 1.00-mm² area of sodium metal by a 500-nm radiation with intensity 1.30kW/m² (the intensity of sunlight above Earth's atmosphere). (b) Given the work function of the metal as 2.28 eV, what power is carried away by these photoelectrons?

Exercise:**Problem:**

A laser with a power output of 2.00 mW at a 400-nm wavelength is used to project a beam of light onto a calcium photoelectrode. (a) How many photoelectrons leave the calcium surface per second? (b) What power is carried away by ejected photoelectrons, given that the work function of calcium is 2.31 eV? (c) Calculate the photocurrent. (d) If the photoelectrode suddenly becomes electrically insulated and the setup of two electrodes in the circuit suddenly starts to act like a 2.00-pF capacitor, how long will current flow before the capacitor voltage stops it?

Solution:

a. 4.02×10^{15} ; b. 0.533 mW; c. 0.644 mA; d. 2.57 ns

Exercise:**Problem:**

The work function for barium is 2.48 eV. Find the maximum kinetic energy of the ejected photoelectrons when the barium surface is illuminated with: (a) radiation emitted by a 100-kW radio station broadcasting at 800 kHz; (b) a 633-nm laser light emitted from a powerful He-Ne laser; and (c) a 434-nm blue light emitted by a small hydrogen gas discharge tube.

Exercise:**Problem:**

- (a) Calculate the wavelength of a photon that has the same momentum as a proton moving with 1% of the speed of light in a vacuum. (b) What is the energy of this photon in MeV? (c) What is the kinetic energy of the proton in MeV?
-

Solution:

a. 0.132 pm; b. 9.39 MeV; c. 0.047 MeV

Exercise:**Problem:**

- (a) Find the momentum of a 100-keV X-ray photon. (b) Find the velocity of a neutron with the same momentum. (c) What is the neutron's kinetic energy in eV?

Exercise:**Problem:**

The momentum of light, as it is for particles, is exactly reversed when a photon is reflected straight back from a mirror, assuming negligible recoil of the mirror. The change in momentum is twice the photon's incident momentum, as it is for the particles. Suppose that a beam of light has an intensity 1.0 kW/m^2 and falls on a 2.0-m^2 area of a mirror and reflects from it. (a) Calculate the energy reflected in 1.00 s. (b) What is the momentum imparted to the mirror? (c) Use Newton's second law to find the force on the mirror. (d) Does the assumption of no-recoil for the mirror seem reasonable?

Solution:

a. 2 kJ; b. $1.33 \times 10^{-5} \text{ kg} \cdot \text{m/s}$; c. $1.33 \times 10^{-5} \text{ N}$; d. yes

Exercise:**Problem:**

A photon of energy 5.0 keV collides with a stationary electron and is scattered at an angle of 60° . What is the energy acquired by the electron in the collision?

Exercise:**Problem:**

A 0.75-nm photon is scattered by a stationary electron. The speed of the electron's recoil is $1.5 \times 10^6 \text{ m/s}$. (a) Find the wavelength shift of the photon. (b) Find the scattering angle of the photon.

Solution:

a. 0.003 nm; b. 105.56°

Exercise:

Problem:

Find the maximum change in X-ray wavelength that can occur due to Compton scattering. Does this change depend on the wavelength of the incident beam?

Exercise:

Problem:

A photon of wavelength 700 nm is incident on a hydrogen atom. When this photon is absorbed, the atom becomes ionized. What is the lowest possible orbit that the electron could have occupied before being ionized?

Solution:

$$n = 3$$

Exercise:

Problem:

What is the maximum kinetic energy of an electron such that a collision between the electron and a stationary hydrogen atom in its ground state is definitely elastic?

Exercise:

Problem:

Singly ionized atomic helium He^{+1} is a hydrogen-like ion. (a) What is its ground-state radius? (b) Calculate the energies of its four lowest energy states. (c) Repeat the calculations for the Li^{2+} ion.

Solution:

$$\text{a. } a_0/2; \text{ b. } -54.4 \text{ eV}/n^2; \text{ c. } a_0/3, -122.4 \text{ eV}/n^2$$

Exercise:

Problem:

A triply ionized atom of beryllium Be^{3+} is a hydrogen-like ion. When Be^{3+} is in one of its excited states, its radius in this n th state is exactly the same as the radius of the first Bohr orbit of hydrogen. Find n and compute the ionization energy for this state of Be^{3+} .

Exercise:

Problem:

In extreme-temperature environments, such as those existing in a solar corona, atoms may be ionized by undergoing collisions with other atoms. One example of such ionization in the solar corona is the presence of C^{5+} ions, detected in the Fraunhofer spectrum. (a) By what factor do the energies of the C^{5+} ion scale compare to the energy spectrum of a hydrogen atom? (b) What is the wavelength of the first line in the Paschen series of C^{5+} ? (c) In what part of the spectrum are these lines located?

Solution:

a. 36; b. 18.2 nm; c. UV

Exercise:**Problem:**

(a) Calculate the ionization energy for He^+ . (b) What is the minimum frequency of a photon capable of ionizing He^+ ?

Exercise:**Problem:**

Experiments are performed with ultracold neutrons having velocities as small as 1.00 m/s. Find the wavelength of such an ultracold neutron and its kinetic energy.

Solution:

396 nm; 5.23 neV

Exercise:**Problem:**

Find the velocity and kinetic energy of a 6.0-fm neutron. (Rest mass energy of neutron is $E_0 = 940 \text{ MeV}$.)

Exercise:**Problem:**

The spacing between crystalline planes in the NaCl crystal is 0.281 nm, as determined by X-ray diffraction with X-rays of wavelength 0.170 nm. What is the energy of neutrons in the neutron beam that produces diffraction peaks at the same locations as the peaks obtained with the X-rays?

Solution:

7.3 keV

Exercise:

Problem:

What is the wavelength of an electron accelerated from rest in a 30.0-kV potential difference?

Exercise:**Problem:**

Calculate the velocity of a 1.0- μm electron and a potential difference used to accelerate it from rest to this velocity.

Solution:

728 m/s; 1.5 μV

Exercise:**Problem:**

In a supercollider at CERN, protons are accelerated to velocities of $0.25c$. What are their wavelengths at this speed? What are their kinetic energies? If a beam of protons were to gain its kinetic energy in only one pass through a potential difference, how high would this potential difference have to be? (Rest mass energy of a proton is $E_0 = 938 \text{ MeV}$).

Exercise:**Problem:**

Find the de Broglie wavelength of an electron accelerated from rest in an X-ray tube in the potential difference of 100 keV. (Rest mass energy of an electron is $E_0 = 511 \text{ keV}$.)

Solution:

$$\lambda = hc / \sqrt{K(2E_0 + K)} = 3.705 \times 10^{-12} \text{ m}, K = 100 \text{ keV}$$

Exercise:**Problem:**

The cutoff wavelength for the emission of photoelectrons from a particular surface is 500 nm. Find the maximum kinetic energy of the ejected photoelectrons when the surface is illuminated with light of wavelength 450 nm.

Exercise:**Problem:**

Compare the wavelength shift of a photon scattered by a free electron to that of a photon scattered at the same angle by a free proton.

Solution:

$$\Delta\lambda_c^{(\text{electron})} / \Delta\lambda_c^{(\text{proton})} = m_p / m_e = 1836$$

Exercise:

Problem:

The spectrometer used to measure the wavelengths of the scattered X-rays in the Compton experiment is accurate to $5.0 \times 10^{-4}\text{nm}$. What is the minimum scattering angle for which the X-rays interacting with the free electrons can be distinguished from those interacting with the atoms?

Exercise:

Problem:

Consider a hydrogen-like ion where an electron is orbiting a nucleus that has charge $q = +Ze$. Derive the formulas for the energy E_n of the electron in n th orbit and the orbital radius r_n .

Solution:

(Proof)

Exercise:

Problem:

Assume that a hydrogen atom exists in the $n = 2$ excited state for 10^{-8}s before decaying to the ground state. How many times does the electron orbit the proton nucleus during this time? How long does it take Earth to orbit the sun this many times?

Exercise:

Problem:

An atom can be formed when a negative muon is captured by a proton. The muon has the same charge as the electron and a mass 207 times that of the electron. Calculate the frequency of the photon emitted when this atom makes the transition from $n = 2$ to the $n = 1$ state. Assume that the muon is orbiting a stationary proton.

Solution:

$$5.1 \times 10^{17}\text{Hz}$$

Glossary

double-slit interference experiment

Young's double-slit experiment, which shows the interference of waves

electron microscopy

microscopy that uses electron waves to "see" fine details of nano-size objects

Heisenberg uncertainty principle

sets the limits on precision in simultaneous measurements of momentum and position of a particle

wave-particle duality

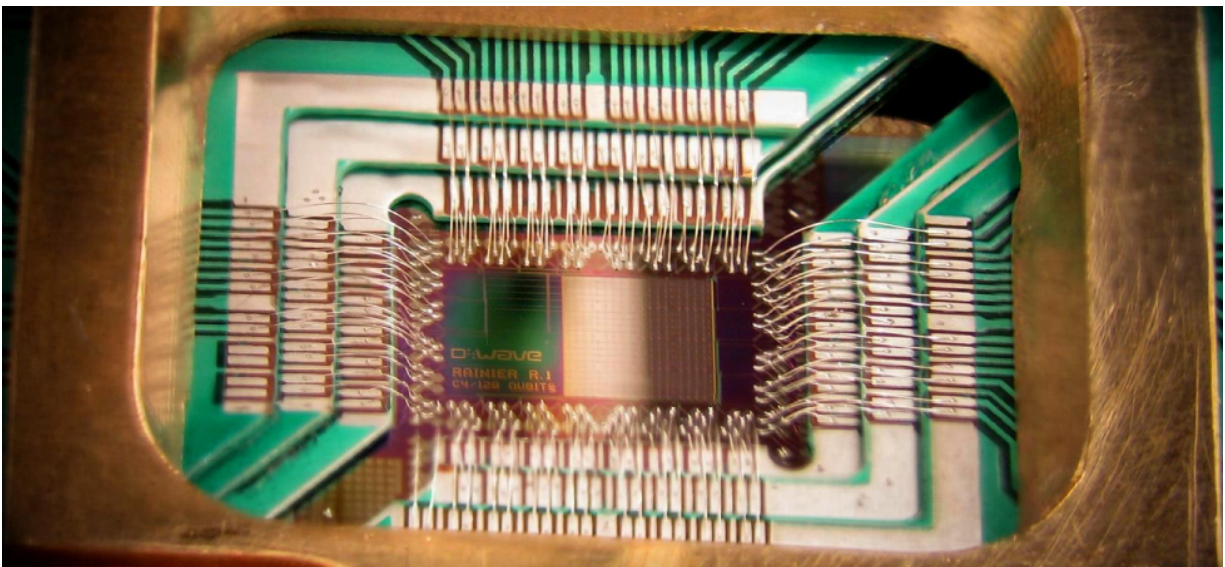
particles can behave as waves and radiation can behave as particles

Introduction

class="introduction"

A D-wave
qubit
processor:
The brain of
a quantum
computer
that encodes
information
in quantum
bits to
perform
complex
calculations.

(credit:
modificatio
n of work
by D-Wave
Systems,
Inc.)



Quantum mechanics is a powerful framework for understanding the motions and interactions of particles at small scales, such as atoms and molecules. The ideas behind quantum mechanics often appear quite strange. In many ways, our everyday experience with the macroscopic physical world does not prepare us for the microscopic world of quantum mechanics. The purpose of this chapter is to introduce you to this exciting world.

Pictured above is a quantum-computer processor. This device is the “brain” of a quantum computer that operates at near-absolute zero temperatures. Unlike a digital computer, which encodes information in binary digits (definite states of either zero or one), a quantum computer encodes information in quantum bits or qubits (mixed states of zero *and* one). Quantum computers are discussed in the first section of this chapter.

Wave Functions

By the end of this section, you will be able to:

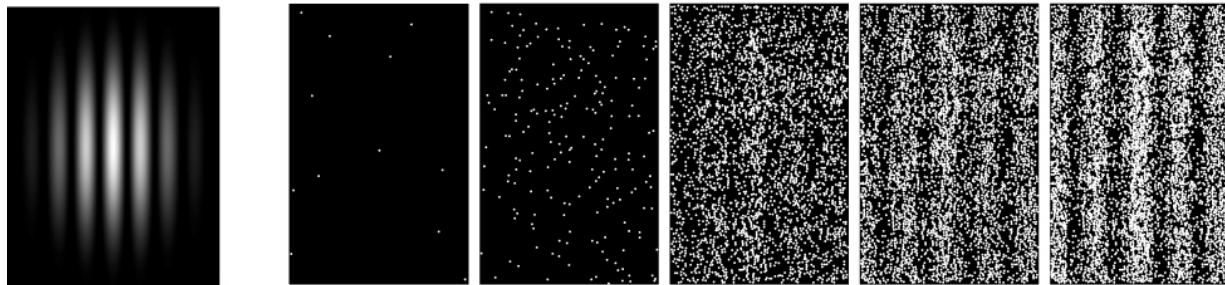
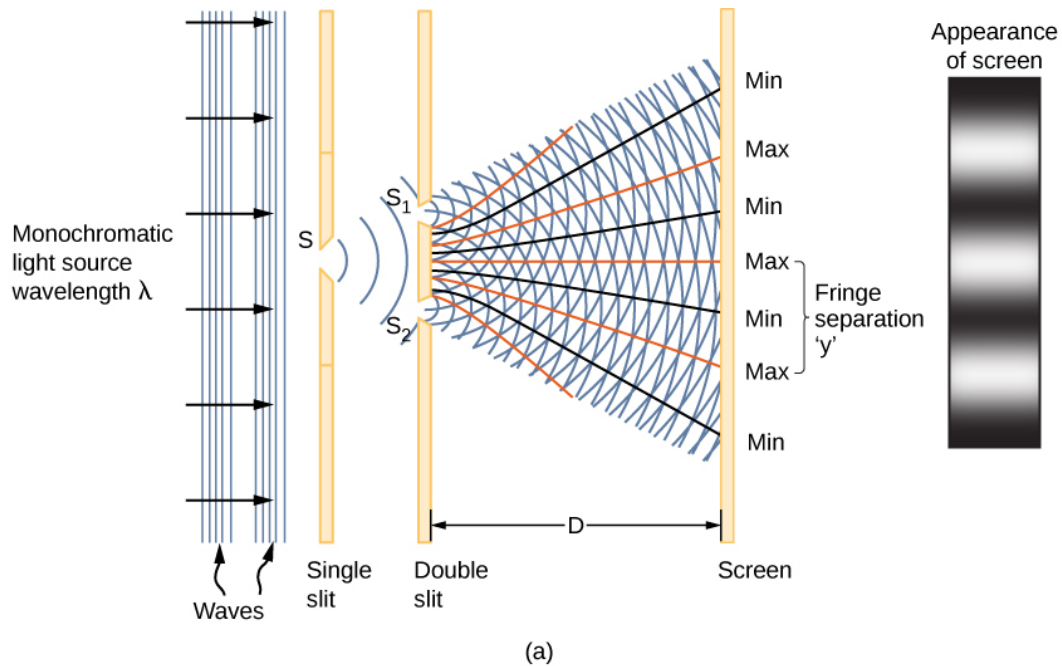
- Describe the statistical interpretation of the wave function
- Use the wave function to determine probabilities
- Calculate expectation values of position, momentum, and kinetic energy

In the preceding chapter, we saw that particles act in some cases like particles and in other cases like waves. But what does it mean for a particle to “act like a wave”? What precisely is “waving”? What rules govern how this wave changes and propagates? How is the wave function used to make predictions? For example, if the amplitude of an electron wave is given by a function of position and time, $\Psi(x, t)$, defined for all x , *where* exactly is the electron? The purpose of this chapter is to answer these questions.

Using the Wave Function

A clue to the physical meaning of the wave function $\Psi(x, t)$ is provided by the two-slit interference of monochromatic light ([link](#)). (See also [Electromagnetic Waves](#) and [Interference](#).) The **wave function** of a light wave is given by $E(x, t)$, and its energy density is given by $|E|^2$, where E is the electric field strength. The energy of an individual photon depends only on the frequency of light, $\varepsilon_{\text{photon}} = hf$, so $|E|^2$ is proportional to the number of photons. When light waves from S_1 interfere with light waves from S_2 at the viewing screen (a distance D away), an interference pattern is produced (part (a) of the figure). Bright fringes correspond to points of constructive interference of the light waves, and dark fringes correspond to points of destructive interference of the light waves (part (b)).

Suppose the screen is initially unexposed to light. If the screen is exposed to very weak light, the interference pattern appears gradually ([link](#))(c), left to right). Individual photon hits on the screen appear as dots. The dot density is expected to be large at locations where the interference pattern will be, ultimately, the most intense. In other words, the probability (per unit area) that a single photon will strike a particular spot on the screen is proportional to the square of the total electric field, $|E|^2$ at that point. Under the right conditions, the same interference pattern develops for matter particles, such as electrons.



Two-slit interference of monochromatic light. (a) Schematic of two-slit interference; (b) light interference pattern; (c) interference pattern built up gradually under low-intensity light (left to right).

Note:

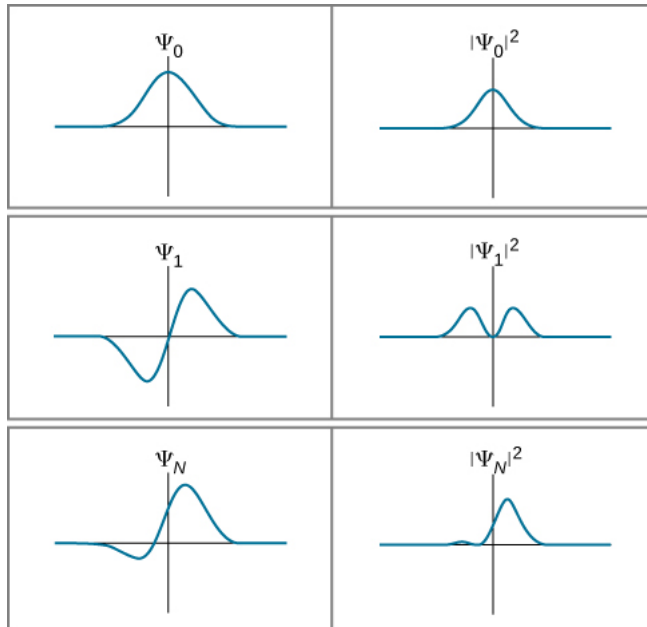
Visit this [interactive simulation](#) to learn more about quantum wave interference.

The square of the matter wave $|\Psi|^2$ in one dimension has a similar interpretation as the square of the electric field $|E|^2$. It gives the probability that a particle will be found at a particular position and time per unit length, also called the **probability density**. The probability (P) a particle is found in a narrow interval $(x, x + dx)$ at time t is therefore

Equation:

$$P(x, x + dx) = |\Psi(x, t)|^2 dx.$$

(Later, we define the magnitude squared for the general case of a function with “imaginary parts.”) This probabilistic interpretation of the wave function is called the **Born interpretation**. Examples of wave functions and their squares for a particular time t are given in [\[link\]](#).



Several examples of wave functions and the corresponding square of their wave functions.

If the wave function varies slowly over the interval Δx , the probability a particle is found in the interval is approximately

Equation:

$$P(x, x + \Delta x) \approx |\Psi(x, t)|^2 \Delta x.$$

Notice that squaring the wave function ensures that the probability is positive. (This is analogous to squaring the electric field strength—which may be positive or negative—to obtain a positive value of intensity.) However, if the wave function does not vary slowly, we must integrate:

Equation:

$$P(x, x + \Delta x) = \int_x^{x+\Delta x} |\Psi(x, t)|^2 dx.$$

This probability is just the area under the function $|\Psi(x, t)|^2$ between x and $x + \Delta x$. The probability of finding the particle “somewhere” (the **normalization condition**) is

Note:

Equation:

$$P(-\infty, +\infty) = \int_{-\infty}^{\infty} |\Psi(x, t)|^2 dx = 1.$$

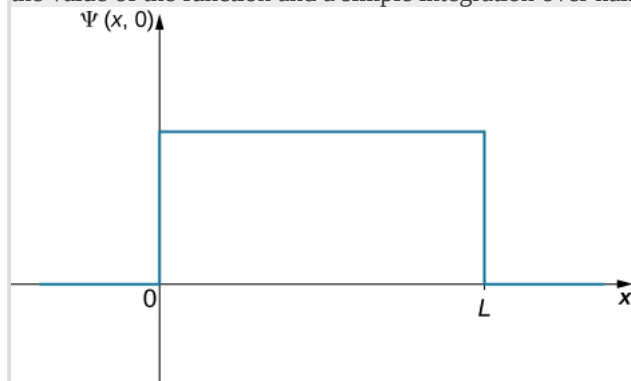
For a particle in two dimensions, the integration is over an area and requires a double integral; for a particle in three dimensions, the integration is over a volume and requires a triple integral. For now, we stick to the simple one-dimensional case.

Example:**Where Is the Ball? (Part I)**

A ball is constrained to move along a line inside a tube of length L . The ball is equally likely to be found anywhere in the tube at some time t . What is the probability of finding the ball in the left half of the tube at that time? (The answer is 50%, of course, but how do we get this answer by using the probabilistic interpretation of the quantum mechanical wave function?)

Strategy

The first step is to write down the wave function. The ball is equally likely to be found anywhere in the box, so one way to describe the ball with a *constant* wave function ([link](#)). The normalization condition can be used to find the value of the function and a simple integration over half of the box yields the final answer.



Wave function for a ball in a tube of length L .

Solution

The wave function of the ball can be written as $\Psi(x, t) = C(0 < x < L)$, where C is a constant, and $\Psi(x, t) = 0$ otherwise. We can determine the constant C by applying the normalization condition (we set $t = 0$ to simplify the notation):

Equation:

$$P(x = -\infty, +\infty) = \int_{-\infty}^{\infty} |C|^2 dx = 1.$$

This integral can be broken into three parts: (1) negative infinity to zero, (2) zero to L , and (3) L to infinity. The particle is constrained to be in the tube, so $C = 0$ outside the tube and the first and last integrations are zero. The above equation can therefore be written

Equation:

$$P(x = 0, L) = \int_0^L |C|^2 dx = 1.$$

The value C does not depend on x and can be taken out of the integral, so we obtain

Equation:

$$|C|^2 \int_0^L dx = 1.$$

Integration gives

Equation:

$$C = \sqrt{\frac{1}{L}}.$$

To determine the probability of finding the ball in the first half of the box ($0 < x < L$), we have

Equation:

$$P(x = 0, L/2) = \int_0^{L/2} \left| \sqrt{\frac{1}{L}} \right|^2 dx = \left(\frac{1}{L} \right) \frac{L}{2} = 0.50.$$

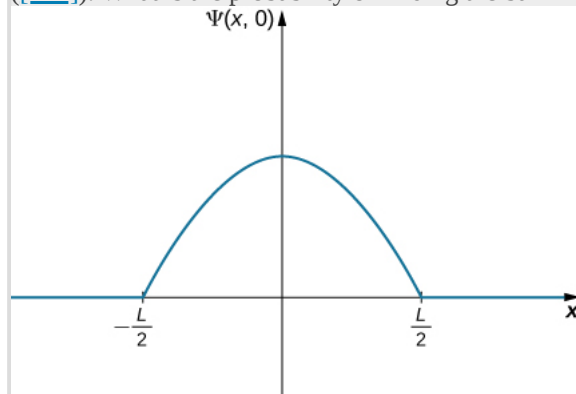
Significance

The probability of finding the ball in the first half of the tube is 50%, as expected. Two observations are noteworthy. First, this result corresponds to the area under the constant function from $x = 0$ to $L/2$ (the area of a square left of $L/2$). Second, this calculation requires an integration of the *square* of the wave function. A common mistake in performing such calculations is to forget to square the wave function before integration.

Example:

Where Is the Ball? (Part II)

A ball is again constrained to move along a line inside a tube of length L . This time, the ball is found preferentially in the middle of the tube. One way to represent its wave function is with a simple cosine function ([link](#)). What is the probability of finding the ball in the last one-quarter of the tube?



Wave function for a ball in a tube of length L , where the ball is preferentially in the middle of the tube.

Strategy

We use the same strategy as before. In this case, the wave function has two unknown constants: One is associated with the wavelength of the wave and the other is the amplitude of the wave. We determine the amplitude by using the boundary conditions of the problem, and we evaluate the wavelength by using the normalization condition. Integration of the square of the wave function over the last quarter of the tube yields the final answer. The calculation is simplified by centering our coordinate system on the peak of the wave function.

Solution

The wave function of the ball can be written

Equation:

$$\Psi(x, 0) = A \cos(kx) (-L/2 < x < L/2),$$

where A is the amplitude of the wave function and $k = 2\pi/\lambda$ is its wave number. Beyond this interval, the amplitude of the wave function is zero because the ball is confined to the tube. Requiring the wave function to terminate at the right end of the tube gives

Equation:

$$\Psi\left(x = \frac{L}{2}, 0\right) = 0.$$

Evaluating the wave function at $x = L/2$ gives

Equation:

$$A \cos(kL/2) = 0.$$

This equation is satisfied if the argument of the cosine is an integral multiple of $\pi/2$, $3\pi/2$, $5\pi/2$, and so on. In this case, we have

Equation:

$$\frac{kL}{2} = \frac{\pi}{2},$$

or

Equation:

$$k = \frac{\pi}{L}.$$

Applying the normalization condition gives $A = \sqrt{2/L}$, so the wave function of the ball is

Equation:

$$\Psi(x, 0) = \sqrt{\frac{2}{L}} \cos(\pi x/L), \quad -L/2 < x < L/2.$$

To determine the probability of finding the ball in the last quarter of the tube, we square the function and integrate:

Equation:

$$P(x = L/4, L/2) = \int_{L/4}^{L/2} \left| \sqrt{\frac{2}{L}} \cos\left(\frac{\pi x}{L}\right) \right|^2 dx = 0.091.$$

Significance

The probability of finding the ball in the last quarter of the tube is 9.1%. The ball has a definite wavelength ($\lambda = 2L$). If the tube is of macroscopic length ($L = 1\text{ m}$), the momentum of the ball is

Equation:

$$p = \frac{h}{\lambda} = \frac{h}{2L} \sim 10^{-36} \text{ m/s}.$$

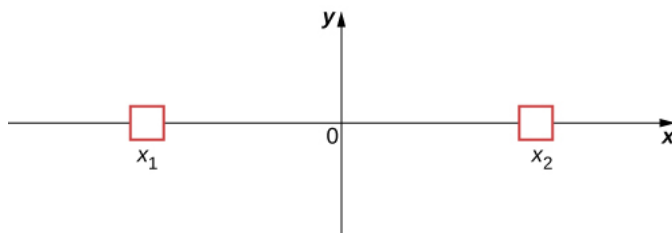
This momentum is much too small to be measured by any human instrument.

An Interpretation of the Wave Function

We are now in position to begin to answer the questions posed at the beginning of this section. First, for a traveling particle described by $\Psi(x, t) = A \sin(kx - \omega t)$, what is “waving?” Based on the above discussion, the answer is a mathematical function that can, among other things, be used to determine where the particle is likely to be when a position measurement is performed. Second, how is the wave function used to make predictions? If it is necessary to find the probability that a particle will be found in a certain interval, square the wave function and integrate over the interval of interest. Soon, you will learn soon that the wave function can be used to make many other kinds of predictions, as well.

Third, if a matter wave is given by the wave function $\Psi(x, t)$, *where* exactly is the particle? Two answers exist: (1) when the observer *is not* looking (or the particle is not being otherwise detected), the particle is everywhere ($x = -\infty, +\infty$); and (2) when the observer *is* looking (the particle is being detected), the particle “jumps into” a particular position state ($x, x + dx$) with a probability given by $P(x, x + dx) = |\Psi(x, t)|^2 dx$ —a process called **state reduction** or **wave function collapse**. This answer is called the **Copenhagen interpretation** of the wave function, or of quantum mechanics.

To illustrate this interpretation, consider the simple case of a particle that can occupy a small container either at x_1 or x_2 ([link](#)). In classical physics, we assume the particle is located either at x_1 or x_2 when the observer is not looking. However, in quantum mechanics, the particle may exist in a state of indefinite position—that is, it may be located at x_1 *and* x_2 when the observer is not looking. The assumption that a particle can only have one value of position (when the observer is not looking) is abandoned. Similar comments can be made of other measurable quantities, such as momentum and energy.

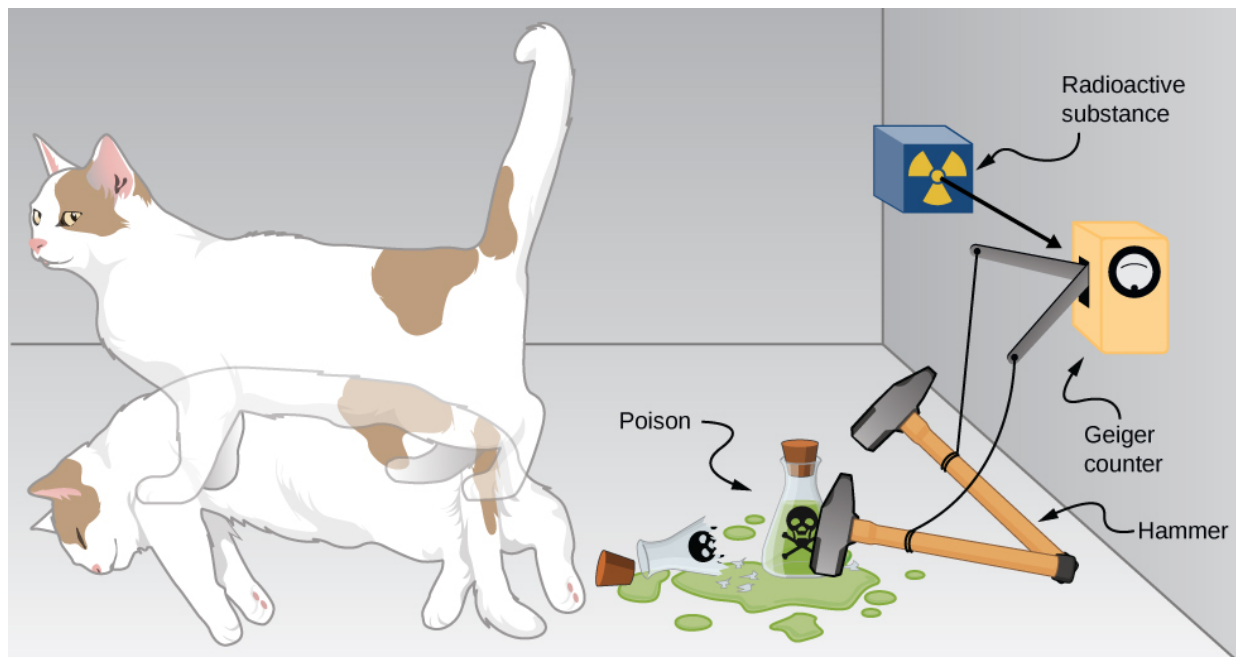


A two-state system of position of a particle.

The bizarre consequences of the Copenhagen interpretation of quantum mechanics are illustrated by a creative thought experiment first articulated by Erwin Schrödinger (*National Geographic*, 2013) ([link](#)):

“A cat is placed in a steel box along with a Geiger counter, a vial of poison, a hammer, and a radioactive substance. When the radioactive substance decays, the Geiger detects it and triggers the hammer to release the poison, which subsequently kills the cat. The radioactive decay is a random [probabilistic] process, and there is no way to predict

when it will happen. Physicists say the atom exists in a state known as a superposition—both decayed and not decayed at the same time. Until the box is opened, an observer doesn't know whether the cat is alive or dead—because the cat's fate is intrinsically tied to whether or not the atom has decayed and the cat would [according to the Copenhagen interpretation] be “living and dead ... in equal parts” until it is observed.”



Schrödinger's cat.

Schrödinger took the absurd implications of this thought experiment (a cat simultaneously dead and alive) as an argument against the Copenhagen interpretation. However, this interpretation remains the most commonly taught view of quantum mechanics.

Two-state systems (left and right, atom decays and does not decay, and so on) are often used to illustrate the principles of quantum mechanics. These systems find many applications in nature, including electron spin and mixed states of particles, atoms, and even molecules. Two-state systems are also finding application in the quantum computer, as mentioned in the introduction of this chapter. Unlike a digital computer, which encodes information in binary digits (zeroes and ones), a quantum computer stores and manipulates data in the form of quantum bits, or qubits. In general, a qubit is not in a state of zero or one, but rather in a mixed state of zero *and* one. If a large number of qubits are placed in the same quantum state, the measurement of an individual qubit would produce a zero with a probability p , and a one with a probability $q = 1 - p$. Many scientists believe that quantum computers are the future of the computer industry.

Complex Conjugates

Later in this section, you will see how to use the wave function to describe particles that are “free” or bound by forces to other particles. The specific form of the wave function depends on the details of the physical system. A peculiarity of quantum theory is that these functions are usually **complex functions**. A complex function is one that contains one or more imaginary numbers ($i = \sqrt{-1}$). Experimental measurements produce real (nonimaginary) numbers only, so the above procedure to use the wave function must be slightly modified. In general, the probability that a particle is found in the narrow interval $(x, x + dx)$ at time t is given by

Note:

Equation:

$$P(x, x + dx) = |\Psi(x, t)|^2 dx = \Psi^*(x, t) \Psi(x, t) dx,$$

where $\Psi^*(x, t)$ is the complex conjugate of the wave function. The complex conjugate of a function is obtained by replacing every occurrence of $i = \sqrt{-1}$ in that function with $-i$. This procedure eliminates complex numbers in all predictions because the product $\Psi^*(x, t) \Psi(x, t)$ is always a real number.

Note:

Exercise:

Problem: Check Your Understanding If $a = 3 + 4i$, what is the product $a^* a$?

Solution:

$$(3 + 4i)(3 - 4i) = 9 - 16i^2 = 25$$

Consider the motion of a free particle that moves along the x -direction. As the name suggests, a free particle experiences no forces and so moves with a constant velocity. As we will see in a later section of this chapter, a formal quantum mechanical treatment of a free particle indicates that its wave function has real *and* complex parts. In particular, the wave function is given by

Equation:

$$\Psi(x, t) = A \cos(kx - \omega t) + iA \sin(kx - \omega t),$$

where A is the amplitude, k is the wave number, and ω is the angular frequency. Using Euler's formula, $e^{i\phi} = \cos(\phi) + i \sin(\phi)$, this equation can be written in the form

Equation:

$$\Psi(x, t) = Ae^{i(kx - \omega t)} = Ae^{i\phi},$$

where ϕ is the phase angle. If the wave function varies slowly over the interval Δx , the probability of finding the particle in that interval is

Equation:

$$P(x, x + \Delta x) \approx \Psi^*(x, t) \Psi(x, t) \Delta x = (Ae^{i\phi}) (A^* e^{-i\phi}) \Delta x = (A^* A) \Delta x.$$

If A has real and complex parts ($a + ib$, where a and b are real constants), then

Equation:

$$A^* A = (a + ib)(a - ib) = a^2 + b^2.$$

Notice that the complex numbers have vanished. Thus,

Equation:

$$P(x, x + \Delta x) \approx |A|^2 \Delta x$$

is a real quantity. The interpretation of $\Psi^*(x, t)\Psi(x, t)$ as a probability density ensures that the predictions of quantum mechanics can be checked in the “real world.”

Note:

Exercise:

Problem:

Check Your Understanding Suppose that a particle with energy E is moving along the x -axis and is confined in the region between 0 and L . One possible wave function is

Equation:

$$\psi(x, t) = \begin{cases} Ae^{-iEt/\hbar} \sin \frac{\pi x}{L}, & \text{when } 0 \leq x \leq L \\ 0, & \text{otherwise} \end{cases}.$$

Determine the normalization constant.

Solution:

$$A = \sqrt{2/L}$$

Expectation Values

In classical mechanics, the solution to an equation of motion is a function of a measurable quantity, such as $x(t)$, where x is the position and t is the time. Note that the particle has one value of position for any time t . In quantum mechanics, however, the solution to an equation of motion is a wave function, $\Psi(x, t)$. The particle has many values of position for any time t , and only the probability density of finding the particle, $|\Psi(x, t)|^2$, can be known. The average value of position for a large number of particles with the same wave function is expected to be

Equation:

$$\langle x \rangle = \int_{-\infty}^{\infty} xP(x, t)dx = \int_{-\infty}^{\infty} x\Psi^*(x, t)\Psi(x, t)dx.$$

This is called the **expectation value** of the position. It is usually written

Note:

Equation:

$$\langle x \rangle = \int_{-\infty}^{\infty} \Psi^*(x, t)x\Psi(x, t)dx,$$

where the x is sandwiched between the wave functions. The reason for this will become apparent soon. Formally, x is called the **position operator**.

At this point, it is important to stress that a wave function can be written in terms of other quantities as well, such as velocity (v), momentum (p), and kinetic energy (K). The expectation value of momentum, for example, can be written

Equation:

$$\langle p \rangle = \int_{-\infty}^{\infty} \Psi^*(p, t) p \Psi(p, t) dp,$$

Where dp is used instead of dx to indicate an infinitesimal interval in momentum. In some cases, we know the wave function in position, $\Psi(x, t)$, but seek the expectation of momentum. The procedure for doing this is

Equation:

$$\langle p \rangle = \int_{-\infty}^{\infty} \Psi^*(x, t) \left(-i\hbar \frac{d}{dx} \right) \Psi(x, t) dx,$$

where the quantity in parentheses, sandwiched between the wave functions, is called the **momentum operator** in the x -direction. [The momentum operator in [\[link\]](#) is said to be the position-space representation of the momentum operator.] The momentum operator must act (operate) on the wave function to the right, and then the result must be multiplied by the complex conjugate of the wave function on the left, before integration. The momentum operator in the x -direction is sometimes denoted

Equation:

$$(p_x)_{\text{op}} = -i\hbar \frac{d}{dx},$$

Momentum operators for the y - and z -directions are defined similarly. This operator and many others are derived in a more advanced course in modern physics. In some cases, this derivation is relatively simple. For example, the kinetic energy operator is just

Equation:

$$(K)_{\text{op}} = \frac{1}{2} m (v_x)_{\text{op}}^2 = \frac{(p_x)_{\text{op}}^2}{2m} = \frac{(-i\hbar \frac{d}{dx})^2}{2m} = \frac{-\hbar^2}{2m} \left(\frac{d}{dx} \right) \left(\frac{d}{dx} \right).$$

Thus, if we seek an expectation value of kinetic energy of a particle in one dimension, two successive ordinary derivatives of the wave function are required before integration.

Expectation-value calculations are often simplified by exploiting the symmetry of wave functions. Symmetric wave functions can be even or odd. An **even function** is a function that satisfies

Equation:

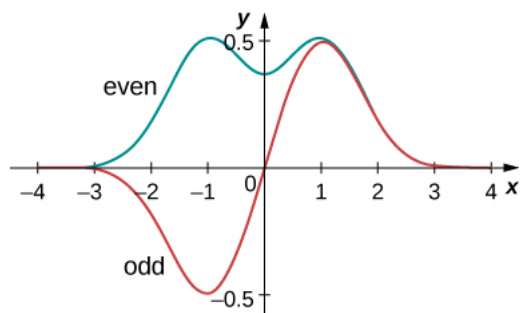
$$\psi(x) = \psi(-x).$$

In contrast, an **odd function** is a function that satisfies

Equation:

$$\psi(x) = -\psi(-x).$$

An example of even and odd functions is shown in [\[link\]](#). An even function is symmetric about the y-axis. This function is produced by reflecting $\psi(x)$ for $x > 0$ about the vertical y-axis. By comparison, an odd function is generated by reflecting the function about the y-axis and then about the x-axis. (An odd function is also referred to as an **anti-symmetric function**.)



Examples of even and odd wave functions.

In general, an even function times an even function produces an even function. A simple example of an even function is the product $x^2 e^{-x^2}$ (even times even is even). Similarly, an odd function times an odd function produces an even function, such as $x \sin x$ (odd times odd is even). However, an odd function times an even function produces an odd function, such as $x e^{-x^2}$ (odd times even is odd). The integral over all space of an odd function is zero, because the total area of the function above the x-axis cancels the (negative) area below it. As the next example shows, this property of odd functions is very useful.

Example:

Expectation Value (Part I)

The normalized wave function of a particle is

Equation:

$$\psi(x) = e^{-|x|/x_0} / \sqrt{x_0}.$$

Find the expectation value of position.

Strategy

Substitute the wave function into [\[link\]](#) and evaluate. The position operator introduces a multiplicative factor only, so the position operator need not be “sandwiched.”

Solution

First multiply, then integrate:

Equation:

$$\langle x \rangle = \int_{-\infty}^{+\infty} dx x |\psi(x)|^2 = \int_{-\infty}^{+\infty} dx x \left| \frac{e^{-|x|/x_0}}{\sqrt{x_0}} \right|^2 = \frac{1}{x_0} \int_{-\infty}^{+\infty} dx x e^{-2|x|/x_0} = 0.$$

Significance

The function in the integrand ($x e^{-2|x|/x_0}$) is odd since it is the product of an odd function (x) and an even function ($e^{-2|x|/x_0}$). The integral vanishes because the total area of the function above the x-axis cancels the (negative) area

below it. The result ($\langle x \rangle = 0$) is not surprising since the probability density function is symmetric about $x = 0$.

Example:

Expectation Value (Part II)

The time-dependent wave function of a particle confined to a region between 0 and L is

Equation:

$$\psi(x, t) = Ae^{-i\omega t} \sin(\pi x/L)$$

where ω is angular frequency and E is the energy of the particle. (*Note:* The function varies as a sine because of the limits (0 to L). When $x = 0$, the sine factor is zero and the wave function is zero, consistent with the boundary conditions.) Calculate the expectation values of position, momentum, and kinetic energy.

Strategy

We must first normalize the wave function to find A . Then we use the operators to calculate the expectation values.

Solution

Computation of the normalization constant:

Equation:

$$1 = \int_0^L dx \psi^*(x) \psi(x) = \int_0^L dx \left(Ae^{+i\omega t} \sin \frac{\pi x}{L} \right) \left(Ae^{-i\omega t} \sin \frac{\pi x}{L} \right) = A^2 \int_0^L dx \sin^2 \frac{\pi x}{L} = A^2 \frac{L}{2} \Rightarrow A = \sqrt{\frac{2}{L}}$$

The expectation value of position is

Equation:

$$\langle x \rangle = \int_0^L dx \psi^*(x) x \psi(x) = \int_0^L dx \left(Ae^{+i\omega t} \sin \frac{\pi x}{L} \right) x \left(Ae^{-i\omega t} \sin \frac{\pi x}{L} \right) = A^2 \int_0^L dx x \sin^2 \frac{\pi x}{L} = A^2 \frac{L^2}{4} = \frac{L}{2}.$$

The expectation value of momentum in the x -direction also requires an integral. To set this integral up, the associated operator must—by rule—act to the right on the wave function $\psi(x)$:

Equation:

$$-i\hbar \frac{d}{dx} \psi(x) = -i\hbar \frac{d}{dx} Ae^{-i\omega t} \sin \frac{\pi x}{L} = -i \frac{Ah}{2L} e^{-i\omega t} \cos \frac{\pi x}{L}.$$

Therefore, the expectation value of momentum is

Equation:

$$\langle p \rangle = \int_0^L dx \left(Ae^{+i\omega t} \sin \frac{\pi x}{L} \right) \left(-i \frac{Ah}{2L} e^{-i\omega t} \cos \frac{\pi x}{L} \right) = -i \frac{A^2 h}{4L} \int_0^L dx \sin \frac{2\pi x}{L} = 0.$$

The function in the integral is a sine function with a wavelength equal to the width of the well, L —an odd function about $x = L/2$. As a result, the integral vanishes.

The expectation value of kinetic energy in the x -direction requires the associated operator to act on the wave function:

Equation:

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} Ae^{-i\omega t} \sin \frac{\pi x}{L} = -\frac{\hbar^2}{2m} Ae^{-i\omega t} \frac{d^2}{dx^2} \sin \frac{\pi x}{L} = \frac{Ah^2}{8mL^2} e^{-i\omega t} \sin \frac{\pi x}{L}.$$

Thus, the expectation value of the kinetic energy is

Equation:

$$\begin{aligned}
\langle K \rangle &= \int_0^L dx \left(A e^{+i\omega t} \sin \frac{\pi x}{L} \right) \left(\frac{A h^2}{8mL^2} e^{-i\omega t} \sin \frac{\pi x}{L} \right) \\
&= \frac{A^2 h^2}{8mL^2} \int_0^L dx \sin^2 \frac{\pi x}{L} = \frac{A^2 h^2}{8mL^2} \frac{L}{2} = \frac{h^2}{8mL^2}.
\end{aligned}$$

Significance

The average position of a large number of particles in this state is $L/2$. The average momentum of these particles is zero because a given particle is equally likely to be moving right or left. However, the particle is not at rest because its average kinetic energy is not zero. Finally, the probability density is

Equation:

$$|\psi|^2 = (2/L) \sin^2(\pi x/L).$$

This probability density is largest at location $L/2$ and is zero at $x = 0$ and at $x = L$. Note that these conclusions do not depend explicitly on time.

Note:**Exercise:****Problem:**

Check Your Understanding For the particle in the above example, find the probability of locating it between positions 0 and $L/4$

Solution:

$$(1/2 - 1/\pi)/2 = 9\%$$

Quantum mechanics makes many surprising predictions. However, in 1920, Niels Bohr (founder of the Niels Bohr Institute in Copenhagen, from which we get the term “Copenhagen interpretation”) asserted that the predictions of quantum mechanics and classical mechanics must agree for all macroscopic systems, such as orbiting planets, bouncing balls, rocking chairs, and springs. This **correspondence principle** is now generally accepted. It suggests the rules of classical mechanics are an approximation of the rules of quantum mechanics for systems with very large energies. Quantum mechanics describes both the microscopic and macroscopic world, but classical mechanics describes only the latter.

Summary

- In quantum mechanics, the state of a physical system is represented by a wave function.
- In Born’s interpretation, the square of the particle’s wave function represents the probability density of finding the particle around a specific location in space.
- Wave functions must first be normalized before using them to make predictions.
- The expectation value is the average value of a quantity that requires a wave function and an integration.

Conceptual Questions

Exercise:**Problem:**

What is the physical unit of a wave function, $\Psi(x, t)$? What is the physical unit of the square of this wave function?

Solution:

$1/\sqrt{L}$, where $L = \text{length}$; $1/L$, where $L = \text{length}$

Exercise:

Problem: Can the magnitude of a wave function ($\Psi^*(x, t) \Psi(x, t)$) be a negative number? Explain.

Exercise:

Problem: What kind of physical quantity does a wave function of an electron represent?

Solution:

The wave function does not correspond directly to any measured quantity. It is a tool for predicting the values of physical quantities.

Exercise:

Problem: What is the physical meaning of a wave function of a particle?

Exercise:

Problem: What is the meaning of the expression “expectation value?” Explain.

Solution:

The average value of the physical quantity for a large number of particles with the same wave function.

Problems**Exercise:**

Problem: Compute $|\Psi(x, t)|^2$ for the function $\Psi(x, t) = \psi(x) \sin \omega t$, where ω is a real constant.

Solution:

$$|\psi(x)|^2 \sin^2 \omega t$$

Exercise:

Problem: Given the complex-valued function $f(x, y) = (x - iy)/(x + iy)$, calculate $|f(x, y)|^2$.

Exercise:

Problem:

Which one of the following functions, and why, qualifies to be a wave function of a particle that can move along the entire real axis? (a) $\psi(x) = Ae^{-x^2}$; (b) $\psi(x) = Ae^{-x}$; (c) $\psi(x) = A \tan x$; (d) $\psi(x) = A(\sin x)/x$; (e) $\psi(x) = Ae^{-|x|}$.

Solution:

(a) and (e), can be normalized

Exercise:**Problem:**

A particle with mass m moving along the x -axis and its quantum state is represented by the following wave function:

$$\Psi(x, t) = \begin{cases} 0, & x < 0, \\ Axe^{-\alpha x} e^{-iEt/\hbar}, & x \geq 0, \end{cases}$$

where $\alpha = 2.0 \times 10^{10} \text{ m}^{-1}$. (a) Find the normalization constant. (b) Find the probability that the particle can be found on the interval $0 \leq x \leq L$. (c) Find the expectation value of position. (d) Find the expectation value of kinetic energy.

Exercise:

Problem: A wave function of a particle with mass m is given by

$$\psi(x) = \begin{cases} A \cos \alpha x, & -\frac{\pi}{2\alpha} \leq x \leq +\frac{\pi}{2\alpha}, \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha = 1.00 \times 10^{10} \text{ m}^{-1}$. (a) Find the normalization constant. (b) Find the probability that the particle can be found on the interval $0 \leq x \leq 0.5 \times 10^{-10} \text{ m}$. (c) Find the particle's average position. (d) Find its average momentum. (e) Find its average kinetic energy $-0.5 \times 10^{-10} \text{ m} \leq x \leq +0.5 \times 10^{-10} \text{ m}$.

Solution:

a. $A = \sqrt{2\alpha/\pi}$; b. probability = 29.3%; c. $\langle x \rangle = 0$; d. $\langle p \rangle = 0$; e. $\langle K \rangle = \alpha^2 \hbar^2 / 2m$

Glossary

anti-symmetric function
odd function

Born interpretation
states that the square of a wave function is the probability density

complex function
function containing both real and imaginary parts

Copenhagen interpretation
states that when an observer *is not* looking or when a measurement is not being made, the particle has many values of measurable quantities, such as position

correspondence principle

in the limit of large energies, the predictions of quantum mechanics agree with the predictions of classical mechanics

expectation value

average value of the physical quantity assuming a large number of particles with the same wave function

even function

in one dimension, a function symmetric with the origin of the coordinate system

momentum operator

operator that corresponds to the momentum of a particle

normalization condition

requires that the probability density integrated over the entire physical space results in the number one

odd function

in one dimension, a function antisymmetric with the origin of the coordinate system

position operator

operator that corresponds to the position of a particle

probability density

square of the particle's wave function

state reduction

hypothetical process in which an observed or detected particle "jumps into" a definite state, often described in terms of the collapse of the particle's wave function

wave function

function that represents the quantum state of a particle (quantum system)

wave function collapse

equivalent to state reduction

The Heisenberg Uncertainty Principle

By the end of this section, you will be able to:

- Describe the physical meaning of the position-momentum uncertainty relation
- Explain the origins of the uncertainty principle in quantum theory
- Describe the physical meaning of the energy-time uncertainty relation

Heisenberg's uncertainty principle is a key principle in quantum mechanics. Very roughly, it states that if we know *everything* about where a particle is located (the uncertainty of position is small), we know *nothing* about its momentum (the uncertainty of momentum is large), and vice versa. Versions of the uncertainty principle also exist for other quantities as well, such as energy and time. We discuss the momentum-position and energy-time uncertainty principles separately.

Momentum and Position

To illustrate the momentum-position uncertainty principle, consider a free particle that moves along the x -direction. The particle moves with a constant velocity u and momentum $p = mu$. According to de Broglie's relations, $p = \hbar k$ and $E = \hbar\omega$. As discussed in the previous section, the wave function for this particle is given by

Equation:

$$\psi_k(x, t) = A[\cos(\omega t - kx) - i \sin(\omega t - kx)] = Ae^{-i(\omega t - kx)} = Ae^{-i\omega t}e^{ikx}$$

and the probability density $|\psi_k(x, t)|^2 = A^2$ is *uniform* and independent of time. The particle is equally likely to be found anywhere along the x -axis but has definite values of wavelength and wave number, and therefore momentum. The uncertainty of position is infinite (we are completely uncertain about position) and the uncertainty of the momentum is zero (we are completely certain about momentum). This account of a free particle is consistent with Heisenberg's uncertainty principle.

Similar statements can be made of localized particles. In quantum theory, a localized particle is modeled by a linear superposition of free-particle (or plane-wave) states called a **wave packet**. An example of a wave packet is shown in [\[link\]](#). A wave packet contains many wavelengths and therefore by de Broglie's relations many momenta—possible in quantum mechanics! This particle also has many values of position, although the particle is confined mostly to the interval Δx . The particle can be better localized (Δx can be decreased) if more plane-wave states of different wavelengths or momenta are added together in the right way (Δp is increased). According to Heisenberg, these uncertainties obey the following relation.

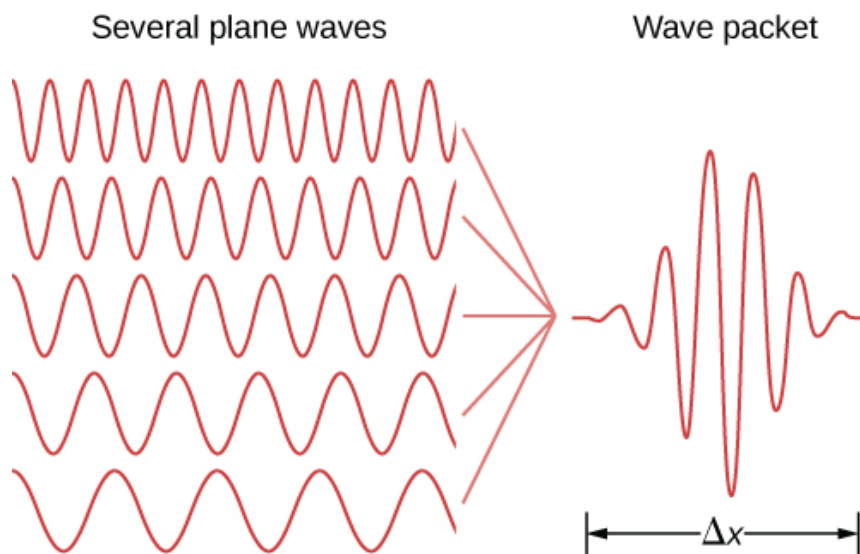
Note:**The Heisenberg Uncertainty Principle**

The product of the uncertainty in position of a particle and the uncertainty in its momentum can never be less than one-half of the reduced Planck constant:

Equation:

$$\Delta x \Delta p \geq \hbar/2.$$

This relation expresses Heisenberg's uncertainty principle. It places limits on what we can know about a particle from simultaneous measurements of position and momentum. If Δx is large, Δp is small, and vice versa. [\[link\]](#) can be derived in a more advanced course in modern physics. Reflecting on this relation in his work *The Physical Principles of the Quantum Theory*, Heisenberg wrote "Any use of the words 'position' and 'velocity' with accuracy exceeding that given by [the relation] is just as meaningless as the use of words whose sense is not defined."



Adding together several plane waves of different wavelengths can produce a wave that is relatively localized.

Note that the uncertainty principle has nothing to do with the precision of an experimental apparatus. Even for perfect measuring devices, these uncertainties would remain because they originate in the wave-like nature of matter. The precise value of the product $\Delta x \Delta p$ depends on the specific form of the wave function. Interestingly, the Gaussian function (or bell-curve distribution) gives the minimum value of the uncertainty product: $\Delta x \Delta p = \hbar/2$.

Example:**The Uncertainty Principle Large and Small**

Determine the minimum uncertainties in the positions of the following objects if their speeds are known with a precision of $1.0 \times 10^{-3} \text{ m/s}$: (a) an electron and (b) a bowling ball of mass 6.0 kg.

Strategy

Given the uncertainty in speed $\Delta u = 1.0 \times 10^{-3} \text{ m/s}$, we have to first determine the uncertainty in momentum $\Delta p = m \Delta u$ and then invert [\[link\]](#) to find the uncertainty in position $\Delta x = \hbar/(2\Delta p)$.

Solution

- a. For the electron:

Equation:

$$\begin{aligned}\Delta p &= m\Delta u = (9.1 \times 10^{-31} \text{ kg})(1.0 \times 10^{-3} \text{ m/s}) = 9.1 \times 10^{-34} \text{ kg} \cdot \text{m/s}, \\ \Delta x &= \frac{\hbar}{2\Delta p} = 5.8 \text{ cm}.\end{aligned}$$

- b. For the bowling ball:

Equation:

$$\begin{aligned}\Delta p &= m\Delta u = (6.0 \text{ kg})(1.0 \times 10^{-3} \text{ m/s}) = 6.0 \times 10^{-3} \text{ kg} \cdot \text{m/s}, \\ \Delta x &= \frac{\hbar}{2\Delta p} = 8.8 \times 10^{-33} \text{ m}.\end{aligned}$$

Significance

Unlike the position uncertainty for the electron, the position uncertainty for the bowling ball is immeasurably small. Planck's constant is very small, so the limitations imposed by the uncertainty principle are not noticeable in macroscopic systems such as a bowling ball.

Example:**Uncertainty and the Hydrogen Atom**

Estimate the ground-state energy of a hydrogen atom using Heisenberg's uncertainty principle. (*Hint:* According to early experiments, the size of a hydrogen atom is approximately 0.1 nm.)

Strategy

An electron bound to a hydrogen atom can be modeled by a particle bound to a one-dimensional box of length $L = 0.1$ nm. The ground-state wave function of this system is a half wave, like that given in [\[link\]](#). This is the largest wavelength that can “fit” in the box, so the wave function corresponds to the lowest energy state. Note that this function is very similar in shape to a Gaussian (bell curve) function. We can take the average energy of a particle described by this function (E) as a good estimate of the ground state energy (E_0). This average energy of a particle is related to its average of the momentum squared, which is related to its momentum uncertainty.

Solution

To solve this problem, we must be specific about what is meant by “uncertainty of position” and “uncertainty of momentum.” We identify the uncertainty of position (Δx) with the standard deviation of position (σ_x), and the uncertainty of momentum (Δp) with the standard deviation of momentum (σ_p). For the Gaussian function, the uncertainty product is

Equation:

$$\sigma_x \sigma_p = \frac{\hbar}{2},$$

where

Equation:

$$\sigma_x^2 = x^2 - \bar{x}^2 \text{ and } \sigma_p^2 = p^2 - \bar{p}^2.$$

The particle is equally likely to be moving left as moving right, so $\bar{p} = 0$. Also, the uncertainty of position is comparable to the size of the box, so $\sigma_x = L$. The estimated ground state energy is therefore

Equation:

$$E_0 = E_{\text{Gaussian}} = \frac{p^2}{m} = \frac{\sigma_p^2}{2m} = \frac{1}{2m} \left(\frac{\hbar}{2\sigma_x} \right)^2 = \frac{1}{2m} \left(\frac{\hbar}{2L} \right)^2 = \frac{\hbar^2}{8mL^2}.$$

Multiplying numerator and denominator by c^2 gives

Equation:

$$E_0 = \frac{(\hbar c)^2}{8(mc^2)L^2} = \frac{(197.3 \text{ eV} \cdot \text{nm})^2}{8 (0.511 \cdot 10^6 \text{ eV})(0.1 \text{ nm})^2} = 0.952 \text{ eV} \approx 1 \text{ eV}.$$

Significance

Based on early estimates of the size of a hydrogen atom and the uncertainty principle, the ground-state energy of a hydrogen atom is in the eV range. The ionization energy of an electron in the ground-state energy is approximately 10 eV, so this prediction is roughly confirmed. (*Note:* The product $\hbar c$ is often a useful value in performing calculations in quantum mechanics.)

Energy and Time

Another kind of uncertainty principle concerns uncertainties in simultaneous measurements of the energy of a quantum state and its lifetime,

Note:
Equation:

$$\Delta E \Delta t \geq \frac{\hbar}{2},$$

where ΔE is the uncertainty in the energy measurement and Δt is the uncertainty in the lifetime measurement. The **energy-time uncertainty principle** does not result from a relation of the type expressed by [\[link\]](#) for technical reasons beyond this discussion. Nevertheless, the general meaning of the energy-time principle is that a quantum state that exists for only a short time cannot have a definite energy. The reason is that the frequency of a state is inversely proportional to time and the frequency connects with the energy of the state, so to measure the energy with good precision, the state must be observed for many cycles.

To illustrate, consider the excited states of an atom. The finite lifetimes of these states can be deduced from the shapes of spectral lines observed in atomic emission spectra. Each time an excited state decays, the emitted energy is slightly different and, therefore, the emission line is characterized by a *distribution* of spectral frequencies (or wavelengths) of the emitted photons. As a result, all spectral lines are characterized by spectral widths. The average energy of the emitted photon corresponds to the theoretical energy of the excited state and gives the spectral location of the peak of the emission line. Short-lived states have broad spectral widths and long-lived states have narrow spectral widths.

Example:**Atomic Transitions**

An atom typically exists in an excited state for about $\Delta t = 10^{-8}$ s. Estimate the uncertainty Δf in the frequency of emitted photons when an atom makes a transition from an excited state with the simultaneous emission of a photon with an average frequency of $f = 7.1 \times 10^{14}$ Hz. Is the emitted radiation monochromatic?

Strategy

We invert [\[link\]](#) to obtain the energy uncertainty $\Delta E \approx \hbar/2\Delta t$ and combine it with the photon energy $E = hf$ to obtain Δf . To estimate whether or not the emission is monochromatic, we evaluate $\Delta f/f$.

Solution

The spread in photon energies is $\Delta E = h\Delta f$. Therefore,

Equation:

$$\begin{aligned}\Delta E &\approx \frac{\hbar}{2\Delta t} \Rightarrow h\Delta f \approx \frac{\hbar}{2\Delta t} \Rightarrow \Delta f \approx \frac{1}{4\pi\Delta t} = \frac{1}{4\pi(10^{-8}\text{s})} = 8.0 \times 10^6 \text{ Hz}, \\ \frac{\Delta f}{f} &= \frac{8.0 \times 10^6 \text{ Hz}}{7.1 \times 10^{14} \text{ Hz}} = 1.1 \times 10^{-8}.\end{aligned}$$

Significance

Because the emitted photons have their frequencies within 1.1×10^{-6} percent of the average frequency, the emitted radiation can be considered monochromatic.

Note:**Exercise:****Problem:**

Check Your Understanding A sodium atom makes a transition from the first excited state to the ground state, emitting a 589.0-nm photon with energy 2.105 eV. If the lifetime of this excited state is 1.6×10^{-8} s, what is the uncertainty in energy of this excited state? What is the width of the corresponding spectral line?

Solution:

$$4.1 \times 10^{-8} \text{ eV}; 1.1 \times 10^{-5} \text{ nm}$$

Summary

- The Heisenberg uncertainty principle states that it is impossible to simultaneously measure the x-components of position and of momentum of a particle with an

arbitrarily high precision. The product of experimental uncertainties is always larger than or equal to $\hbar/2$.

- The limitations of this principle have nothing to do with the quality of the experimental apparatus but originate in the wave-like nature of matter.
- The energy-time uncertainty principle expresses the experimental observation that a quantum state that exists only for a short time cannot have a definite energy.

Conceptual Questions

Exercise:

Problem:

If the formalism of quantum mechanics is ‘more exact’ than that of classical mechanics, why don’t we use quantum mechanics to describe the motion of a leaping frog? Explain.

Exercise:

Problem:

Can the de Broglie wavelength of a particle be known precisely? Can the position of a particle be known precisely?

Solution:

Yes, if its position is completely unknown. Yes, if its momentum is completely unknown.

Exercise:

Problem:

Can we measure the energy of a free localized particle with complete precision?

Exercise:

Problem:

Can we measure both the position and momentum of a particle with complete precision?

Solution:

No. According to the uncertainty principle, if the uncertainty on the particle’s position is small, the uncertainty on its momentum is large. Similarly, if the uncertainty on the particle’s position is large, the uncertainty on its momentum is small.

Problems

Exercise:

Problem:

A velocity measurement of an α -particle has been performed with a precision of 0.02 mm/s. What is the minimum uncertainty in its position?

Exercise:

Problem:

A gas of helium atoms at 273 K is in a cubical container with 25.0 cm on a side. (a) What is the minimum uncertainty in momentum components of helium atoms? (b) What is the minimum uncertainty in velocity components? (c) Find the ratio of the uncertainties in (b) to the mean speed of an atom in each direction.

Solution:

$$\text{a. } \Delta p \geq 2.11 \times 10^{-34} \text{ N} \cdot \text{s}; \text{ b. } \Delta v \geq 6.31 \times 10^{-8} \text{ m/s}; \text{ c. } \Delta v / \sqrt{k_B T / m_\alpha} = 5.94 \times 10^{-11}$$

Exercise:

Problem:

If the uncertainty in the y -component of a proton's position is 2.0 pm, find the minimum uncertainty in the simultaneous measurement of the proton's y -component of velocity. What is the minimum uncertainty in the simultaneous measurement of the proton's x -component of velocity?

Exercise:

Problem:

Some unstable elementary particle has a rest energy of 80.41 GeV and an uncertainty in rest energy of 2.06 GeV. Estimate the lifetime of this particle.

Solution:

$$\Delta \tau \geq 1.6 \times 10^{-25} \text{ s}$$

Exercise:

Problem:

An atom in a metastable state has a lifetime of 5.2 ms. Find the minimum uncertainty in the measurement of energy of the excited state.

Exercise:**Problem:**

Measurements indicate that an atom remains in an excited state for an average time of 50.0 ns before making a transition to the ground state with the simultaneous emission of a 2.1-eV photon. (a) Estimate the uncertainty in the frequency of the photon. (b) What fraction of the photon's average frequency is this?

Solution:

a. $\Delta f \geq 1.59 \text{ MHz}$; b. $\Delta\omega/\omega_0 = 3.135 \times 10^{-9}$

Exercise:**Problem:**

Suppose an electron is confined to a region of length 0.1 nm (of the order of the size of a hydrogen atom). (a) What is the minimum uncertainty of its momentum? (b) What would the uncertainty in momentum be if the confined length region doubled to 0.2 nm?

Glossary

Heisenberg's uncertainty principle

places limits on what can be known from a simultaneous measurements of position and momentum; states that if the uncertainty on position is small then the uncertainty on momentum is large, and vice versa

wave packet

superposition of many plane matter waves that can be used to represent a localized particle

energy-time uncertainty principle

energy-time relation for uncertainties in the simultaneous measurements of the energy of a quantum state and of its lifetime

The Schrödinger Equation

By the end of this section, you will be able to:

- Describe the role Schrödinger's equation plays in quantum mechanics
- Explain the difference between time-dependent and -independent Schrödinger's equations
- Interpret the solutions of Schrödinger's equation

In the preceding two sections, we described how to use a quantum mechanical wave function and discussed Heisenberg's uncertainty principle. In this section, we present a complete and formal theory of quantum mechanics that can be used to make predictions. In developing this theory, it is helpful to review the wave theory of light. For a light wave, the electric field $E(x,t)$ obeys the relation

Equation:

$$\frac{\partial^2 E}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2},$$

where c is the speed of light and the symbol ∂ represents a *partial derivative*. (Recall from [Oscillations](#) that a partial derivative is closely related to an ordinary derivative, but involves functions of more than one variable. When taking the partial derivative of a function by a certain variable, all other variables are held constant.) A light wave consists of a very large number of photons, so the quantity $|E(x, t)|^2$ can be interpreted as a probability density of finding a single photon at a particular point in space (for example, on a viewing screen).

There are many solutions to this equation. One solution of particular importance is

Equation:

$$E(x, t) = A \sin(kx - \omega t),$$

where A is the amplitude of the electric field, k is the wave number, and ω is the angular frequency. Combining this equation with [\[link\]](#) gives

Equation:

$$k^2 = \frac{\omega^2}{c^2}.$$

According to de Broglie's equations, we have $p = \hbar k$ and $E = \hbar \omega$. Substituting these equations in [\[link\]](#) gives

Equation:

$$p = \frac{E}{c},$$

or

Equation:

$$E = pc.$$

Therefore, according to Einstein's general energy-momentum equation ([\[link\]](#)), [\[link\]](#) describes a particle with a zero rest mass. This is consistent with our knowledge of a photon.

This process can be reversed. We can begin with the energy-momentum equation of a particle and then ask what wave equation corresponds to it. The energy-momentum equation of a nonrelativistic particle in one dimension is

Equation:

$$E = \frac{p^2}{2m} + U(x, t),$$

where p is the momentum, m is the mass, and U is the potential energy of the particle. The wave equation that goes with it turns out to be a key

equation in quantum mechanics, called **Schrödinger's time-dependent equation**.

Note:

The Schrödinger Time-Dependent Equation

The equation describing the energy and momentum of a wave function is known as the Schrödinger equation:

Equation:

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} + U(x, t) \Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}.$$

As described in [Potential Energy and Conservation of Energy](#), the force on the particle described by this equation is given by

Equation:

$$F = -\frac{\partial U(x, t)}{\partial x}.$$

This equation plays a role in quantum mechanics similar to Newton's second law in classical mechanics. Once the potential energy of a particle is specified—or, equivalently, once the force on the particle is specified—we can solve this differential equation for the wave function. The solution to Newton's second law equation (also a differential equation) in one dimension is a function $x(t)$ that specifies where an object is at any time t . The solution to Schrödinger's time-dependent equation provides a tool—the wave function—that can be used to determine where the particle is *likely* to be. This equation can be also written in two or three dimensions. Solving Schrödinger's time-dependent equation often requires the aid of a computer.

Consider the special case of a free particle. A free particle experiences no force ($F = 0$). Based on [\[link\]](#), this requires only that

Equation:

$$U(x, t) = U_0 = \text{constant}.$$

For simplicity, we set $U_0 = 0$. Schrödinger's equation then reduces to

Equation:

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} = i\hbar \frac{\partial \Psi(x, t)}{\partial t}.$$

A valid solution to this equation is

Equation:

$$\Psi(x, t) = Ae^{i(kx - \omega t)}.$$

Not surprisingly, this solution contains an imaginary number ($i = \sqrt{-1}$) because the differential equation itself contains an imaginary number. As stressed before, however, quantum-mechanical predictions depend only on $|\Psi(x, t)|^2$, which yields completely real values. Notice that the real plane-wave solutions, $\Psi(x, t) = A \sin(kx - \omega t)$ and $\Psi(x, t) = A \cos(kx - \omega t)$, do not obey Schrödinger's equation. The temptation to think that a wave function can be seen, touched, and felt in nature is eliminated by the appearance of an imaginary number. In Schrödinger's theory of quantum mechanics, the wave function is merely a tool for calculating things.

If the potential energy function (U) does not depend on time, it is possible to show that

Note:

Equation:

$$\Psi(x, t) = \psi(x)e^{-i\omega t}$$

satisfies Schrödinger's time-dependent equation, where $\psi(x)$ is a *time-independent* function and $e^{-i\omega t}$ is a *space-independent* function. In other words, the wave function is *separable* into two parts: a space-only part and a time-only part. The factor $e^{-i\omega t}$ is sometimes referred to as a **time-modulation factor** since it modifies the space-only function. According to de Broglie, the energy of a matter wave is given by $E = \hbar\omega$, where E is its total energy. Thus, the above equation can also be written as

Equation:

$$\Psi(x, t) = \psi(x)e^{-iEt/\hbar}.$$

Any linear combination of such states (mixed state of energy or momentum) is also valid solution to this equation. Such states can, for example, describe a localized particle (see [\[link\]](#))

Note:

Exercise:

Problem:

Check Your Understanding A particle with mass m is moving along the x -axis in a potential given by the potential energy function $U(x) = 0.5m\omega^2x^2$. Compute the product $\Psi(x, t)^* U(x) \Psi(x, t)$. Express your answer in terms of the time-independent wave function, $\psi(x)$.

Solution:

$$0.5m\omega^2x^2\psi(x)^*\psi(x)$$

Combining [\[link\]](#) and [\[link\]](#), Schrödinger's time-dependent equation reduces to

Note:

Equation:

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x),$$

where E is the total energy of the particle (a real number). This equation is called **Schrödinger's time-independent equation**. Notice that we use “big psi” (Ψ) for the time-dependent wave function and “little psi” (ψ) for the time-independent wave function. The wave-function solution to this equation must be multiplied by the time-modulation factor to obtain the time-dependent wave function.

In the next sections, we solve Schrödinger's time-independent equation for three cases: a quantum particle in a box, a simple harmonic oscillator, and a quantum barrier. These cases provide important lessons that can be used to solve more complicated systems. The time-independent wave function $\psi(x)$ solutions must satisfy three conditions:

- $\psi(x)$ must be a continuous function.
- The first derivative of $\psi(x)$ with respect to space, $d\psi(x)/dx$, must be continuous, unless $V(x) = \infty$.
- $\psi(x)$ must not diverge (“blow up”) at $x = \pm\infty$.

The first condition avoids sudden jumps or gaps in the wave function. The second condition requires the wave function to be smooth at all points, except in special cases. (In a more advanced course on quantum mechanics, for example, potential spikes of infinite depth and height are used to model solids). The third condition requires the wave function be normalizable. This third condition follows from Born's interpretation of quantum

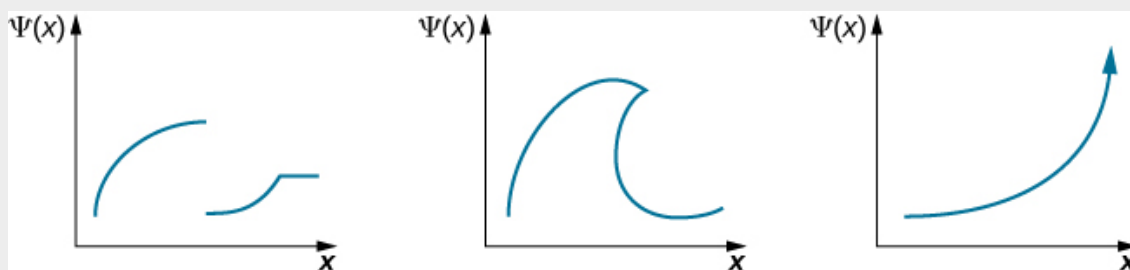
mechanics. It ensures that $|\psi(x)|^2$ is a finite number so we can use it to calculate probabilities.

Note:

Exercise:

Problem:

Check Your Understanding Which of the following wave functions is a valid wave-function solution for Schrödinger's equation?



Solution:

None. The first function has a discontinuity; the second function is double-valued; and the third function diverges so is not normalizable.

Summary

- The Schrödinger equation is the fundamental equation of wave quantum mechanics. It allows us to make predictions about wave functions.
- When a particle moves in a time-independent potential, a solution of the time-dependent Schrödinger equation is a product of a time-independent wave function and a time-modulation factor.
- The Schrödinger equation can be applied to many physical situations.

Conceptual Questions

Exercise:

Problem:

What is the difference between a wave function $\psi(x, y, z)$ and a wave function $\Psi(x, y, z, t)$ for the same particle?

Exercise:

Problem:

If a quantum particle is in a stationary state, does it mean that it does not move?

Solution:

No, it means that predictions about the particle (expressed in terms of probabilities) are time-independent.

Exercise:

Problem:

Explain the difference between time-dependent and -independent Schrödinger's equations.

Exercise:

Problem:

Suppose a wave function is discontinuous at some point. Can this function represent a quantum state of some physical particle? Why? Why not?

Solution:

No, because the probability of the particle existing in a narrow (infinitesimally small) interval at the discontinuity is undefined.

Problems

Exercise:

Problem: Combine [\[link\]](#) and [\[link\]](#) to show $k^2 = \frac{\omega^2}{c^2}$.

Solution:

Carrying out the derivatives yields $k^2 = \frac{\omega^2}{c^2}$.

Exercise:

Problem:

Show that $\Psi(x, t) = Ae^{i(kx - \omega t)}$ is a valid solution to Schrödinger's time-dependent equation.

Exercise:

Problem:

Show that $\Psi(x, t) = A \sin(kx - \omega t)$ and $\Psi(x, t) = A \cos(kx - \omega t)$ do not obey Schrödinger's time-dependent equation.

Solution:

Carrying out the derivatives (as above) for the sine function gives a cosine on the right side the equation, so the equality fails. The same occurs for the cosine solution.

Exercise:

Problem:

Show that when $\Psi_1(x, t)$ and $\Psi_2(x, t)$ are solutions to the time-dependent Schrödinger equation and A, B are numbers, then a function $\Psi(x, t)$ that is a superposition of these functions is also a solution:
 $\Psi(x, t) = A\Psi_1(x, t) + B\Psi_2(x, t)$.

Exercise:**Problem:**

A particle with mass m is described by the following wave function: $\psi(x) = A \cos kx + B \sin kx$, where A , B , and k are constants. Assuming that the particle is free, show that this function is the solution of the stationary Schrödinger equation for this particle and find the energy that the particle has in this state.

Solution:

$$E = \hbar^2 k^2 / 2m$$

Exercise:**Problem:**

Find the expectation value of the kinetic energy for the particle in the state, $\Psi(x, t) = Ae^{i(kx - \omega t)}$. What conclusion can you draw from your solution?

Exercise:**Problem:**

Find the expectation value of the square of the momentum squared for the particle in the state, $\Psi(x, t) = Ae^{i(kx - \omega t)}$. What conclusion can you draw from your solution?

Solution:

$\hbar^2 k^2$; The particle has definite momentum and therefore definite momentum squared.

Exercise:**Problem:**

A free proton has a wave function given by $\Psi(x, t) = Ae^{i(5.02 \times 10^{11}x - 8.00 \times 10^{15}t)}$.

The coefficient of x is inverse meters (m^{-1}) and the coefficient on t is inverse seconds (s^{-1}). Find its momentum and energy.

Glossary

Schrödinger's time-dependent equation

equation in space and time that allows us to determine wave functions of a quantum particle

Schrödinger's time-independent equation

equation in space that allows us to determine wave functions of a quantum particle; this wave function must be multiplied by a time-modulation factor to obtain the time-dependent wave function

time-modulation factor

factor $e^{-i\omega t}$ that multiplies the time-independent wave function when the potential energy of the particle is time independent

The Quantum Particle in a Box

By the end of this section, you will be able to:

- Describe how to set up a boundary-value problem for the stationary Schrödinger equation
- Explain why the energy of a quantum particle in a box is quantized
- Describe the physical meaning of stationary solutions to Schrödinger's equation and the connection of these solutions with time-dependent quantum states
- Explain the physical meaning of Bohr's correspondence principle

In this section, we apply Schrödinger's equation to a particle bound to a one-dimensional box. This special case provides lessons for understanding quantum mechanics in more complex systems. The energy of the particle is quantized as a consequence of a standing wave condition inside the box.

Consider a particle of mass m that is allowed to move only along the x -direction and its motion is confined to the region between hard and rigid walls located at $x = 0$ and at $x = L$ ([link](#)). Between the walls, the particle moves freely. This physical situation is called the **infinite square well**, described by the potential energy function

Equation:

$$U(x) = \begin{cases} 0, & 0 \leq x \leq L, \\ \infty, & \text{otherwise.} \end{cases}$$

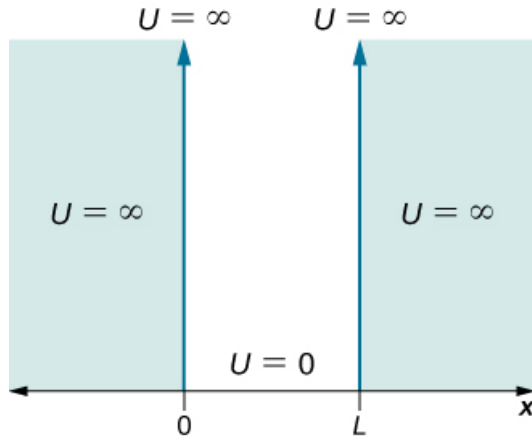
Combining this equation with Schrödinger's time-independent wave equation gives

Note:

Equation:

$$\frac{-\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} = E\psi(x), \text{ for } 0 \leq x \leq L$$

where E is the total energy of the particle. What types of solutions do we expect? The energy of the particle is a positive number, so if the value of the wave function is positive (right side of the equation), the curvature of the wave function is negative, or concave down (left side of the equation). Similarly, if the value of the wave function is negative (right side of the equation), the curvature of the wave function is positive or concave up (left side of equation). This condition is met by an oscillating wave function, such as a sine or cosine wave. Since these waves are confined to the box, we envision standing waves with fixed endpoints at $x = 0$ and $x = L$.



The potential energy function that confines the particle in a one-dimensional box.

Solutions $\psi(x)$ to this equation have a probabilistic interpretation. In particular, the square $|\psi(x)|^2$ represents the probability density of finding the particle at a particular location x . This function must be integrated to determine the probability of finding the particle in some interval of space. We are therefore looking for a normalizable solution that satisfies the following normalization condition:

Equation:

$$\int_0^L dx |\psi(x)|^2 = 1.$$

The walls are rigid and impenetrable, which means that the particle is never found beyond the wall. Mathematically, this means that the solution must vanish at the walls:

Equation:

$$\psi(0) = \psi(L) = 0.$$

We expect oscillating solutions, so the most general solution to this equation is

Equation:

$$\psi_k(x) = A_k \cos kx + B_k \sin kx$$

where k is the wave number, and A_k and B_k are constants. Applying the boundary condition expressed by [\[link\]](#) gives

Equation:

$$\psi_k(0) = A_k \cos(k \cdot 0) + B_k \sin(k \cdot 0) = A_k = 0.$$

Because we have $A_k = 0$, the solution must be

Equation:

$$\psi_k(x) = B_k \sin kx.$$

If B_k is zero, $\psi_k(x) = 0$ for all values of x and the normalization condition, [\[link\]](#), cannot be satisfied. Assuming $B_k \neq 0$, [\[link\]](#) for $x = L$ then gives

Equation:

$$0 = B_k \sin(kL) \Rightarrow \sin(kL) = 0 \Rightarrow kL = n\pi, n = 1, 2, 3, \dots$$

We discard the $n = 0$ solution because $\psi(x)$ for this quantum number would be zero everywhere—an un-normalizable and therefore unphysical solution. Substituting [\[link\]](#) into [\[link\]](#) gives

Equation:

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} (B_k \sin(kx)) = E (B_k \sin(kx)).$$

Computing these derivatives leads to

Equation:

$$E = E_k = \frac{\hbar^2 k^2}{2m}.$$

According to de Broglie, $p = \hbar k$, so this expression implies that the total energy is equal to the kinetic energy, consistent with our assumption that the “particle moves freely.” Combining the results of [\[link\]](#) and [\[link\]](#) gives

Note:

Equation:

$$E_n = n^2 \frac{\pi^2 \hbar^2}{2mL^2}, n = 1, 2, 3, \dots$$

Strange! A particle bound to a one-dimensional box can only have certain discrete (quantized) values of energy. Further, the particle cannot have a zero kinetic energy—it is impossible for a

particle bound to a box to be “at rest.”

To evaluate the allowed wave functions that correspond to these energies, we must find the normalization constant B_n . We impose the normalization condition [\[link\]](#) on the wave function

Equation:

$$\psi_n(x) = B_n \sin n\pi x / L$$

Equation:

$$1 = \int_0^L dx |\psi_n(x)|^2 = \int_0^L dx B_n^2 \sin^2 \frac{n\pi}{L} x = B_n^2 \int_0^L dx \sin^2 \frac{n\pi}{L} x = B_n^2 \frac{L}{2} \Rightarrow B_n = \sqrt{\frac{2}{L}}.$$

Hence, the wave functions that correspond to the energy values given in [\[link\]](#) are

Note:

Equation:

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L}, n = 1, 2, 3, \dots$$

For the lowest energy state or **ground state energy**, we have

Equation:

$$E_1 = \frac{\pi^2 \hbar^2}{2mL^2}, \psi_1(x) = \sqrt{\frac{2}{L}} \sin \left(\frac{\pi x}{L} \right).$$

All other energy states can be expressed as

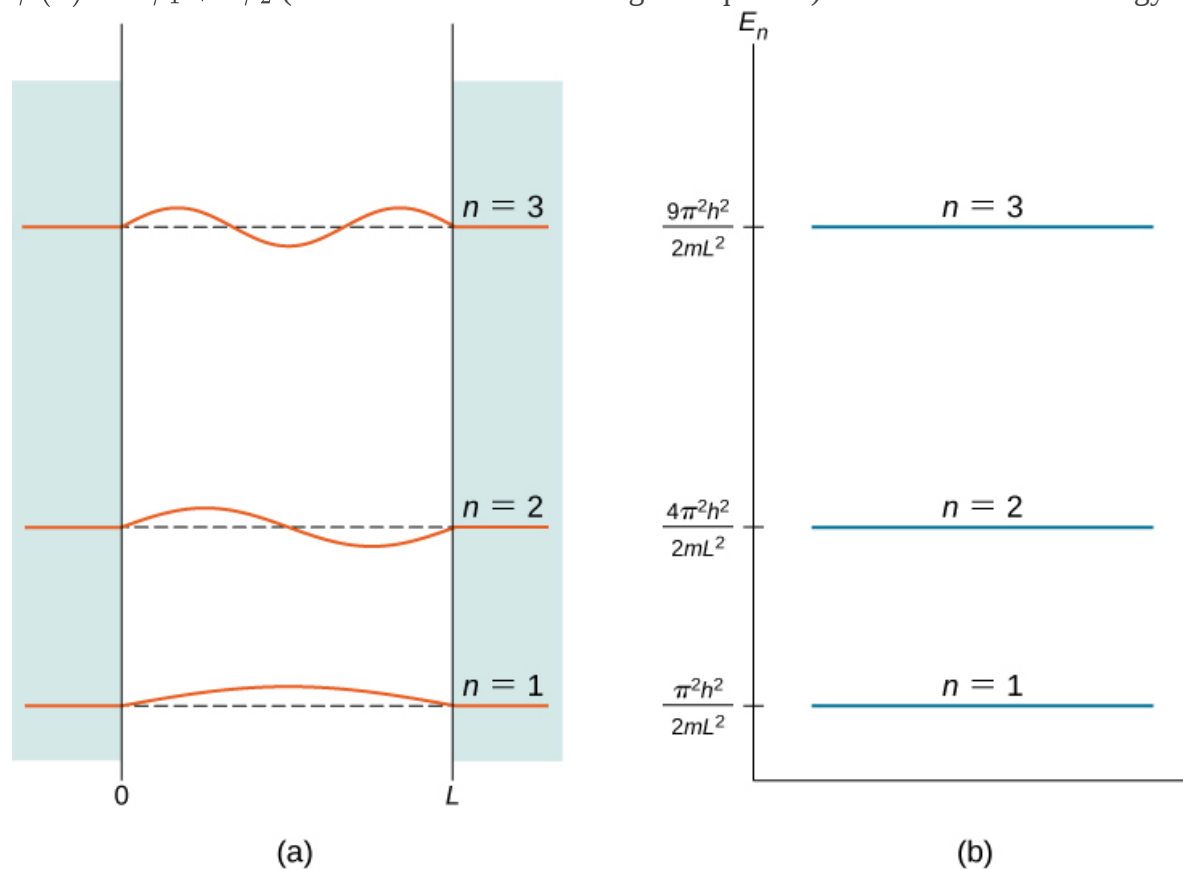
Equation:

$$E_n = n^2 E_1, \psi_n(x) = \sqrt{\frac{2}{L}} \sin \left(\frac{n\pi x}{L} \right).$$

The index n is called the **energy quantum number** or **principal quantum number**. The state for $n = 2$ is the first excited state, the state for $n = 3$ is the second excited state, and so on. The first three quantum states (for $n = 1, 2$, and 3) of a particle in a box are shown in [\[link\]](#).

The wave functions in [\[link\]](#) are sometimes referred to as the “states of definite energy.” Particles in these states are said to occupy **energy levels**, which are represented by the horizontal lines in [\[link\]](#). Energy levels are analogous to rungs of a ladder that the particle can “climb” as it gains or loses energy.

The wave functions in [\[link\]](#) are also called **stationary states** and **standing wave states**. These functions are “stationary,” because their probability density functions, $|\Psi(x, t)|^2$, do not vary in time, and “standing waves” because their real and imaginary parts oscillate up and down like a standing wave—like a rope waving between two children on a playground. Stationary states are states of definite energy [\[link\]](#), but linear combinations of these states, such as $\psi(x) = a\psi_1 + b\psi_2$ (also solutions to Schrödinger’s equation) are states of mixed energy.



The first three quantum states of a quantum particle in a box for principal quantum numbers $n = 1, 2$, and 3 : (a) standing wave solutions and (b) allowed energy states.

Energy quantization is a consequence of the boundary conditions. If the particle is not confined to a box but wanders freely, the allowed energies are continuous. However, in this case, only certain energies ($E_1, 4E_1, 9E_1, \dots$) are allowed. The energy difference between adjacent energy levels is given by

Equation:

$$\Delta E_{n+1,n} = E_{n+1} - E_n = (n+1)^2 E_1 - n^2 E_1 = (2n+1)E_1.$$

Conservation of energy demands that if the energy of the system changes, the energy difference is carried in some other form of energy. For the special case of a charged particle confined to a small volume (for example, in an atom), energy changes are often carried away by photons. The frequencies of the emitted photons give us information about the energy differences (spacings) of the system and the volume of containment—the size of the “box” [see [link](#)].

Example:

A Simple Model of the Nucleus

Suppose a proton is confined to a box of width $L = 1.00 \times 10^{-14} \text{ m}$ (a typical nuclear radius). What are the energies of the ground and the first excited states? If the proton makes a transition from the first excited state to the ground state, what are the energy and the frequency of the emitted photon?

Strategy

If we assume that the proton confined in the nucleus can be modeled as a quantum particle in a box, all we need to do is to use [link](#) to find its energies E_1 and E_2 . The mass of a proton is $m = 1.67 \times 10^{-27} \text{ kg}$. The emitted photon carries away the energy difference $\Delta E = E_2 - E_1$. We can use the relation $E_f = hf$ to find its frequency f .

Solution

The ground state:

Equation:

$$E_1 = \frac{\pi^2 \hbar^2}{2m L^2} = \frac{\pi^2 (1.05 \times 10^{-34} \text{ J} \cdot \text{s})^2}{2(1.67 \times 10^{-27} \text{ kg}) (1.00 \times 10^{-14} \text{ m})^2} = 3.28 \times 10^{-13} \text{ J} = 2.05 \text{ MeV}.$$

The first excited state: $E_2 = 2^2 E_1 = 4(2.05 \text{ MeV}) = 8.20 \text{ MeV}$.

The energy of the emitted photon is

$$E_f = \Delta E = E_2 - E_1 = 8.20 \text{ MeV} - 2.05 \text{ MeV} = 6.15 \text{ MeV}.$$

The frequency of the emitted photon is

Equation:

$$f = \frac{E_f}{h} = \frac{6.15 \text{ MeV}}{4.14 \times 10^{-21} \text{ MeV} \cdot \text{s}} = 1.49 \times 10^{21} \text{ Hz}.$$

Significance

This is the typical frequency of a gamma ray emitted by a nucleus. The energy of this photon is about 10 million times greater than that of a visible light photon.

The expectation value of the position for a particle in a box is given by

Equation:

$$\langle x \rangle = \int_0^L dx \psi_n^*(x) x \psi_n(x) = \int_0^L dx x |\psi_n^*(x)|^2 = \int_0^L dx x \frac{2}{L} \sin^2 \frac{n\pi x}{L} = \frac{L}{2}.$$

We can also find the expectation value of the momentum or average momentum of a large number of particles in a given state:

Equation:

$$\begin{aligned} \langle p \rangle &= \int_0^L dx \psi_n^*(x) \left[-i\hbar \frac{d}{dx} \psi_n(x) \right] \\ &= -i\hbar \int_0^L dx \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \left[\frac{d}{dx} \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \right] = -i\frac{2\hbar}{L} \int_0^L dx \sin \frac{n\pi x}{L} \left[\frac{n\pi}{L} \cos \frac{n\pi x}{L} \right] \\ &= -i\frac{2n\pi\hbar}{L^2} \int_0^L dx \frac{1}{2} \sin \frac{2n\pi x}{L} = -i\frac{n\pi\hbar}{L^2} \frac{L}{2n\pi} \int_0^{2\pi n} d\varphi \sin \varphi = -i\frac{\hbar}{2L} \cdot 0 = 0. \end{aligned}$$

Thus, for a particle in a state of definite energy, the average position is in the middle of the box and the average momentum of the particle is zero—as it would also be for a classical particle. Note that while the minimum energy of a classical particle can be zero (the particle can be at rest in the middle of the box), the minimum energy of a quantum particle is nonzero and given by [\[link\]](#). The average particle energy in the n th quantum state—its expectation value of energy—is

Equation:

$$E_n = \langle E \rangle = n^2 \frac{\pi^2 \hbar^2}{2m}.$$

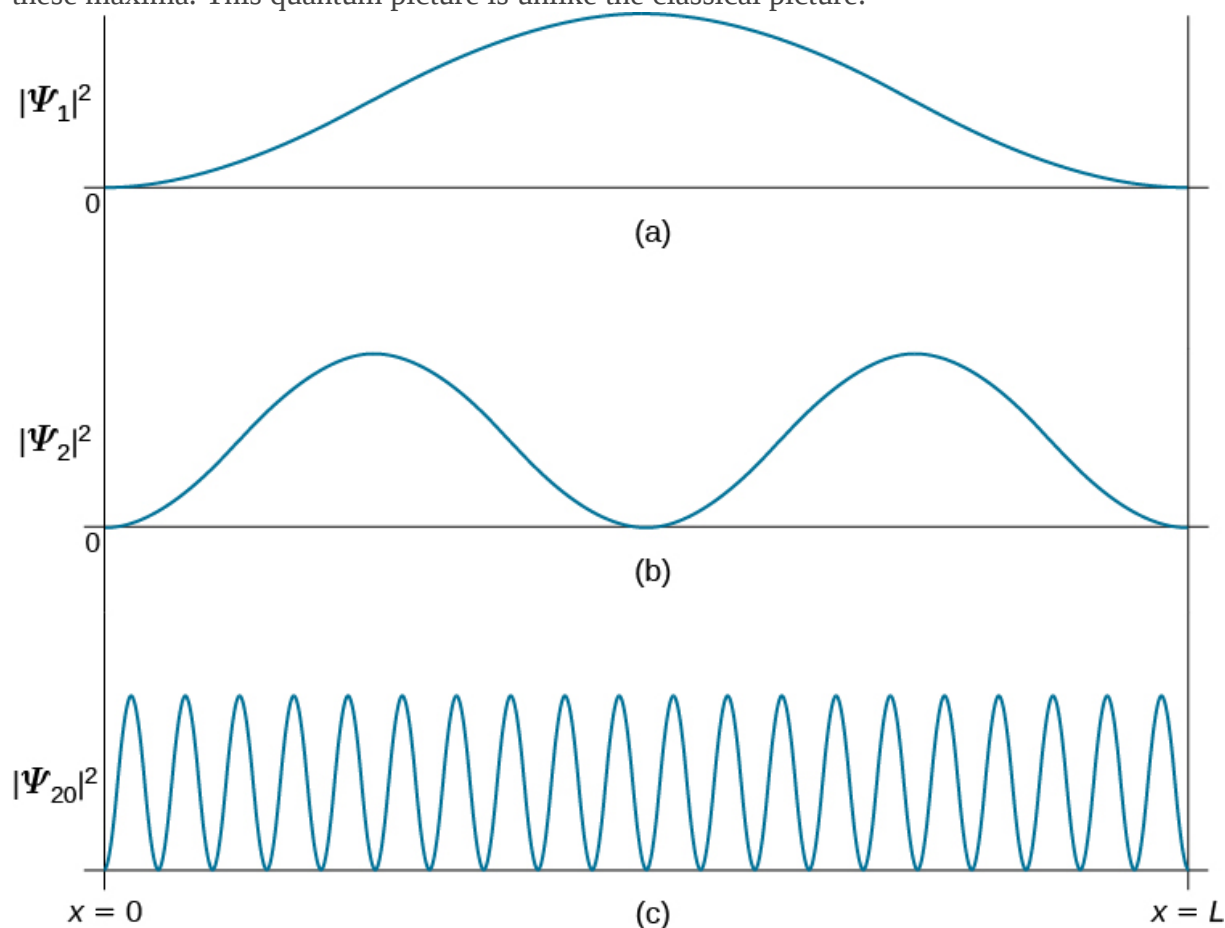
The result is not surprising because the standing wave state is a state of definite energy. Any energy measurement of this system must return a value equal to one of these allowed energies.

Our analysis of the quantum particle in a box would not be complete without discussing Bohr's correspondence principle. This principle states that for large quantum numbers, the laws of quantum physics must give identical results as the laws of classical physics. To illustrate how this principle works for a quantum particle in a box, we plot the probability density distribution

Equation:

$$|\psi_n(x)|^2 = \frac{2}{L} \sin^2(n\pi x/L)$$

for finding the particle around location x between the walls when the particle is in quantum state ψ_n . [\[link\]](#) shows these probability distributions for the ground state, for the first excited state, and for a highly excited state that corresponds to a large quantum number. We see from these plots that when a quantum particle is in the ground state, it is most likely to be found around the middle of the box, where the probability distribution has the largest value. This is not so when the particle is in the first excited state because now the probability distribution has the zero value in the middle of the box, so there is no chance of finding the particle there. When a quantum particle is in the first excited state, the probability distribution has two maxima, and the best chance of finding the particle is at positions close to the locations of these maxima. This quantum picture is unlike the classical picture.



The probability density distribution $|\psi_n(x)|^2$ for a quantum particle in a box for: (a) the ground state, $n = 1$; (b) the first excited state, $n = 2$; and, (c) the nineteenth excited state, $n = 20$.

The probability density of finding a classical particle between x and $x + \Delta x$ depends on how much time Δt the particle spends in this region. Assuming that its speed u is constant, this time is $\Delta t = \Delta x/u$, which is also constant for any location between the walls. Therefore, the

probability density of finding the classical particle at x is uniform throughout the box, and there is no preferable location for finding a classical particle. This classical picture is matched in the limit of large quantum numbers. For example, when a quantum particle is in a highly excited state, shown in [\[link\]](#), the probability density is characterized by rapid fluctuations and then the probability of finding the quantum particle in the interval Δx does not depend on where this interval is located between the walls.

Example:

A Classical Particle in a Box

A small 0.40-kg cart is moving back and forth along an air track between two bumpers located 2.0 m apart. We assume no friction; collisions with the bumpers are perfectly elastic so that between the bumpers, the car maintains a constant speed of 0.50 m/s. Treating the cart as a quantum particle, estimate the value of the principal quantum number that corresponds to its classical energy.

Strategy

We find the kinetic energy K of the cart and its ground state energy E_1 as though it were a quantum particle. The energy of the cart is completely kinetic, so $K = n^2 E_1$ ([\[link\]](#)). Solving for n gives $n = (K/E_1)^{1/2}$.

Solution

The kinetic energy of the cart is

Equation:

$$K = \frac{1}{2}mu^2 = \frac{1}{2}(0.40 \text{ kg})(0.50 \text{ m/s})^2 = 0.050 \text{ J}.$$

The ground state of the cart, treated as a quantum particle, is

Equation:

$$E_1 = \frac{\pi^2 \hbar^2}{2mL^2} = \frac{\pi^2 (1.05 \times 10^{-34} \text{ J} \cdot \text{s})^2}{2(0.40 \text{ kg})(2.0 \text{ m})^2} = 1.700 \times 10^{-68} \text{ J}.$$

Therefore, $n = (K/E_1)^{1/2} = (0.050/1.700 \times 10^{-68})^{1/2} = 1.2 \times 10^{33}$.

Significance

We see from this example that the energy of a classical system is characterized by a very large quantum number. Bohr's correspondence principle concerns this kind of situation. We can apply the formalism of quantum mechanics to any kind of system, quantum or classical, and the results are correct in each case. In the limit of high quantum numbers, there is no advantage in using quantum formalism because we can obtain the same results with the less complicated formalism of classical mechanics. However, we cannot apply classical formalism to a quantum system in a low-number energy state.

Note:

Exercise:**Problem:**

Check Your Understanding (a) Consider an infinite square well with wall boundaries $x = 0$ and $x = L$. What is the probability of finding a quantum particle in its ground state somewhere between $x = 0$ and $x = L/4$? (b) Repeat question (a) for a classical particle.

Solution:

a. 9.1%; b. 25%

Having found the stationary states $\psi_n(x)$ and the energies E_n by solving the time-independent Schrödinger equation [\[link\]](#), we use [\[link\]](#) to write wave functions $\Psi_n(x, t)$ that are solutions of the time-dependent Schrödinger's equation given by [\[link\]](#). For a particle in a box this gives **Equation:**

$$\Psi_n(x, t) = e^{-i\omega_n t} \psi_n(x) = \sqrt{\frac{2}{L}} e^{-iE_n t/\hbar} \sin \frac{n\pi x}{L}, n = 1, 2, 3, \dots$$

where the energies are given by [\[link\]](#).

The quantum particle in a box model has practical applications in a relatively newly emerged field of optoelectronics, which deals with devices that convert electrical signals into optical signals. This model also deals with nanoscale physical phenomena, such as a nanoparticle trapped in a low electric potential bounded by high-potential barriers.

Summary

- Energy states of a quantum particle in a box are found by solving the time-independent Schrödinger equation.
- To solve the time-independent Schrödinger equation for a particle in a box and find the stationary states and allowed energies, we require that the wave function terminate at the box wall.
- Energy states of a particle in a box are quantized and indexed by principal quantum number.
- The quantum picture differs significantly from the classical picture when a particle is in a low-energy state of a low quantum number.
- In the limit of high quantum numbers, when the quantum particle is in a highly excited state, the quantum description of a particle in a box coincides with the classical description, in the spirit of Bohr's correspondence principle.

Conceptual Questions

Exercise:

Problem:

Using the quantum particle in a box model, describe how the possible energies of the particle are related to the size of the box.

Exercise:

Problem:

Is it possible that when we measure the energy of a quantum particle in a box, the measurement may return a smaller value than the ground state energy? What is the highest value of the energy that we can measure for this particle?

Solution:

No. For an infinite square well, the spacing between energy levels increases with the quantum number n . The *smallest* energy measured corresponds to the transition from $n = 2$ to 1, which is three times the ground state energy. The largest *energy* measured corresponds to a transition from $n = \infty$ to 1, which is infinity. (Note: Even particles with extremely large energies remain bound to an infinite square well—they can never “escape”)

Exercise:

Problem:

For a quantum particle in a box, the first excited state (Ψ_2) has zero value at the midpoint position in the box, so that the probability density of finding a particle at this point is exactly zero. Explain what is wrong with the following reasoning: “If the probability of finding a quantum particle at the midpoint is zero, the particle is never at this point, right? How does it come then that the particle can cross this point on its way from the left side to the right side of the box?”

Problems

Exercise:

Problem:

Assume that an electron in an atom can be treated as if it were confined to a box of width 2.0 \AA . What is the ground state energy of the electron? Compare your result to the ground state kinetic energy of the hydrogen atom in the Bohr’s model of the hydrogen atom.

Solution:

9.4 eV, 64%

Exercise:**Problem:**

Assume that a proton in a nucleus can be treated as if it were confined to a one-dimensional box of width 10.0 fm. (a) What are the energies of the proton when it is in the states corresponding to $n = 1$, $n = 2$, and $n = 3$? (b) What are the energies of the photons emitted when the proton makes the transitions from the first and second excited states to the ground state?

Exercise:**Problem:**

An electron confined to a box has the ground state energy of 2.5 eV. What is the width of the box?

Solution:

0.38 nm

Exercise:**Problem:**

What is the ground state energy (in eV) of a proton confined to a one-dimensional box the size of the uranium nucleus that has a radius of approximately 15.0 fm?

Exercise:**Problem:**

What is the ground state energy (in eV) of an α -particle confined to a one-dimensional box the size of the uranium nucleus that has a radius of approximately 15.0 fm?

Solution:

1.82 MeV

Exercise:**Problem:**

To excite an electron in a one-dimensional box from its first excited state to its third excited state requires 20.0 eV. What is the width of the box?

Exercise:**Problem:**

An electron confined to a box of width 0.15 nm by infinite potential energy barriers emits a photon when it makes a transition from the first excited state to the ground state. Find the wavelength of the emitted photon.

Solution:

24.7 nm

Exercise:**Problem:**

If the energy of the first excited state of the electron in the box is 25.0 eV, what is the width of the box?

Exercise:**Problem:**

Suppose an electron confined to a box emits photons. The longest wavelength that is registered is 500.0 nm. What is the width of the box?

Solution:

6.03 Å

Exercise:**Problem:**

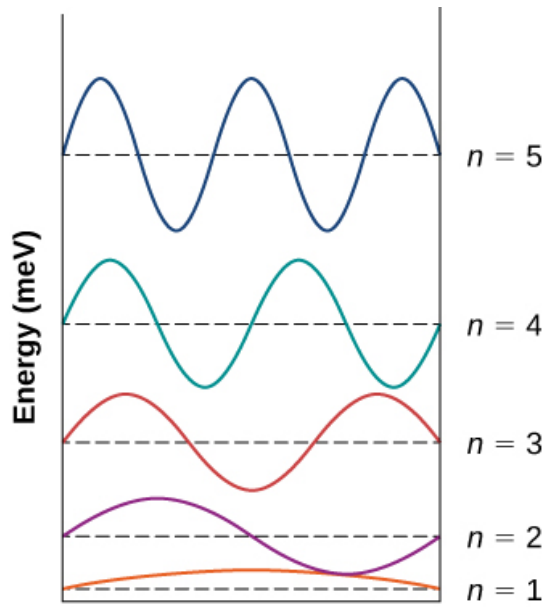
Hydrogen H_2 molecules are kept at 300.0 K in a cubical container with a side length of 20.0 cm. Assume that you can treat the molecules as though they were moving in a one-dimensional box. (a) Find the ground state energy of the hydrogen molecule in the container. (b) Assume that the molecule has a thermal energy given by $k_{\text{B}}T/2$ and find the corresponding quantum number n of the quantum state that would correspond to this thermal energy.

Exercise:**Problem:**

An electron is confined to a box of width 0.25 nm. (a) Draw an energy-level diagram representing the first five states of the electron. (b) Calculate the wavelengths of the emitted photons when the electron makes transitions between the fourth and the second excited states, between the second excited state and the ground state, and between the third and the second excited states.

Solution:

a.



;
b. $\lambda_{5 \rightarrow 3} = 12.9 \text{ nm}$, $\lambda_{3 \rightarrow 1} = 25.8 \text{ nm}$, $\lambda_{4 \rightarrow 3} = 29.4 \text{ nm}$

Exercise:

Problem:

An electron in a box is in the ground state with energy 2.0 eV. (a) Find the width of the box. (b) How much energy is needed to excite the electron to its first excited state? (c) If the electron makes a transition from an excited state to the ground state with the simultaneous emission of 30.0-eV photon, find the quantum number of the excited state?

Glossary

energy levels

states of definite energy, often represented by horizontal lines in an energy “ladder” diagram

energy quantum number

index that labels the allowed energy states

ground state energy

lowest energy state in the energy spectrum

infinite square well

potential function that is zero in a fixed range and infinitely beyond this range

principal quantum number

energy quantum number

standing wave state

stationary state for which the real and imaginary parts of $\Psi(x, t)$ oscillate up and down like a standing wave (often modeled with sine and cosine functions)

stationary state

state for which the probability density function, $|\Psi(x, t)|^2$, does not vary in time

The Quantum Harmonic Oscillator

By the end of this section, you will be able to:

- Describe the model of the quantum harmonic oscillator
- Identify differences between the classical and quantum models of the harmonic oscillator
- Explain physical situations where the classical and the quantum models coincide

Oscillations are found throughout nature, in such things as electromagnetic waves, vibrating molecules, and the gentle back-and-forth sway of a tree branch. In previous chapters, we used Newtonian mechanics to study macroscopic oscillations, such as a block on a spring and a simple pendulum. In this chapter, we begin to study oscillating systems using quantum mechanics. We begin with a review of the classic harmonic oscillator.

The Classic Harmonic Oscillator

A simple harmonic oscillator is a particle or system that undergoes harmonic motion about an equilibrium position, such as an object with mass vibrating on a spring. In this section, we consider oscillations in one-dimension only. Suppose a mass moves back-and-forth along the

x -direction about the equilibrium position, $x = 0$. In classical mechanics, the particle moves in response to a linear restoring force given by $F_x = -kx$, where x is the displacement of the particle from its equilibrium position. The motion takes place between two turning points, $x = \pm A$, where A denotes the amplitude of the motion. The position of the object varies periodically in time with angular frequency $\omega = \sqrt{k/m}$, which depends on the mass m of the oscillator and on the force constant k of the net force, and can be written as

Equation:

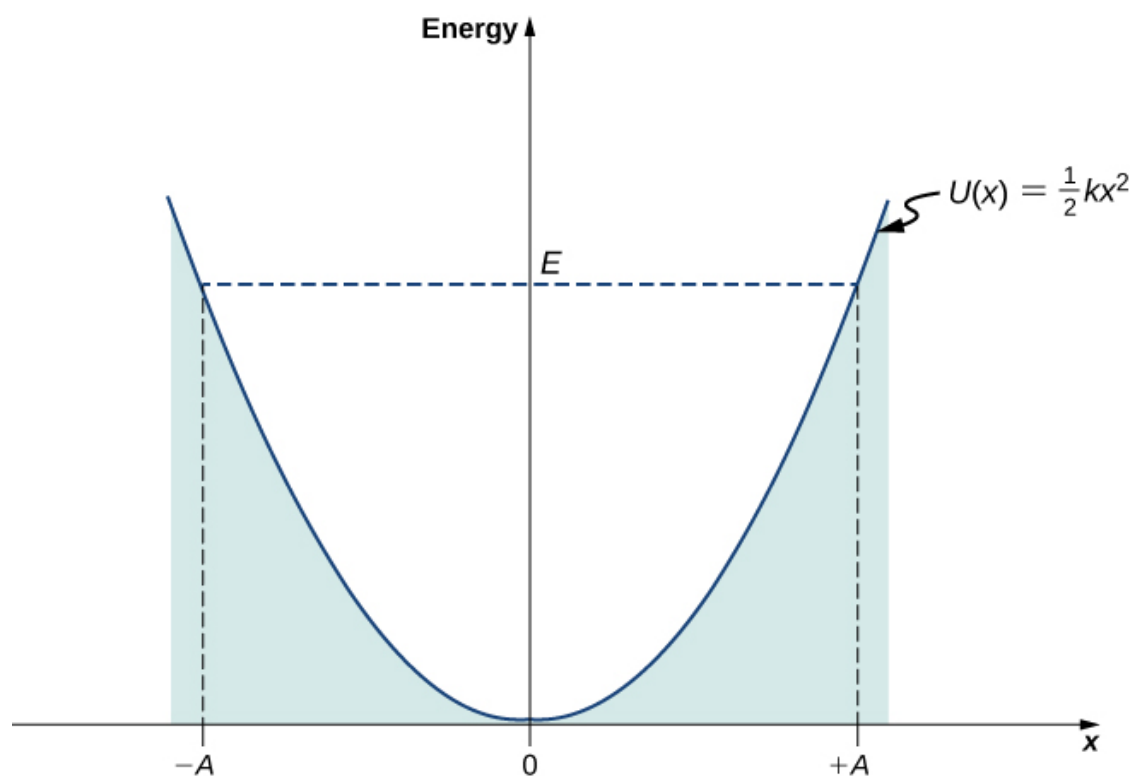
$$x(t) = A \cos(\omega t + \phi).$$

The total energy E of an oscillator is the sum of its kinetic energy $K = mu^2/2$ and the elastic potential energy of the force $U(x) = kx^2/2$,

Equation:

$$E = \frac{1}{2}mu^2 + \frac{1}{2}kx^2.$$

At turning points $x = \pm A$, the speed of the oscillator is zero; therefore, at these points, the energy of oscillation is solely in the form of potential energy $E = k A^2/2$. The plot of the potential energy $U(x)$ of the oscillator versus its position x is a parabola ([link](#)). The potential-energy function is a quadratic function of x , measured with respect to the equilibrium position. On the same graph, we also plot the total energy E of the oscillator, as a horizontal line that intercepts the parabola at $x = \pm A$. Then the kinetic energy K is represented as the vertical distance between the line of total energy and the potential energy parabola.



The potential energy well of a classical harmonic oscillator: The motion is confined between turning points at $x = -A$ and at $x = +A$. The energy of oscillations is $E = kA^2/2$.

In this plot, the motion of a classical oscillator is confined to the region where its kinetic energy is nonnegative, which is what the energy relation [link](#) says. Physically, it means that a classical oscillator can never be found beyond its turning points, and its energy depends only on how far the turning points are from its equilibrium position. The energy of a classical oscillator changes in a continuous way. The lowest energy that a classical oscillator may have is zero, which corresponds to a situation where an object is at rest at

its equilibrium position. The zero-energy state of a classical oscillator simply means no oscillations and no motion at all (a classical particle sitting at the bottom of the potential well in [\[link\]](#)). When an object oscillates, no matter how big or small its energy may be, it spends the longest time near the turning points, because this is where it slows down and reverses its direction of motion. Therefore, the probability of finding a classical oscillator between the turning points is highest near the turning points and lowest at the equilibrium position. (Note that this is not a statement of preference of the object to go to lower energy. It is a statement about how quickly the object moves through various regions.)

The Quantum Harmonic Oscillator

One problem with this classical formulation is that it is not general. We cannot use it, for example, to describe vibrations of diatomic molecules, where quantum effects are important. A first step toward a quantum formulation is to use the classical expression $k = m\omega^2$ to limit mention of a “spring” constant between the atoms. In this way the potential energy function can be written in a more general form,

Note:

Equation:

$$U(x) = \frac{1}{2}m\omega^2x^2.$$

Combining this expression with the time-independent Schrödinger equation gives

Note:

Equation:

$$-\frac{\hbar}{2m} \frac{d^2\psi(x)}{dx^2} + \frac{1}{2}m\omega^2x^2\psi(x) = E\psi(x).$$

To solve [\[link\]](#)—that is, to find the allowed energies E and their corresponding wave functions $\psi(x)$ —we require the wave functions to be symmetric about $x = 0$ (the bottom of the potential well) and to be normalizable. These conditions ensure that the

probability density $|\psi(x)|^2$ must be finite when integrated over the entire range of x from $-\infty$ to $+\infty$. How to solve [\[link\]](#) is the subject of a more advanced course in quantum mechanics; here, we simply cite the results. The allowed energies are

Note:

Equation:

$$E_n = \left(n + \frac{1}{2}\right) \hbar \omega = \frac{2n+1}{2} \hbar \omega, \quad n = 0, 1, 2, 3, \dots$$

The wave functions that correspond to these energies (the stationary states or states of definite energy) are

Note:

Equation:

$$\psi_n(x) = N_n e^{-\beta^2 x^2 / 2} H_n(\beta x), \quad n = 0, 1, 2, 3, \dots$$

where $\beta = \sqrt{m\omega/\hbar}$, N_n is the normalization constant, and $H_n(y)$ is a polynomial of degree n called a *Hermite polynomial*. The first four Hermite polynomials are

Equation:

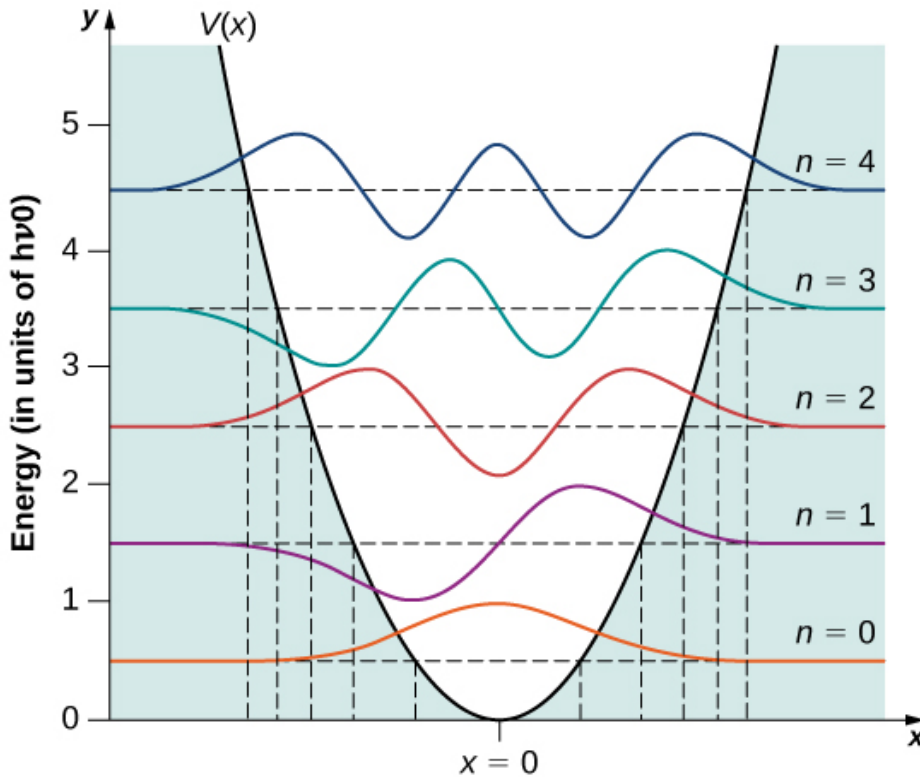
$$H_0(y) = 1$$

$$H_1(y) = 2y$$

$$H_2(y) = 4y^2 - 2$$

$$H_3(y) = 8y^3 - 12y.$$

A few sample wave functions are given in [\[link\]](#). As the value of the principal number increases, the solutions alternate between even functions and odd functions about $x = 0$.



The first five wave functions of the quantum harmonic oscillator. The classical limits of the oscillator's motion are indicated by vertical lines, corresponding to the classical turning points at $x = \pm A$ of a classical particle with the same energy as the energy of a quantum oscillator in the state indicated in the figure.

Example:

Classical Region of Harmonic Oscillations

Find the amplitude A of oscillations for a classical oscillator with energy equal to the energy of a quantum oscillator in the quantum state n .

Strategy

To determine the amplitude A , we set the classical energy $E = kx^2/2 = m\omega^2 A^2/2$ equal to E_n given by [\[link\]](#).

Solution

We obtain

Equation:

$$E_n = m\omega^2 A_n^2/2 \Rightarrow A_n = \sqrt{\frac{2}{m\omega^2} E_n} = \sqrt{\frac{2}{m\omega^2} \frac{2n+1}{2} \hbar\omega} = \sqrt{(2n+1) \frac{\hbar}{m\omega}}.$$

Significance

As the quantum number n increases, the energy of the oscillator and therefore the amplitude of oscillation increases (for a fixed natural angular frequency). For large n , the amplitude is approximately proportional to the square root of the quantum number.

Several interesting features appear in this solution. Unlike a classical oscillator, the measured energies of a quantum oscillator can have only energy values given by [\[link\]](#). Moreover, unlike the case for a quantum particle in a box, the allowable energy levels are evenly spaced,

Equation:

$$\Delta E = E_{n+1} - E_n = \frac{2(n+1)+1}{2} \hbar\omega - \frac{2n+1}{2} \hbar\omega = \hbar\omega = hf.$$

When a particle bound to such a system makes a transition from a higher-energy state to a lower-energy state, the smallest-energy quantum carried by the emitted photon is necessarily hf . Similarly, when the particle makes a transition from a lower-energy state to a higher-energy state, the smallest-energy quantum that can be absorbed by the particle is hf . A quantum oscillator can absorb or emit energy only in multiples of this smallest-energy quantum. This is consistent with Planck's hypothesis for the energy exchanges between radiation and the cavity walls in the blackbody radiation problem.

Example:

Vibrational Energies of the Hydrogen Chloride Molecule

The HCl diatomic molecule consists of one chlorine atom and one hydrogen atom. Because the chlorine atom is 35 times more massive than the hydrogen atom, the vibrations of the HCl molecule can be quite well approximated by assuming that the Cl atom is motionless and the H atom performs harmonic oscillations due to an elastic molecular force modeled by Hooke's law. The infrared vibrational spectrum measured for hydrogen chloride has the lowest-frequency line centered at $f = 8.88 \times 10^{13}$ Hz. What is the spacing between the vibrational energies of this molecule? What is the force constant k of the atomic bond in the HCl molecule?

Strategy

The lowest-frequency line corresponds to the emission of lowest-frequency photons. These photons are emitted when the molecule makes a transition between two adjacent vibrational energy levels. Assuming that energy levels are equally spaced, we use [\[link\]](#)

to estimate the spacing. The molecule is well approximated by treating the Cl atom as being infinitely heavy and the H atom as the mass m that performs the oscillations. Treating this molecular system as a classical oscillator, the force constant is found from the classical relation $k = m\omega^2$.

Solution

The energy spacing is

Equation:

$$\Delta E = hf = (4.14 \times 10^{-15} \text{ eV} \cdot \text{s})(8.88 \times 10^{13} \text{ Hz}) = 0.368 \text{ eV}.$$

The force constant is

Equation:

$$k = m\omega^2 = m(2\pi f)^2 = (1.67 \times 10^{-27} \text{ kg})(2\pi \times 8.88 \times 10^{13} \text{ Hz})^2 = 520 \text{ N/m}.$$

Significance

The force between atoms in an HCl molecule is surprisingly strong. The typical energy released in energy transitions between vibrational levels is in the infrared range. As we will see later, transitions in between vibrational energy levels of a diatomic molecule often accompany transitions between rotational energy levels.

Note:

Exercise:

Problem:

Check Your Understanding The vibrational frequency of the hydrogen iodide HI diatomic molecule is $6.69 \times 10^{13} \text{ Hz}$. (a) What is the force constant of the molecular bond between the hydrogen and the iodine atoms? (b) What is the energy of the emitted photon when this molecule makes a transition between adjacent vibrational energy levels?

Solution:

a. 295 N/m; b. 0.277 eV

The quantum oscillator differs from the classic oscillator in three ways:

First, the ground state of a quantum oscillator is $E_0 = \hbar\omega/2$, not zero. In the classical view, the lowest energy is zero. The nonexistence of a zero-energy state is common for

all quantum-mechanical systems because of omnipresent fluctuations that are a consequence of the Heisenberg uncertainty principle. If a quantum particle sat motionless at the bottom of the potential well, its momentum as well as its position would have to be simultaneously exact, which would violate the Heisenberg uncertainty principle. Therefore, the lowest-energy state must be characterized by uncertainties in momentum and in position, so the ground state of a quantum particle must lie above the bottom of the potential well.

Second, a particle in a quantum harmonic oscillator potential can be found with nonzero probability outside the interval $-A \leq x \leq +A$. In a classic formulation of the problem, the particle would not have any energy to be in this region. The probability of finding a ground-state quantum particle in the classically forbidden region is about 16%.

Third, the probability density distributions $|\psi_n(x)|^2$ for a quantum oscillator in the ground low-energy state, $\psi_0(x)$, is largest at the middle of the well ($x = 0$). For the particle to be found with greatest probability at the center of the well, we expect that the particle spends the most time there as it oscillates. This is opposite to the behavior of a classical oscillator, in which the particle spends most of its time moving with relative small speeds near the turning points.

Note:

Exercise:

Problem:

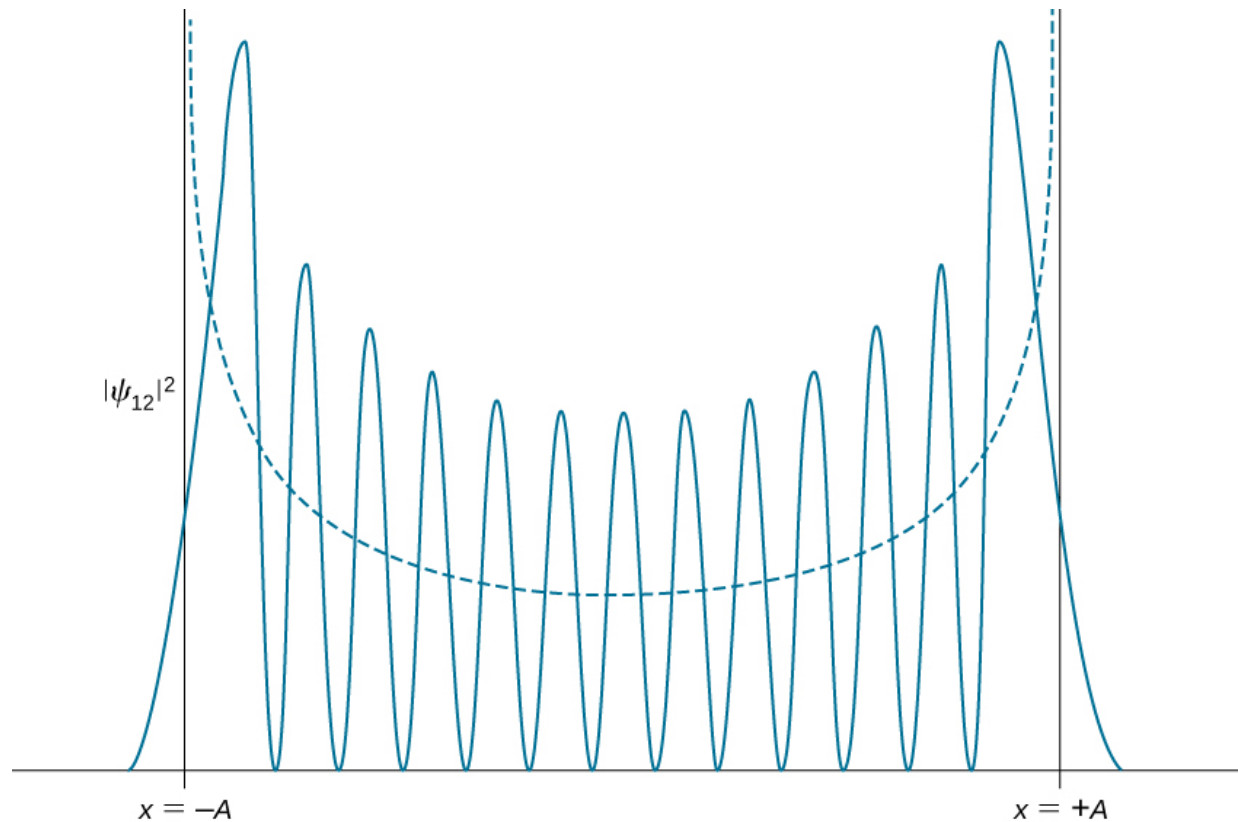
Check Your Understanding Find the expectation value of the position for a particle in the ground state of a harmonic oscillator using symmetry.

Solution:

$$\langle x \rangle = 0$$

Quantum probability density distributions change in character for excited states, becoming more like the classical distribution when the quantum number gets higher. We observe this change already for the first excited state of a quantum oscillator because the distribution $|\psi_1(x)|^2$ peaks up around the turning points and vanishes at the equilibrium position, as seen in [\[link\]](#). In accordance with Bohr's correspondence principle, in the limit of high quantum numbers, the quantum description of a harmonic oscillator converges to the classical description, which is illustrated in [\[link\]](#). The classical probability density distribution corresponding to the quantum energy of the $n = 12$ state

is a reasonably good approximation of the quantum probability distribution for a quantum oscillator in this excited state. This agreement becomes increasingly better for highly excited states.



The probability density distribution for finding the quantum harmonic oscillator in its $n = 12$ quantum state. The dashed curve shows the probability density distribution of a classical oscillator with the same energy.

Summary

- The quantum harmonic oscillator is a model built in analogy with the model of a classical harmonic oscillator. It models the behavior of many physical systems, such as molecular vibrations or wave packets in quantum optics.
- The allowed energies of a quantum oscillator are discrete and evenly spaced. The energy spacing is equal to Planck's energy quantum.
- The ground state energy is larger than zero. This means that, unlike a classical oscillator, a quantum oscillator is never at rest, even at the bottom of a potential

well, and undergoes quantum fluctuations.

- The stationary states (states of definite energy) have nonzero values also in regions beyond classical turning points. When in the ground state, a quantum oscillator is most likely to be found around the position of the minimum of the potential well, which is the least-likely position for a classical oscillator.
- For high quantum numbers, the motion of a quantum oscillator becomes more similar to the motion of a classical oscillator, in accordance with Bohr's correspondence principle.

Conceptual Questions

Exercise:

Problem:

Is it possible to measure energy of $0.75\hbar\omega$ for a quantum harmonic oscillator? Why? Why not? Explain.

Solution:

No. This energy corresponds to $n = 0.25$, but n must be an integer.

Exercise:

Problem:

Explain the connection between Planck's hypothesis of energy quanta and the energies of the quantum harmonic oscillator.

Exercise:

Problem:

If a classical harmonic oscillator can be at rest, why can the quantum harmonic oscillator never be at rest? Does this violate Bohr's correspondence principle?

Solution:

Because the smallest allowed value of the quantum number n for a simple harmonic oscillator is 0. No, because quantum mechanics and classical mechanics agree only in the limit of large n .

Exercise:

Problem:

Use an example of a quantum particle in a box or a quantum oscillator to explain the physical meaning of Bohr's correspondence principle.

Exercise:**Problem:**

Can we simultaneously measure position and energy of a quantum oscillator? Why? Why not?

Solution:

Yes, within the constraints of the uncertainty principle. If the oscillating particle is localized, the momentum and therefore energy of the oscillator are distributed.

Problems**Exercise:****Problem:**

Show that the two lowest energy states of the simple harmonic oscillator, $\psi_0(x)$ and $\psi_1(x)$ from [\[link\]](#), satisfy [\[link\]](#).

Solution:

proof

Exercise:**Problem:**

If the ground state energy of a simple harmonic oscillator is 1.25 eV, what is the frequency of its motion?

Exercise:**Problem:**

When a quantum harmonic oscillator makes a transition from the $(n + 1)$ state to the n state and emits a 450-nm photon, what is its frequency?

Solution:

$$6.662 \times 10^{14} \text{ Hz}$$

Exercise:

Problem:

Vibrations of the hydrogen molecule H_2 can be modeled as a simple harmonic oscillator with the spring constant $k = 1.13 \times 10^3 \text{ N/m}$ and mass $m = 1.67 \times 10^{-27} \text{ kg}$. (a) What is the vibrational frequency of this molecule? (b) What are the energy and the wavelength of the emitted photon when the molecule makes transition between its third and second excited states?

Exercise:**Problem:**

A particle with mass 0.030 kg oscillates back-and-forth on a spring with frequency 4.0 Hz . At the equilibrium position, it has a speed of 0.60 m/s . If the particle is in a state of definite energy, find its energy quantum number.

Solution:

$$n \approx 2.037 \times 10^{30}$$

Exercise:**Problem:**

Find the expectation value $\langle x^2 \rangle$ of the square of the position for a quantum harmonic oscillator in the ground state. Note: $\int_{-\infty}^{+\infty} dx x^2 e^{-ax^2} = \sqrt{\pi} (2a^{3/2})^{-1}$.

Exercise:**Problem:**

Determine the expectation value of the potential energy for a quantum harmonic oscillator in the ground state. Use this to calculate the expectation value of the kinetic energy.

Solution:

$$\langle x \rangle = 0.5m\omega^2 \langle x^2 \rangle = \hbar\omega/4; \langle K \rangle = \langle E \rangle - \langle U \rangle = \hbar\omega/4$$

Exercise:**Problem:**

Verify that $\psi_1(x)$ given by [\[link\]](#) is a solution of Schrödinger's equation for the quantum harmonic oscillator.

Exercise:

Problem:

Estimate the ground state energy of the quantum harmonic oscillator by Heisenberg's uncertainty principle. Start by assuming that the product of the uncertainties Δx and Δp is at its minimum. Write Δp in terms of Δx and assume that for the ground state $x \approx \Delta x$ and $p \approx \Delta p$, then write the ground state energy in terms of x . Finally, find the value of x that minimizes the energy and find the minimum of the energy.

Solution:

proof

Exercise:**Problem:**

A mass of 0.250 kg oscillates on a spring with the force constant 110 N/m. Calculate the ground energy level and the separation between the adjacent energy levels. Express the results in joules and in electron-volts. Are quantum effects important?

The Quantum Tunneling of Particles through Potential Barriers

By the end of this section, you will be able to:

- Describe how a quantum particle may tunnel across a potential barrier
- Identify important physical parameters that affect the tunneling probability
- Identify the physical phenomena where quantum tunneling is observed
- Explain how quantum tunneling is utilized in modern technologies

Quantum tunneling is a phenomenon in which particles penetrate a potential energy barrier with a height greater than the total energy of the particles. The phenomenon is interesting and important because it violates the principles of classical mechanics. Quantum tunneling is important in models of the Sun and has a wide range of applications, such as the scanning tunneling microscope and the tunnel diode.

Tunneling and Potential Energy

To illustrate quantum tunneling, consider a ball rolling along a surface with a kinetic energy of 100 J. As the ball rolls, it encounters a hill. The potential energy of the ball placed atop the hill is 10 J. Therefore, the ball (with 100 J of kinetic energy) easily rolls over the hill and continues on. In classical mechanics, the probability that the ball passes over the hill is exactly 1—it makes it over every time. If, however, the height of the hill is increased—a ball placed atop the hill has a potential energy of 200 J—the ball proceeds only part of the way up the hill, stops, and returns in the direction it came. The total energy of the ball is converted entirely into potential energy before it can reach the top of the hill. We do not expect, even after repeated attempts, for the 100-J ball to ever be found beyond the hill. Therefore, the probability that the ball passes over the hill is exactly 0, and probability it is turned back or “reflected” by the hill is exactly 1. The ball *never* makes it over the hill. The existence of the ball beyond the hill is an impossibility or “energetically forbidden.”

However, according to quantum mechanics, the ball has a wave function and this function is defined over all space. The wave function may be highly localized, but there is always a chance that as the ball encounters the hill, the ball will suddenly be found beyond it. Indeed, this probability is appreciable if the “wave packet” of the ball is wider than the barrier.

Note:

View this [interactive simulation](#) for a simulation of tunneling.

In the language of quantum mechanics, the hill is characterized by a **potential barrier**. A finite-height square barrier is described by the following potential-energy function:

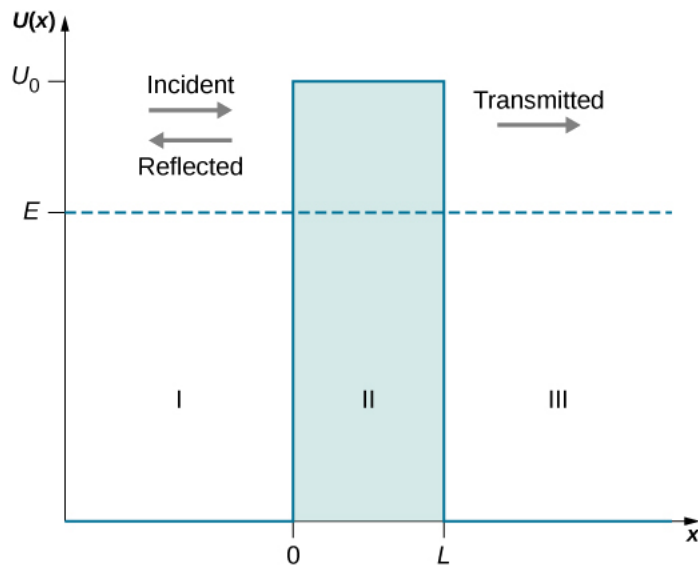
Note:

Equation:

$$U(x) = \begin{cases} 0, & \text{when } x < 0 \\ U_0, & \text{when } 0 \leq x \leq L \\ 0, & \text{when } x > L. \end{cases}$$

The potential barrier is illustrated in [\[link\]](#). When the height U_0 of the barrier is infinite, the wave packet representing an incident quantum particle is unable to penetrate it, and the quantum particle bounces back from the barrier boundary, just like a classical particle. When the width L of the barrier is infinite and its height is finite, a

part of the wave packet representing an incident quantum particle can filter through the barrier boundary and eventually perish after traveling some distance inside the barrier.



A potential energy barrier of height U_0 creates three physical regions with three different wave behaviors. In region I where $x < 0$, an incident wave packet (incident particle) moves in a potential-free zone and coexists with a reflected wave packet (reflected particle). In region II, a part of the incident wave that has not been reflected at $x = 0$ moves as a transmitted wave in a constant potential $U(x) = +U_0$ and tunnels through to region III at $x = L$.

In region III for $x > L$, a wave packet (transmitted particle) that has tunneled through the potential barrier moves as a free particle in potential-free zone. The energy E of the incident particle is indicated by the horizontal line.

When both the width L and the height U_0 are finite, a part of the quantum wave packet incident on one side of the barrier can penetrate the barrier boundary and continue its motion inside the barrier, where it is gradually attenuated on its way to the other side. A part of the incident quantum wave packet eventually emerges on the other side of the barrier in the form of the transmitted wave packet that tunneled through the barrier. How much of the incident wave can tunnel through a barrier depends on the barrier width L and its height U_0 , and on the energy E of the quantum particle incident on the barrier. This is the physics of tunneling.

Barrier penetration by quantum wave functions was first analyzed theoretically by Friedrich Hund in 1927, shortly after Schrödinger published the equation that bears his name. A year later, George Gamow used the formalism of quantum mechanics to explain the radioactive α -decay of atomic nuclei as a quantum-tunneling phenomenon. The invention of the tunnel diode in 1957 made it clear that quantum tunneling is important to the semiconductor industry. In modern nanotechnologies, individual atoms are manipulated using a knowledge of quantum tunneling.

Tunneling and the Wave Function

Suppose a uniform and time-independent beam of electrons or other quantum particles with energy E traveling along the x -axis (in the positive direction to the right) encounters a potential barrier described by [\[link\]](#). The question is: What is the probability that an individual particle in the beam will tunnel through the potential barrier? The answer can be found by solving the boundary-value problem for the time-independent Schrödinger equation for a particle in the beam. The general form of this equation is given by [\[link\]](#), which we reproduce here:

Equation:

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x), \text{ where } -\infty < x < +\infty.$$

In [\[link\]](#), the potential function $U(x)$ is defined by [\[link\]](#). We assume that the given energy E of the incoming particle is smaller than the height U_0 of the potential barrier, $E < U_0$, because this is the interesting physical case. Knowing the energy E of the incoming particle, our task is to solve [\[link\]](#) for a function $\psi(x)$ that is continuous and has continuous first derivatives for all x . In other words, we are looking for a “smooth-looking” solution (because this is how wave functions look) that can be given a probabilistic interpretation so that $|\psi(x)|^2 = \psi^*(x)\psi(x)$ is the probability density.

We divide the real axis into three regions with the boundaries defined by the potential function in [\[link\]](#) (illustrated in [\[link\]](#)) and transcribe [\[link\]](#) for each region. Denoting by $\psi_I(x)$ the solution in region I for $x < 0$, by $\psi_{II}(x)$ the solution in region II for $0 \leq x \leq L$, and by $\psi_{III}(x)$ the solution in region III for $x > L$, the stationary Schrödinger equation has the following forms in these three regions:

Equation:

$$-\frac{\hbar^2}{2m} \frac{d^2\psi_I(x)}{dx^2} = E\psi_I(x), \text{ in region I: } -\infty < x < 0,$$

Equation:

$$-\frac{\hbar^2}{2m} \frac{d^2\psi_{II}(x)}{dx^2} + U_0\psi_{II}(x) = E\psi_{II}(x), \text{ in region II: } 0 \leq x \leq L,$$

Equation:

$$-\frac{\hbar^2}{2m} \frac{d^2\psi_{III}(x)}{dx^2} = E\psi_{III}(x), \text{ in region III: } L < x < +\infty.$$

The continuity condition at region boundaries requires that:

Equation:

$$\psi_I(0) = \psi_{II}(0), \text{ at the boundary between regions I and II and}$$

and

Equation:

$$\psi_{II}(L) = \psi_{III}(L), \text{ at the boundary between regions II and III.}$$

The “smoothness” condition requires the first derivative of the solution be continuous at region boundaries:

Equation:

$$\left. \frac{d\psi_I(x)}{dx} \right|_{x=0} = \left. \frac{d\psi_{II}(x)}{dx} \right|_{x=0}, \text{ at the boundary between regions I and II;}$$

and

Equation:

$$\left. \frac{d\psi_{\text{II}}(x)}{dx} \right|_{x=L} = \left. \frac{d\psi_{\text{III}}(x)}{dx} \right|_{x=L}, \text{ at the boundary between regions II and III.}$$

In what follows, we find the functions $\psi_{\text{I}}(x)$, $\psi_{\text{II}}(x)$, and $\psi_{\text{III}}(x)$.

We can easily verify (by substituting into the original equation and differentiating) that in regions I and III, the solutions must be in the following general forms:

Equation:

$$\psi_{\text{I}}(x) = Ae^{+ikx} + Be^{-ikx}$$

Equation:

$$\psi_{\text{III}}(x) = Fe^{+ikx} + Ge^{-ikx}$$

where $k = \sqrt{2mE}/\hbar$ is a wave number and the complex exponent denotes oscillations,

Equation:

$$e^{\pm ikx} = \cos kx \pm i \sin kx.$$

The constants A , B , F , and G in [\[link\]](#) and [\[link\]](#) may be complex. These solutions are illustrated in [\[link\]](#). In region I, there are two waves—one is incident (moving to the right) and one is reflected (moving to the left)—so none of the constants A and B in [\[link\]](#) may vanish. In region III, there is only one wave (moving to the right), which is the transmitted wave, so the constant G must be zero in [\[link\]](#), $G = 0$. We can write explicitly that the incident wave is $\psi_{\text{in}}(x) = Ae^{+ikx}$ and that the reflected wave is $\psi_{\text{ref}}(x) = Be^{-ikx}$, and that the transmitted wave is $\psi_{\text{tra}}(x) = Fe^{+ikx}$. The amplitude of the incident wave is

Equation:

$$\left| \psi_{\text{in}}(x) \right|^2 = \psi_{\text{in}}^*(x) \psi_{\text{in}}(x) = (Ae^{+ikx})^* Ae^{+ikx} = A^* e^{-ikx} Ae^{+ikx} = A^* A = |A|^2.$$

Similarly, the amplitude of the reflected wave is $|\psi_{\text{ref}}(x)|^2 = |B|^2$ and the amplitude of the transmitted wave is $|\psi_{\text{tra}}(x)|^2 = |F|^2$. We know from the theory of waves that the square of the wave amplitude is directly proportional to the wave intensity. If we want to know how much of the incident wave tunnels through the barrier, we need to compute the square of the amplitude of the transmitted wave. The **transmission probability** or **tunneling probability** is the ratio of the transmitted intensity ($|F|^2$) to the incident intensity ($|A|^2$), written as

Note:

Equation:

$$T(L, E) = \frac{|\psi_{\text{tra}}(x)|^2}{|\psi_{\text{in}}(x)|^2} = \frac{|F|^2}{|A|^2} = \left| \frac{F}{A} \right|^2$$

where L is the width of the barrier and E is the total energy of the particle. This is the probability an individual particle in the incident beam will tunnel through the potential barrier. Intuitively, we understand that this

probability must depend on the barrier height U_0 .

In region II, the terms in equation [\[link\]](#) can be rearranged to

Equation:

$$\frac{d^2\psi_{\text{II}}(x)}{dx^2} = \beta^2\psi_{\text{II}}(x)$$

where β^2 is positive because $U_0 > E$ and the parameter β is a real number,

Note:

Equation:

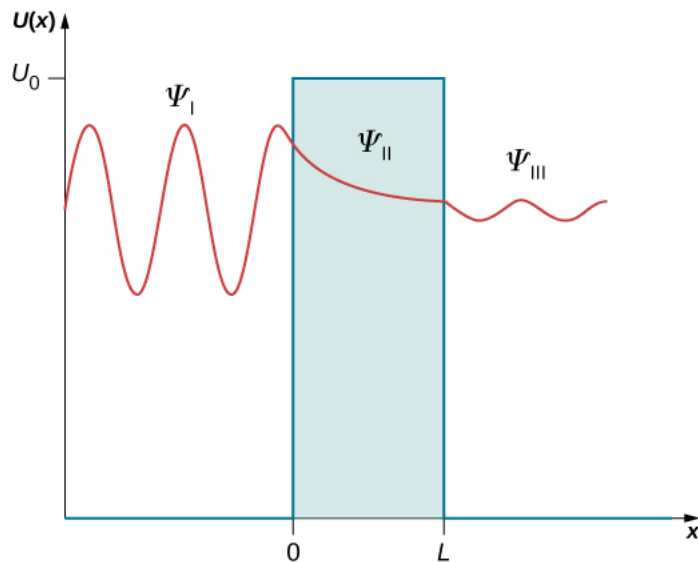
$$\beta^2 = \frac{2m}{\hbar^2}(U_0 - E).$$

The general solution to [\[link\]](#) is not oscillatory (unlike in the other regions) and is in the form of exponentials that describe a gradual attenuation of $\psi_{\text{II}}(x)$,

Equation:

$$\psi_{\text{II}}(x) = Ce^{-\beta x} + De^{+\beta x}.$$

The two types of solutions in the three regions are illustrated in [\[link\]](#).



Three types of solutions to the stationary Schrödinger equation for the quantum-tunneling problem: Oscillatory behavior in regions I and III where a quantum particle moves freely, and exponential-decay behavior in region II (the barrier region) where the particle moves in the potential U_0 .

Now we use the boundary conditions to find equations for the unknown constants. [\[link\]](#) and [\[link\]](#) are substituted into [\[link\]](#) to give

Equation:

$$A + B = C + D.$$

[\[link\]](#) and [\[link\]](#) are substituted into [\[link\]](#) to give

Equation:

$$Ce^{-\beta L} + De^{+\beta L} = Fe^{+ikL}.$$

Similarly, we substitute [\[link\]](#) and [\[link\]](#) into [\[link\]](#), differentiate, and obtain

Equation:

$$-ik(A - B) = \beta(D - C).$$

Similarly, the boundary condition [\[link\]](#) reads explicitly

Equation:

$$\beta(De^{+\beta L} - Ce^{-\beta L}) = -ikFe^{+ikL}.$$

We now have four equations for five unknown constants. However, because the quantity we are after is the transmission coefficient, defined in [\[link\]](#) by the fraction F/A , the number of equations is exactly right because when we divide each of the above equations by A , we end up having only four unknown fractions: B/A , C/A , D/A , and F/A , three of which can be eliminated to find F/A . The actual algebra that leads to expression for F/A is pretty lengthy, but it can be done either by hand or with a help of computer software. The end result is

Equation:

$$\frac{F}{A} = \frac{e^{-ikL}}{\cosh(\beta L) + i(\gamma/2)\sinh(\beta L)}.$$

In deriving [\[link\]](#), to avoid the clutter, we use the substitutions $\gamma \equiv \beta/k - k/\beta$,

Equation:

$$\cosh y = \frac{e^y + e^{-y}}{2}, \text{ and } \sinh y = \frac{e^y - e^{-y}}{2}.$$

We substitute [\[link\]](#) into [\[link\]](#) and obtain the exact expression for the transmission coefficient for the barrier,

Equation:

$$T(L, E) = \left(\frac{F}{A}\right)^* \frac{F}{A} = \frac{e^{+ikL}}{\cosh(\beta L) - i(\gamma/2)\sinh(\beta L)} \cdot \frac{e^{-ikL}}{\cosh(\beta L) + i(\gamma/2)\sinh(\beta L)}$$

or

Note:

Equation:

$$T(L, E) = \frac{1}{\cosh^2(\beta L) + (\gamma/2)^2 \sinh^2(\beta L)}$$

where

Equation:

$$\left(\frac{\gamma}{2}\right)^2 = \frac{1}{4} \left(\frac{1 - E/U_0}{E/U_0} + \frac{E/U_0}{1 - E/U_0} - 2 \right).$$

For a wide and high barrier that transmits poorly, [\[link\]](#) can be approximated by

Note:

Equation:

$$T(L, E) = 16 \frac{E}{U_0} \left(1 - \frac{E}{U_0} \right) e^{-2\beta L}.$$

Whether it is the exact expression [\[link\]](#) or the approximate expression [\[link\]](#), we see that the tunneling effect very strongly depends on the width L of the potential barrier. In the laboratory, we can adjust both the potential height U_0 and the width L to design nano-devices with desirable transmission coefficients.

Example:

Transmission Coefficient

Two copper nanowires are insulated by a copper oxide nano-layer that provides a 10.0-eV potential barrier. Estimate the tunneling probability between the nanowires by 7.00-eV electrons through a 5.00-nm thick oxide layer. What if the thickness of the layer were reduced to just 1.00 nm? What if the energy of electrons were increased to 9.00 eV?

Strategy

Treating the insulating oxide layer as a finite-height potential barrier, we use [\[link\]](#). We identify $U_0 = 10.0$ eV, $E_1 = 7.00$ eV, $E_2 = 9.00$ eV, $L_1 = 5.00$ nm, and $L_2 = 1.00$ nm. We use [\[link\]](#) to compute the exponent. Also, we need the rest mass of the electron $m = 511$ keV/ c^2 and Planck's constant $\hbar = 0.1973$ keV · nm/ c . It is typical for this type of estimate to deal with very small quantities that are often not suitable for handheld calculators. To make correct estimates of orders, we make the conversion $e^y = 10^{y/\ln 10}$.

Solution

Constants:

Equation:

$$\frac{2m}{\hbar^2} = \frac{2(511 \text{ keV}/c^2)}{(0.1973 \text{ keV} \cdot \text{nm}/c)^2} = 26,254 \frac{1}{\text{keV} \cdot (\text{nm})^2},$$

Equation:

$$\beta = \sqrt{\frac{2m}{\hbar^2}(U_0 - E)} = \sqrt{26,254 \frac{(10.0 \text{ eV} - E)}{\text{keV} \cdot (\text{nm})^2}} = \sqrt{26.254(10.0 \text{ eV} - E)/\text{eV}} \frac{1}{\text{nm}}.$$

For a lower-energy electron with $E_1 = 7.00 \text{ eV}$:

Equation:

$$\beta_1 = \sqrt{26.254(10.00 \text{ eV} - E_1)/\text{eV}} \frac{1}{\text{nm}} = \sqrt{26.254(10.00 - 7.00)} \frac{1}{\text{nm}} = \frac{8.875}{\text{nm}},$$

Equation:

$$T(L, E_1) = 16 \frac{E_1}{U_0} \left(1 - \frac{E_1}{U_0}\right) e^{-2\beta_1 L} = 16 \frac{7}{10} \left(1 - \frac{7}{10}\right) e^{-17.75 L/\text{nm}} = 3.36 e^{-17.75 L/\text{nm}}.$$

For a higher-energy electron with $E_2 = 9.00 \text{ eV}$:

Equation:

$$\beta_2 = \sqrt{26.254(10.00 \text{ eV} - E_2)/\text{eV}} \frac{1}{\text{nm}} = \sqrt{26.254(10.00 - 9.00)} \frac{1}{\text{nm}} = \frac{5.124}{\text{nm}},$$

Equation:

$$T(L, E_2) = 16 \frac{E_2}{U_0} \left(1 - \frac{E_2}{U_0}\right) e^{-2\beta_2 L} = 16 \frac{9}{10} \left(1 - \frac{9}{10}\right) e^{-10.25 L/\text{nm}} = 1.44 e^{-10.25 L/\text{nm}}.$$

For a broad barrier with $L_1 = 5.00 \text{ nm}$:

Equation:

$$T(L_1, E_1) = 3.36 e^{-17.75 L_1/\text{nm}} = 3.36 e^{-17.75 \cdot 5.00 \text{ nm}/\text{nm}} = 3.36 e^{-88} = 3.36(6.2 \times 10^{-39}) = 2.1\% \times 10^{-36},$$

Equation:

$$T(L_1, E_2) = 1.44 e^{-10.25 L_1/\text{nm}} = 1.44 e^{-10.25 \cdot 5.00 \text{ nm}/\text{nm}} = 1.44 e^{-51.2} = 1.44(5.81 \times 10^{-12}) = 8.36\% \times 10^{-2}$$

For a narrower barrier with $L_2 = 1.00 \text{ nm}$:

Equation:

$$T(L_2, E_1) = 3.36 e^{-17.75 L_2/\text{nm}} = 3.36 e^{-17.75 \cdot 1.00 \text{ nm}/\text{nm}} = 3.36 e^{-17.75} = 3.36(5.1 \times 10^{-7}) = 1.7\% \times 10^{-4},$$

Equation:

$$T(L_2, E_2) = 1.44 e^{-10.25 L_2/\text{nm}} = 1.44 e^{-10.25 \cdot 1.00 \text{ nm}/\text{nm}} = 1.44 e^{-10.25} = 1.44(3.53 \times 10^{-5}) = 5.09\% \times 10^{-7}.$$

Significance

We see from these estimates that the probability of tunneling is affected more by the width of the potential barrier than by the energy of an incident particle. In today's technologies, we can manipulate individual atoms on metal surfaces to create potential barriers that are fractions of a nanometer, giving rise to measurable tunneling currents. One of many applications of this technology is the scanning tunneling microscope (STM), which we discuss later in this section.

Note:

Exercise:

Problem:

Check Your Understanding A proton with kinetic energy 1.00 eV is incident on a square potential barrier with height 10.00 eV. If the proton is to have the same transmission probability as an electron of the same energy, what must the width of the barrier be relative to the barrier width encountered by an electron?

Solution:

$$L_{\text{proton}}/L_{\text{electron}} = \sqrt{m_e/m_p} = 2.3\%$$

Radioactive Decay

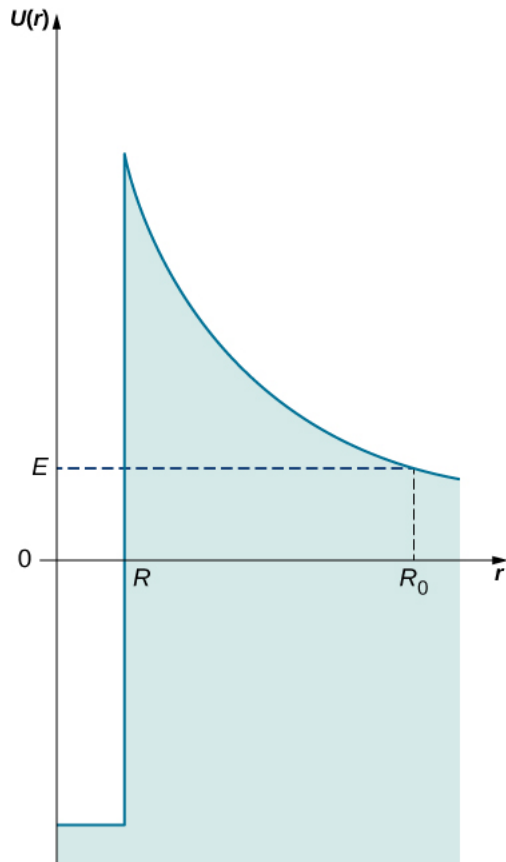
In 1928, Gamow identified quantum tunneling as the mechanism responsible for the radioactive decay of atomic nuclei. He observed that some isotopes of thorium, uranium, and bismuth disintegrate by emitting α -particles (which are doubly ionized helium atoms or, simply speaking, helium nuclei). In the process of emitting an α -particle, the original nucleus is transformed into a new nucleus that has two fewer neutrons and two fewer protons than the original nucleus. The α -particles emitted by one isotope have approximately the same kinetic energies. When we look at variations of these energies among isotopes of various elements, the lowest kinetic energy is about 4 MeV and the highest is about 9 MeV, so these energies are of the same order of magnitude. This is about where the similarities between various isotopes end.

When we inspect half-lives (a half-life is the time in which a radioactive sample loses half of its nuclei due to decay), different isotopes differ widely. For example, the half-life of polonium-214 is 160 μs and the half-life of uranium is 4.5 billion years. Gamow explained this variation by considering a ‘spherical-box’ model of the nucleus, where α -particles can bounce back and forth between the walls as free particles. The confinement is provided by a strong nuclear potential at a spherical wall of the box. The thickness of this wall, however, is not infinite but finite, so in principle, a nuclear particle has a chance to escape this nuclear confinement. On the inside wall of the confining barrier is a high nuclear potential that keeps the α -particle in a small confinement. But when an α -particle gets out to the other side of this wall, it is subject to electrostatic Coulomb repulsion and moves away from the nucleus. This idea is illustrated in [\[link\]](#). The width L of the potential barrier that separates an α -particle from the outside world depends on the particle’s kinetic energy E . This width is the distance between the point marked by the nuclear radius R and the point R_0 where an α -particle emerges on the other side of the barrier, $L = R_0 - R$. At the distance R_0 , its kinetic energy must at least match the electrostatic energy of repulsion, $E = (4\pi\epsilon_0)^{-1}Ze^2/R_0$ (where $+Ze$ is the charge of the nucleus). In this way we can estimate the width of the nuclear barrier,

Equation:

$$L = \frac{e^2}{4\pi\epsilon_0} \frac{Z}{E} - R.$$

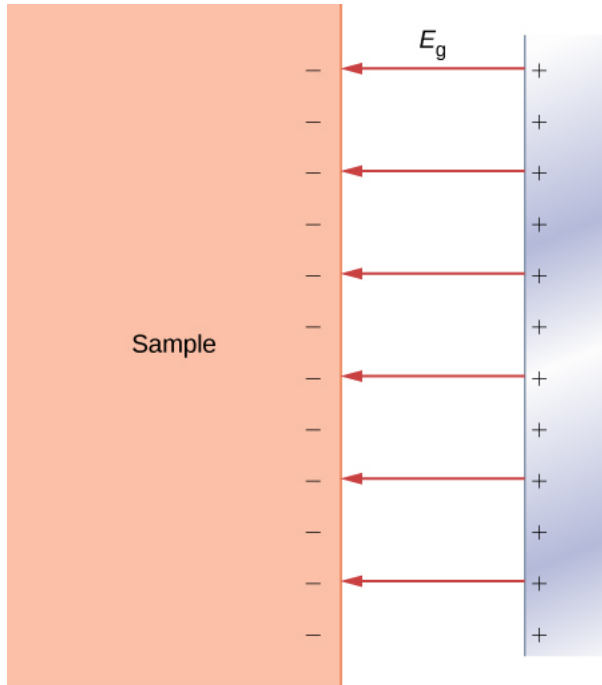
We see from this estimate that the higher the energy of α -particle, the narrower the width of the barrier that it is to tunnel through. We also know that the width of the potential barrier is the most important parameter in tunneling probability. Thus, highly energetic α -particles have a good chance to escape the nucleus, and, for such nuclei, the nuclear disintegration half-life is short. Notice that this process is highly nonlinear, meaning a small increase in the α -particle energy has a disproportionately large enhancing effect on the tunneling probability and, consequently, on shortening the half-life. This explains why the half-life of polonium that emits 8-MeV α -particles is only hundreds of milliseconds and the half-life of uranium that emits 4-MeV α -particles is billions of years.



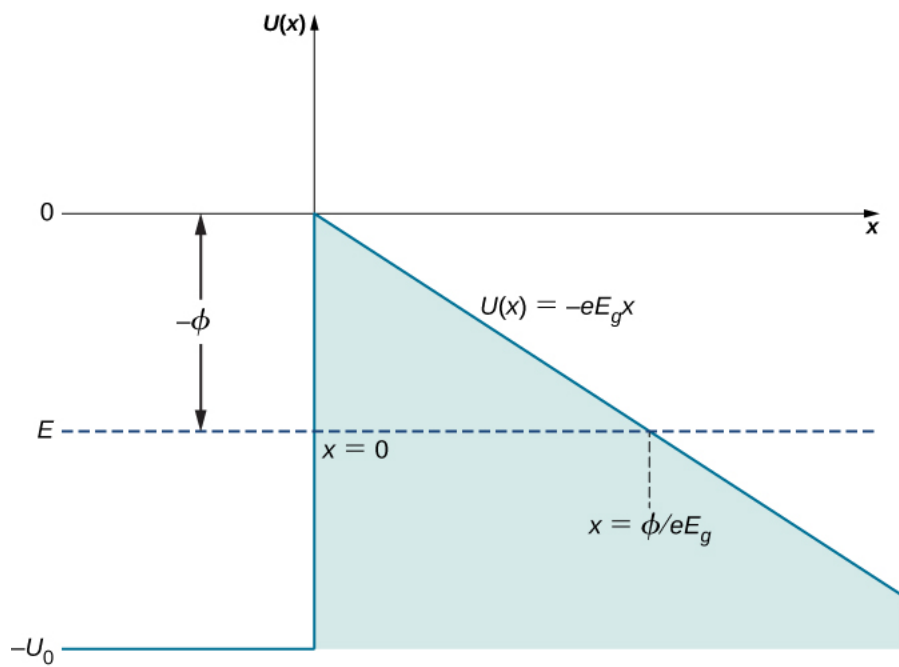
The potential energy barrier for an α -particle bound in the nucleus: To escape from the nucleus, an α -particle with energy E must tunnel across the barrier from distance R to distance R_0 away from the center.

Field Emission

Field emission is a process of emitting electrons from conducting surfaces due to a strong external electric field that is applied in the direction normal to the surface ([\[link\]](#)). As we know from our study of electric fields in earlier chapters, an applied external electric field causes the electrons in a conductor to move to its surface and stay there as long as the present external field is not excessively strong. In this situation, we have a constant electric potential throughout the inside of the conductor, including its surface. In the language of potential energy, we say that an electron inside the conductor has a constant potential energy $U(x) = -U_0$ (here, the x means inside the conductor). In the situation represented in [\[link\]](#), where the external electric field is uniform and has magnitude E_g , if an electron happens to be outside the conductor at a distance x away from its surface, its potential energy would have to be $U(x) = -eE_g x$ (here, x denotes distance to the surface). Taking the origin at the surface, so that $x = 0$ is the location of the surface, we can represent the potential energy of conduction electrons in a metal as the potential energy barrier shown in [\[link\]](#). In the absence of the external field, the potential energy becomes a step barrier defined by $U(x \leq 0) = -U_0$ and by $U(x > 0) = 0$.



A normal-direction external electric field at the surface of a conductor: In a strong field, the electrons on a conducting surface may get detached from it and accelerate against the external electric field away from the surface.



The potential energy barrier at the surface of a metallic conductor in the presence of an external uniform electric field E_g normal to the surface: It becomes a step-function barrier when the external field is removed. The work function of the metal is indicated by ϕ .

When an external electric field is strong, conduction electrons at the surface may get detached from it and accelerate along electric field lines in a direction antiparallel to the external field, away from the surface. In short, conduction electrons may escape from the surface. The field emission can be understood as the quantum tunneling of conduction electrons through the potential barrier at the conductor's surface. The physical principle at work here is very similar to the mechanism of α -emission from a radioactive nucleus.

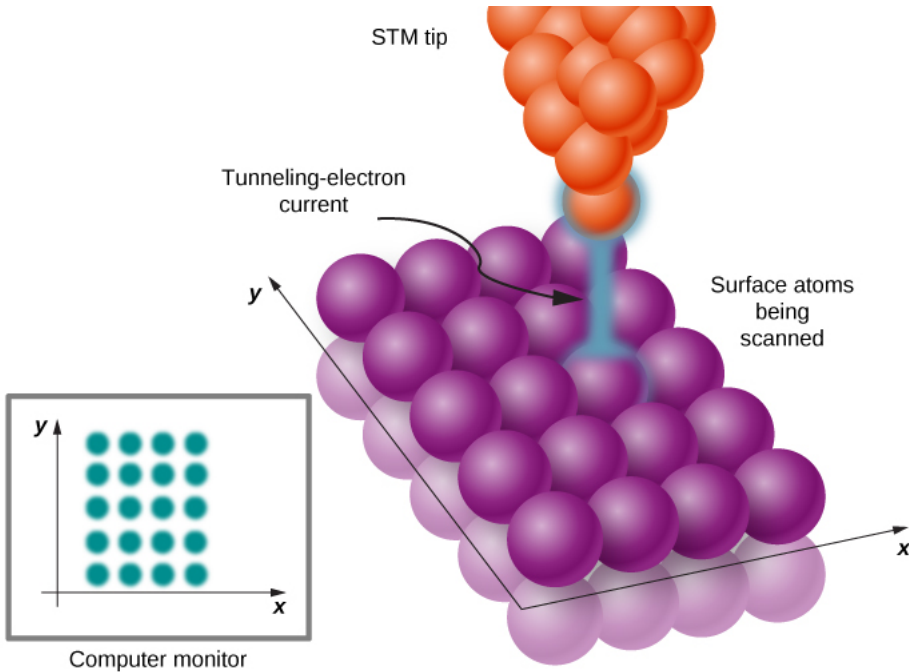
Suppose a conduction electron has a kinetic energy E (the average kinetic energy of an electron in a metal is the work function ϕ for the metal and can be measured, as discussed for the photoelectric effect in [Photons and Matter Waves](#)), and an external electric field can be locally approximated by a uniform electric field of strength E_g . The width L of the potential barrier that the electron must cross is the distance from the conductor's surface to the point outside the surface where its kinetic energy matches the value of its potential energy in the external field. In [\[link\]](#), this distance is measured along the dashed horizontal line $U(x) = E$ from $x = 0$ to the intercept with $U(x) = -eE_g x$, so the barrier width is

Equation:

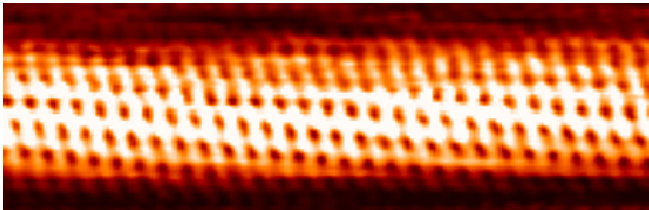
$$L = \frac{e^{-1}E}{E_g} = \frac{e^{-1}\phi}{E_g}.$$

We see that L is inversely proportional to the strength E_g of an external field. When we increase the strength of the external field, the potential barrier outside the conductor becomes steeper and its width decreases for an electron with a given kinetic energy. In turn, the probability that an electron will tunnel across the barrier (conductor surface) becomes exponentially larger. The electrons that emerge on the other side of this barrier form a current (tunneling-electron current) that can be detected above the surface. The tunneling-electron current is proportional to the tunneling probability. The tunneling probability depends nonlinearly on the barrier width L , and L can be changed by adjusting E_g . Therefore, the tunneling-electron current can be tuned by adjusting the strength of an external electric field at the surface. When the strength of an external electric field is constant, the tunneling-electron current has different values at different elevations L above the surface.

The quantum tunneling phenomenon at metallic surfaces, which we have just described, is the physical principle behind the operation of the **scanning tunneling microscope (STM)**, invented in 1981 by Gerd Binnig and Heinrich Rohrer. The STM device consists of a scanning tip (a needle, usually made of tungsten, platinum-iridium, or gold); a piezoelectric device that controls the tip's elevation in a typical range of 0.4 to 0.7 nm above the surface to be scanned; some device that controls the motion of the tip along the surface; and a computer to display images. While the sample is kept at a suitable voltage bias, the scanning tip moves along the surface ([\[link\]](#)), and the tunneling-electron current between the tip and the surface is registered at each position. The amount of the current depends on the probability of electron tunneling from the surface to the tip, which, in turn, depends on the elevation of the tip above the surface. Hence, at each tip position, the distance from the tip to the surface is measured by measuring how many electrons tunnel out from the surface to the tip. This method can give an unprecedented resolution of about 0.001 nm, which is about 1% of the average diameter of an atom. In this way, we can see individual atoms on the surface, as in the image of a carbon nanotube in [\[link\]](#).



In STM, a surface at a constant potential is being scanned by a narrow tip moving along the surface. When the STM tip moves close to surface atoms, electrons can tunnel from the surface to the tip. This tunneling-electron current is continually monitored while the tip is in motion. The amount of current at location (x,y) gives information about the elevation of the tip above the surface at this location. In this way, a detailed topographical map of the surface is created and displayed on a computer monitor.



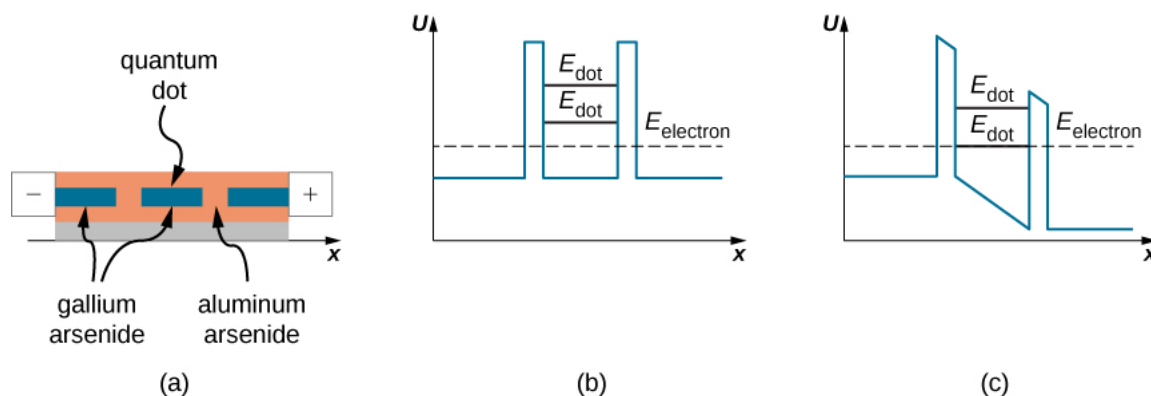
An STM image of a carbon nanotube: Atomic-scale resolution allows us to see individual atoms on the surface. STM images are in gray scale, and coloring is added to bring up details to the human eye. (credit: Taner Yildirim, NIST)

Resonant Quantum Tunneling

Quantum tunneling has numerous applications in semiconductor devices such as electronic circuit components or integrated circuits that are designed at nanoscales; hence, the term '**nanotechnology**.' For example, a diode (an

electric-circuit element that causes an electron current in one direction to be different from the current in the opposite direction, when the polarity of the bias voltage is reversed) can be realized by a tunneling junction between two different types of semiconducting materials. In such a **tunnel diode**, electrons tunnel through a single potential barrier at a contact between two different semiconductors. At the junction, tunneling-electron current changes nonlinearly with the applied potential difference across the junction and may rapidly decrease as the bias voltage is increased. This is unlike the Ohm's law behavior that we are familiar with in household circuits. This kind of rapid behavior (caused by quantum tunneling) is desirable in high-speed electronic devices.

Another kind of electronic nano-device utilizes **resonant tunneling** of electrons through potential barriers that occur in quantum dots. A **quantum dot** is a small region of a semiconductor nanocrystal that is grown, for example, in a silicon or aluminum arsenide crystal. [\[link\]](#)(a) shows a quantum dot of gallium arsenide embedded in an aluminum arsenide wafer. The quantum-dot region acts as a potential well of a finite height (shown in [\[link\]](#)(b)) that has two finite-height potential barriers at dot boundaries. Similarly, as for a quantum particle in a box (that is, an infinite potential well), lower-lying energies of a quantum particle trapped in a finite-height potential well are quantized. The difference between the box and the well potentials is that a quantum particle in a box has an infinite number of quantized energies and is trapped in the box indefinitely, whereas a quantum particle trapped in a potential well has a finite number of quantized energy levels and can tunnel through potential barriers at well boundaries to the outside of the well. Thus, a quantum dot of gallium arsenide sitting in aluminum arsenide is a potential well where low-lying energies of an electron are quantized, indicated as E_{dot} in part (b) in the figure. When the energy E_{electron} of an electron in the outside region of the dot does not match its energy E_{dot} that it would have in the dot, the electron does not tunnel through the region of the dot and there is no current through such a circuit element, even if it were kept at an electric voltage difference (bias). However, when this voltage bias is changed in such a way that one of the barriers is lowered, so that E_{dot} and E_{electron} become aligned, as seen in part (c) of the figure, an electron current flows through the dot. When the voltage bias is now increased, this alignment is lost and the current stops flowing. When the voltage bias is increased further, the electron tunneling becomes improbable until the bias voltage reaches a value for which the outside electron energy matches the next electron energy level in the dot. The word 'resonance' in the device name means that the tunneling-electron current occurs only when a selected energy level is matched by tuning an applied voltage bias, such as in the operation mechanism of the **resonant-tunneling diode** just described. Resonant-tunneling diodes are used as super-fast nano-switches.



Resonant-tunneling diode: (a) A quantum dot of gallium arsenide embedded in aluminum arsenide. (b) Potential well consisting of two potential barriers of a quantum dot with no voltage bias. Electron energies E_{electron} in aluminum arsenide are not aligned with their energy levels E_{dot} in the quantum dot, so electrons do not tunnel through the dot. (c) Potential well of the dot with a voltage bias across the device. A suitably tuned voltage difference distorts the well so that electron-energy levels in the dot are aligned with their energies in aluminum arsenide, causing the electrons to tunnel through the dot.

Summary

- A quantum particle that is incident on a potential barrier of a finite width and height may cross the barrier and appear on its other side. This phenomenon is called ‘quantum tunneling.’ It does not have a classical analog.
- To find the probability of quantum tunneling, we assume the energy of an incident particle and solve the stationary Schrödinger equation to find wave functions inside and outside the barrier. The tunneling probability is a ratio of squared amplitudes of the wave past the barrier to the incident wave.
- The tunneling probability depends on the energy of the incident particle relative to the height of the barrier and on the width of the barrier. It is strongly affected by the width of the barrier in a nonlinear, exponential way so that a small change in the barrier width causes a disproportionately large change in the transmission probability.
- Quantum-tunneling phenomena govern radioactive nuclear decays. They are utilized in many modern technologies such as STM and nano-electronics. STM allows us to see individual atoms on metal surfaces. Electron-tunneling devices have revolutionized electronics and allow us to build fast electronic devices of miniature sizes.

Key Equations

Normalization condition in one dimension	$P(x = -\infty, +\infty) = \int_{-\infty}^{\infty} \Psi(x, t) ^2 dx = 1$
Probability of finding a particle in a narrow interval of position in one dimension $(x, x + dx)$	$P(x, x + dx) = \Psi^*(x, t)\Psi(x, t)dx$
Expectation value of position in one dimension	$\langle x \rangle = \int_{-\infty}^{\infty} \Psi^*(x, t)x\Psi(x, t)dx$
Heisenberg’s position-momentum uncertainty principle	$\Delta x \Delta p \geq \frac{\hbar}{2}$
Heisenberg’s energy-time uncertainty principle	$\Delta E \Delta t \geq \frac{\hbar}{2}$
Schrödinger’s time-dependent equation	$-\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} + U(x, t)\Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}$
General form of the wave function for a time-independent potential in one dimension	$\Psi(x, t) = \psi(x)e^{-i\omega t}$
Schrödinger’s time-independent equation	$-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x)$
Schrödinger’s equation (free particle)	$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} = E\psi(x)$
Allowed energies (particle in box of length L)	$E_n = n^2 \frac{\pi^2 \hbar^2}{2mL^2}, n = 1, 2, 3, \dots$
Stationary states (particle in a box of length L)	$\psi_n(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L}, n = 1, 2, 3, \dots$

Potential-energy function of a harmonic oscillator	$U(x) = \frac{1}{2}m\omega^2 x^2$
Schrödinger equation (harmonic oscillator)	$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + \frac{1}{2}m\omega^2 x^2 \psi(x) = E\psi(x)$
The energy spectrum	$E_n = \left(n + \frac{1}{2}\right)\hbar\omega, n = 0, 1, 2, 3, \dots$
The energy wave functions	$\psi_n(x) = N_n e^{-\beta^2 x^2/2} H_n(\beta x), n = 0, 1, 2, 3, \dots$
Potential barrier	$U(x) = \begin{cases} 0, & \text{when } x < 0 \\ U_0, & \text{when } 0 \leq x \leq L \\ 0, & \text{when } x > L \end{cases}$
Definition of the transmission coefficient	$T(L, E) = \frac{ \psi_{\text{tra}}(x) ^2}{ \psi_{\text{in}}(x) ^2}$
A parameter in the transmission coefficient	$\beta^2 = \frac{2m}{\hbar^2} (U_0 - E)$
Transmission coefficient, exact	$T(L, E) = \frac{1}{\cosh^2 \beta L + (\gamma/2)^2 \sinh^2 \beta L}$
Transmission coefficient, approximate	$T(L, E) = 16 \frac{E}{U_0} \left(1 - \frac{E}{U_0}\right) e^{-2\beta L}$

Conceptual Questions

Exercise:

Problem:

When an electron and a proton of the same kinetic energy encounter a potential barrier of the same height and width, which one of them will tunnel through the barrier more easily? Why?

Exercise:

Problem:

What decreases the tunneling probability most: doubling the barrier width or halving the kinetic energy of the incident particle?

Solution:

doubling the barrier width

Exercise:

Problem: Explain the difference between a box-potential and a potential of a quantum dot.

Exercise:

Problem: Can a quantum particle ‘escape’ from an infinite potential well like that in a box? Why? Why not?

Solution:

No, the restoring force on the particle at the walls of an infinite square well is infinity.

Exercise:

Problem:

A tunnel diode and a resonant-tunneling diode both utilize the same physics principle of quantum tunneling. In what important way are they different?

Problems**Exercise:**

Problem: Show that the wave function in (a) [\[link\]](#) satisfies [\[link\]](#), and (b) [\[link\]](#) satisfies [\[link\]](#).

Solution:

A complex function of the form, $Ae^{i\phi}$, satisfies Schrödinger's time-independent equation. The operators for kinetic and total energy are linear, so any linear combination of such wave functions is also a valid solution to Schrödinger's equation. Therefore, we conclude that [\[link\]](#) satisfies [\[link\]](#), and [\[link\]](#) satisfies [\[link\]](#).

Exercise:**Problem:**

A 6.0-eV electron impacts on a barrier with height 11.0 eV. Find the probability of the electron to tunnel through the barrier if the barrier width is (a) 0.80 nm and (b) 0.40 nm.

Exercise:**Problem:**

A 5.0-eV electron impacts on a barrier of width 0.60 nm. Find the probability of the electron to tunnel through the barrier if the barrier height is (a) 7.0 eV; (b) 9.0 eV; and (c) 13.0 eV.

Solution:

a. 4.21%; b. 0.84%; c. 0.06%

Exercise:**Problem:**

A 12.0-eV electron encounters a barrier of height 15.0 eV. If the probability of the electron tunneling through the barrier is 2.5 %, find its width.

Exercise:**Problem:**

A quantum particle with initial kinetic energy 32.0 eV encounters a square barrier with height 41.0 eV and width 0.25 nm. Find probability that the particle tunnels through this barrier if the particle is (a) an electron and, (b) a proton.

Solution:

a. 0.13%; b. close to 0%

Exercise:

Problem:

A simple model of a radioactive nuclear decay assumes that α -particles are trapped inside a well of nuclear potential that walls are the barriers of a finite width 2.0 fm and height 30.0 MeV. Find the tunneling probability across the potential barrier of the wall for α -particles having kinetic energy (a) 29.0 MeV and (b) 20.0 MeV. The mass of the α -particle is $m = 6.64 \times 10^{-27}$ kg.

Exercise:**Problem:**

A muon, a quantum particle with a mass approximately 200 times that of an electron, is incident on a potential barrier of height 10.0 eV. The kinetic energy of the impacting muon is 5.5 eV and only about 0.10% of the squared amplitude of its incoming wave function filters through the barrier. What is the barrier's width?

Solution:

0.38 nm

Exercise:**Problem:**

A grain of sand with mass 1.0 mg and kinetic energy 1.0 J is incident on a potential energy barrier with height 1.000001 J and width 2500 nm. How many grains of sand have to fall on this barrier before, on the average, one passes through?

Additional Problems**Exercise:****Problem:**

Show that if the uncertainty in the position of a particle is on the order of its de Broglie's wavelength, then the uncertainty in its momentum is on the order of the value of its momentum.

Solution:

proof

Exercise:**Problem:**

The mass of a ρ -meson is measured to be $770 \text{ MeV}/c^2$ with an uncertainty of $100 \text{ MeV}/c^2$. Estimate the lifetime of this meson.

Exercise:**Problem:**

A particle of mass m is confined to a box of width L . If the particle is in the first excited state, what are the probabilities of finding the particle in a region of width $0.020 L$ around the given point x : (a) $x = 0.25L$; (b) $x = 0.40L$; (c) $x = 0.75L$; and (d) $x = 0.90L$.

Solution:

a. 4.0 %; b. 1.4 %; c. 4.0%; d. 1.4%

Exercise:

Problem: A particle in a box $[0;L]$ is in the third excited state. What are its most probable positions?

Exercise:

Problem:

A 0.20-kg billiard ball bounces back and forth without losing its energy between the cushions of a 1.5 m long table. (a) If the ball is in its ground state, how many years does it need to get from one cushion to the other? You may compare this time interval to the age of the universe. (b) How much energy is required to make the ball go from its ground state to its first excited state? Compare it with the kinetic energy of the ball moving at 2.0 m/s.

Solution:

a. $t = mL^2/h = 2.15 \times 10^{26}$ years; b. $E_1 = 1.46 \times 10^{-66}$ J, $K = 0.4$ J

Exercise:

Problem:

Find the expectation value of the position squared when the particle in the box is in its third excited state and the length of the box is L .

Exercise:

Problem:

Consider an infinite square well with wall boundaries $x = 0$ and $x = L$. Show that the function $\psi(x) = A \sin kx$ is the solution to the stationary Schrödinger equation for the particle in a box only if $k = \sqrt{2mE}/\hbar$. Explain why this is an acceptable wave function only if k is an integer multiple of π/L .

Solution:

proof

Exercise:

Problem:

Consider an infinite square well with wall boundaries $x = 0$ and $x = L$. Explain why the function $\psi(x) = A \cos kx$ is not a solution to the stationary Schrödinger equation for the particle in a box.

Exercise:

Problem:

Atoms in a crystal lattice vibrate in simple harmonic motion. Assuming a lattice atom has a mass of 9.4×10^{-26} kg, what is the force constant of the lattice if a lattice atom makes a transition from the ground state to first excited state when it absorbs a 525- μ m photon?

Solution:

1.2 N/m

Exercise:

Problem:

A diatomic molecule behaves like a quantum harmonic oscillator with the force constant 12.0 N/m and mass 5.60×10^{-26} kg. (a) What is the wavelength of the emitted photon when the molecule makes the transition from the third excited state to the second excited state? (b) Find the ground state energy of vibrations for this diatomic molecule.

Exercise:**Problem:**

An electron with kinetic energy 2.0 MeV encounters a potential energy barrier of height 16.0 MeV and width 2.00 nm. What is the probability that the electron emerges on the other side of the barrier?

Solution:

0

Exercise:**Problem:**

A beam of mono-energetic protons with energy 2.0 MeV falls on a potential energy barrier of height 20.0 MeV and of width 1.5 fm. What percentage of the beam is transmitted through the barrier?

Challenge Problems**Exercise:****Problem:**

An electron in a long, organic molecule used in a dye laser behaves approximately like a quantum particle in a box with width 4.18 nm. Find the emitted photon when the electron makes a transition from the first excited state to the ground state and from the second excited state to the first excited state.

Solution:

19.2 μm ; 11.5 μm

Exercise:**Problem:**

In STM, an elevation of the tip above the surface being scanned can be determined with a great precision, because the tunneling-electron current between surface atoms and the atoms of the tip is extremely sensitive to the variation of the separation gap between them from point to point along the surface. Assuming that the tunneling-electron current is in direct proportion to the tunneling probability and that the tunneling probability is to a good approximation expressed by the exponential function $e^{-2\beta L}$ with $\beta = 10.0/\text{nm}$, determine the ratio of the tunneling current when the tip is 0.500 nm above the surface to the current when the tip is 0.515 nm above the surface.

Exercise:**Problem:**

If STM is to detect surface features with local heights of about 0.00200 nm, what percent change in tunneling-electron current must the STM electronics be able to detect? Assume that the tunneling-electron current has characteristics given in the preceding problem.

Solution:

3.92%

Exercise:**Problem:**

Use Heisenberg's uncertainty principle to estimate the ground state energy of a particle oscillating on a spring with angular frequency, $\omega = \sqrt{k/m}$, where k is the spring constant and m is the mass.

Exercise:**Problem:**

Suppose an infinite square well extends from $-L/2$ to $+L/2$. Solve the time-independent Schrödinger's equation to find the allowed energies and stationary states of a particle with mass m that is confined to this well. Then show that these solutions can be obtained by making the coordinate transformation $x' = x - L/2$ for the solutions obtained for the well extending between 0 and L .

Solution:

proof

Exercise:**Problem:**

A particle of mass m confined to a box of width L is in its first excited state $\psi_2(x)$. (a) Find its average position (which is the expectation value of the position). (b) Where is the particle most likely to be found?

Glossary

field emission

electron emission from conductor surfaces when a strong external electric field is applied in normal direction to conductor's surface

nanotechnology

technology that is based on manipulation of nanostructures such as molecules or individual atoms to produce nano-devices such as integrated circuits

potential barrier

potential function that rises and falls with increasing values of position

quantum dot

small region of a semiconductor nanocrystal embedded in another semiconductor nanocrystal, acting as a potential well for electrons

quantum tunneling

phenomenon where particles penetrate through a potential energy barrier with a height greater than the total energy of the particles

resonant tunneling

tunneling of electrons through a finite-height potential well that occurs only when electron energies match an energy level in the well, occurs in quantum dots

resonant-tunneling diode

quantum dot with an applied voltage bias across it

scanning tunneling microscope (STM)

device that utilizes quantum-tunneling phenomenon at metallic surfaces to obtain images of nanoscale structures

transmission probability

also called tunneling probability, the probability that a particle will tunnel through a potential barrier

tunnel diode

electron tunneling-junction between two different semiconductors

tunneling probability

also called transmission probability, the probability that a particle will tunnel through a potential barrier

Introduction

class="introduction"

NGC1763 is
an emission
nebula in
the Large
Magellanic
Cloud,
which is a
satellite
galaxy to
our Milky
Way

Galaxy. The
colors we
see can be
explained
by applying
the ideas of
quantum
mechanics
to atomic
structure.

(credit:
modification
of work
by NASA,
ESA, and
Josh Lake)



In this chapter, we use quantum mechanics to study the structure and properties of atoms. This study introduces ideas and concepts that are necessary to understand more complex systems, such as molecules, crystals, and metals. As we deepen our understanding of atoms, we build on things we already know, such as Rutherford's nuclear model of the atom, Bohr's model of the hydrogen atom, and de Broglie's wave hypothesis.

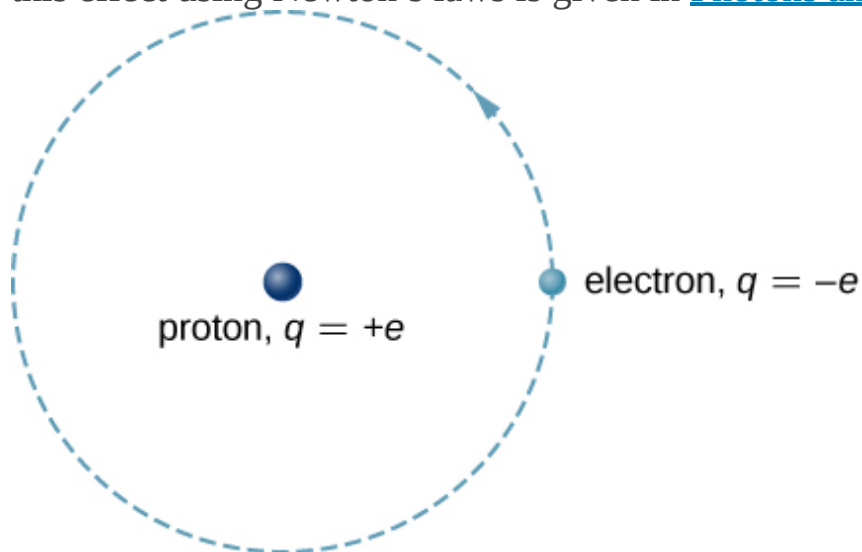
[\[link\]](#) is NGC1763, an emission nebula in the small galaxy known as the Large Magellanic Cloud, which is a satellite of the Milky Way Galaxy. Ultraviolet light from hot stars ionizes the hydrogen atoms in the nebula. As protons and electrons recombine, radiation of different frequencies is emitted. The details of this process can be correctly predicted by quantum mechanics and are examined in this chapter.

The Hydrogen Atom

By the end of this section, you will be able to:

- Describe the hydrogen atom in terms of wave function, probability density, total energy, and orbital angular momentum
- Identify the physical significance of each of the quantum numbers (n, l, m) of the hydrogen atom
- Distinguish between the Bohr and Schrödinger models of the atom
- Use quantum numbers to calculate important information about the hydrogen atom

The hydrogen atom is the simplest atom in nature and, therefore, a good starting point to study atoms and atomic structure. The hydrogen atom consists of a single negatively charged electron that moves about a positively charged proton ([\[link\]](#)). In Bohr's model, the electron is pulled around the proton in a perfectly circular orbit by an attractive Coulomb force. The proton is approximately 1800 times more massive than the electron, so the proton moves very little in response to the force on the proton by the electron. (This is analogous to the Earth-Sun system, where the Sun moves very little in response to the force exerted on it by Earth.) An explanation of this effect using Newton's laws is given in [Photons and Matter Waves](#).



A representation of the Bohr model of the hydrogen atom.

With the assumption of a fixed proton, we focus on the motion of the electron.

In the electric field of the proton, the potential energy of the electron is
Equation:

$$U(r) = -k \frac{e^2}{r},$$

where $k = 1/4\pi\epsilon_0$ and r is the distance between the electron and the proton. As we saw earlier, the force on an object is equal to the negative of the gradient (or slope) of the potential energy function. For the special case of a hydrogen atom, the force between the electron and proton is an attractive Coulomb force.

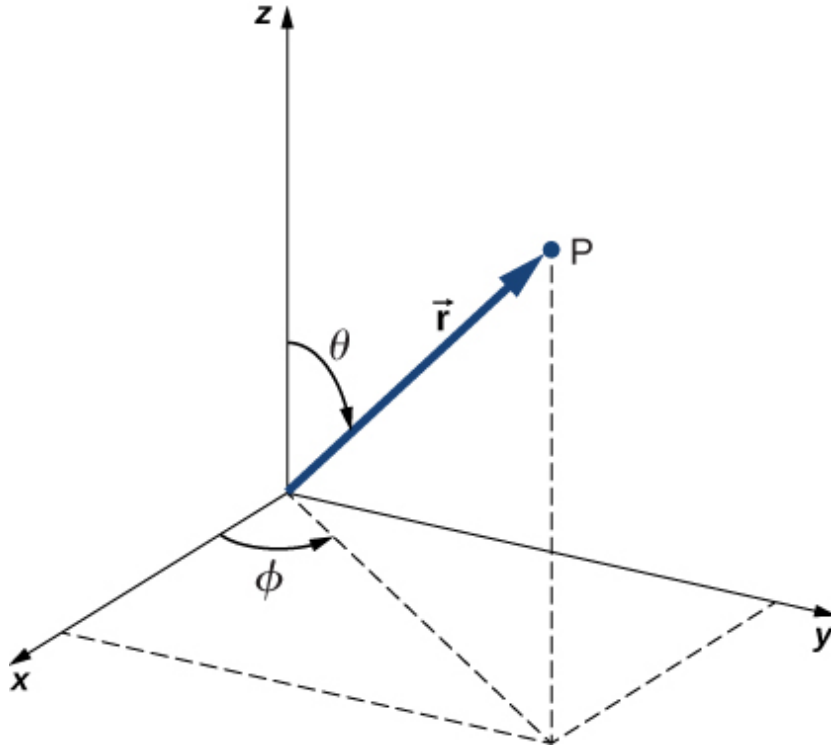
Notice that the potential energy function $U(r)$ does not vary in time. As a result, Schrödinger's equation of the hydrogen atom reduces to two simpler equations: one that depends only on space (x, y, z) and another that depends only on time (t). (The separation of a wave function into space- and time-dependent parts for time-independent potential energy functions is discussed in [Quantum Mechanics](#).) We are most interested in the space-dependent equation:

Equation:

$$\frac{-\hbar^2}{2m_e} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) - k \frac{e^2}{r} \psi = E\psi,$$

where $\psi = \psi(x, y, z)$ is the three-dimensional wave function of the electron, m_e is the mass of the electron, and E is the total energy of the electron. Recall that the total wave function $\Psi(x, y, z, t)$, is the product of the space-dependent wave function $\psi = \psi(x, y, z)$ and the time-dependent wave function $\varphi = \varphi(t)$.

In addition to being time-independent, $U(r)$ is also spherically symmetrical. This suggests that we may solve Schrödinger's equation more easily if we express it in terms of the spherical coordinates (r, θ, ϕ) instead of rectangular coordinates (x, y, z) . A spherical coordinate system is shown in [\[link\]](#). In spherical coordinates, the variable r is the radial coordinate, θ is the polar angle (relative to the vertical z -axis), and ϕ is the azimuthal angle (relative to the x -axis). The relationship between spherical and rectangular coordinates is $x = r \sin \theta \cos \phi$, $y = r \sin \theta \sin \phi$, $z = r \cos \theta$.



The relationship between the spherical and rectangular coordinate systems.

The factor $r \sin \theta$ is the magnitude of a vector formed by the projection of the polar vector onto the xy -plane. Also, the coordinates of x and y are obtained by projecting this vector onto the x - and y -axes, respectively. The inverse transformation gives

Equation:

$$r = \sqrt{x^2 + y^2 + z^2}, \quad \theta = \cos^{-1} \left(\frac{z}{r} \right), \quad \phi = \cos^{-1} \left(\frac{x}{\sqrt{x^2 + y^2}} \right).$$

Schrödinger's wave equation for the hydrogen atom in spherical coordinates is discussed in more advanced courses in modern physics, so we do not consider it in detail here. However, due to the spherical symmetry of $U(r)$, this equation reduces to three simpler equations: one for each of the three coordinates (r , θ , and ϕ). Solutions to the time-independent wave function are written as a product of three functions:

Equation:

$$\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi),$$

where R is the radial function dependent on the radial coordinate r only; Θ is the polar function dependent on the polar coordinate θ only; and Φ is the phi function of ϕ only. Valid solutions to Schrödinger's equation $\psi(r, \theta, \phi)$ are labeled by the quantum numbers n , l , and m .

Equation:

- n : principal quantum number
- l : angular momentum quantum number
- m : angular momentum projection quantum number

(The reasons for these names will be explained in the next section.) The radial function R depends only on n and l ; the polar function Θ depends only on l and m ; and the phi function Φ depends only on m . The dependence of each function on quantum numbers is indicated with subscripts:

Equation:

$$\psi_{nlm}(r, \theta, \phi) = R_{nl}(r)\Theta_{lm}(\theta)\Phi_m(\phi).$$

Not all sets of quantum numbers (n , l , m) are possible. For example, the orbital angular quantum number l can never be greater or equal to the principal quantum number n ($l < n$). Specifically, we have

Equation:

$$\begin{aligned}
n &= 1, 2, 3, \dots \\
l &= 0, 1, 2, \dots, (n-1) \\
m &= -l, (-l+1), \dots, 0, \dots, (+l-1), +l
\end{aligned}$$

Notice that for the ground state, $n = 1$, $l = 0$, and $m = 0$. In other words, there is only one quantum state with the wave function for $n = 1$, and it is ψ_{100} . However, for $n = 2$, we have

Equation:

$$\begin{aligned}
l &= 0, \quad m = 0 \\
l &= 1, \quad m = -1, 0, 1.
\end{aligned}$$

Therefore, the allowed states for the $n = 2$ state are ψ_{200} , ψ_{21-1} , ψ_{210} , and ψ_{211} . Example wave functions for the hydrogen atom are given in [\[link\]](#). Note that some of these expressions contain the letter i , which represents $\sqrt{-1}$. When probabilities are calculated, these complex numbers do not appear in the final answer.

$n = 1, l = 0, m_l = 0$	$\psi_{100} = \frac{1}{\sqrt{\pi}} \frac{1}{a_0^{3/2}} e^{-r/a_0}$
$n = 2, l = 0, m_l = 0$	$\psi_{200} = \frac{1}{4\sqrt{2\pi}} \frac{1}{a_0^{3/2}} \left(2 - \frac{r}{a_0}\right) e^{-r/2a_0}$
$n = 2, l = 1, m_l = -1$	$\psi_{21-1} = \frac{1}{8\sqrt{\pi}} \frac{1}{a_0^{3/2}} \frac{r}{a_0} e^{-r/2a_0} \sin \theta e^{-i\phi}$
$n = 2, l = 1, m_l = 0$	$\psi_{210} = \frac{1}{4\sqrt{2\pi}} \frac{1}{a_0^{3/2}} \frac{r}{a_0} e^{-r/2a_0} \cos \theta$
$n = 2, l = 1, m_l = 1$	

$$\psi_{211} = \frac{1}{8\sqrt{\pi}} \frac{1}{a_0^{3/2}} \frac{r}{a_0} e^{-r/2a_0} \sin \theta e^{i\phi}$$

Wave Functions of the Hydrogen Atom

Physical Significance of the Quantum Numbers

Each of the three quantum numbers of the hydrogen atom (n, l, m) is associated with a different physical quantity. The **principal quantum number** n is associated with the total energy of the electron, E_n . According to Schrödinger's equation:

Equation:

$$E_n = - \left(\frac{m_e k^2 e^4}{2\hbar^2} \right) \left(\frac{1}{n^2} \right) = -E_0 \left(\frac{1}{n^2} \right),$$

where $E_0 = -13.6$ eV. Notice that this expression is identical to that of Bohr's model. As in the Bohr model, the electron in a particular state of energy does not radiate.

Example:

How Many Possible States?

For the hydrogen atom, how many possible quantum states correspond to the principal number $n = 3$? What are the energies of these states?

Strategy

For a hydrogen atom of a given energy, the number of allowed states depends on its orbital angular momentum. We can count these states for each value of the principal quantum number, $n = 1, 2, 3$. However, the total energy depends on the principal quantum number only, which means that we can use [\[link\]](#) and the number of states counted.

Solution

If $n = 3$, the allowed values of l are 0, 1, and 2. If $l = 0$, $m = 0$ (1 state). If $l = 1$, $m = -1, 0, +1$ (3 states); and if $l = 2$, $m = -2, -1, 0, +1, +2$ (5 states). In total, there are $1 + 3 + 5 = 9$ allowed states. Because the total

energy depends only on the principal quantum number, $n = 3$, the energy of each of these states is

Equation:

$$E_{n3} = -E_0 \left(\frac{1}{n^2} \right) = \frac{-13.6 \text{ eV}}{9} = -1.51 \text{ eV}.$$

Significance

An electron in a hydrogen atom can occupy many different angular momentum states with the very same energy. As the orbital angular momentum increases, the number of the allowed states with the same energy increases.

The **angular momentum orbital quantum number** l is associated with the orbital angular momentum of the electron in a hydrogen atom. Quantum theory tells us that when the hydrogen atom is in the state ψ_{nlm} , the magnitude of its orbital angular momentum is

Note:

Equation:

$$L = \sqrt{l(l+1)}\hbar,$$

where

Equation:

$$l = 0, 1, 2, \dots, (n - 1).$$

This result is slightly different from that found with Bohr's theory, which quantizes angular momentum according to the rule

$L = n$, where $n = 1, 2, 3, \dots$

Quantum states with different values of orbital angular momentum are distinguished using spectroscopic notation ([\[link\]](#)). The designations s , p , d , and f result from early historical attempts to classify atomic spectral lines. (The letters stand for sharp, principal, diffuse, and fundamental, respectively.) After f , the letters continue alphabetically.

The ground state of hydrogen is designated as the $1s$ state, where “1” indicates the energy level ($n = 1$) and “s” indicates the orbital angular momentum state ($l = 0$). When $n = 2$, l can be either 0 or 1. The $n = 2$, $l = 0$ state is designated “ $2s$.” The $n = 2$, $l = 1$ state is designated “ $2p$.” When $n = 3$, l can be 0, 1, or 2, and the states are $3s$, $3p$, and $3d$, respectively. Notation for other quantum states is given in [\[link\]](#).

The **angular momentum projection quantum number** m is associated with the azimuthal angle ϕ (see [\[link\]](#)) and is related to the z -component of orbital angular momentum of an electron in a hydrogen atom. This component is given by

Note:

Equation:

$$L_z = m\hbar,$$

where

Equation:

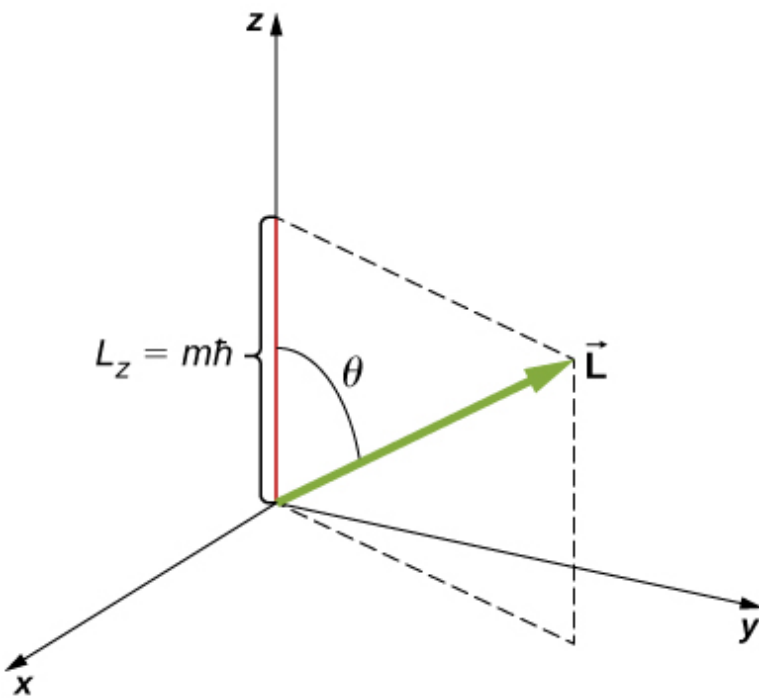
$$m = -l, -l + 1, \dots, 0, \dots, +l - 1, l.$$

The z -component of angular momentum is related to the magnitude of angular momentum by

Equation:

$$L_z = L \cos \theta,$$

where θ is the angle between the angular momentum vector and the z-axis. Note that the direction of the z-axis is determined by experiment—that is, along any direction, the experimenter decides to measure the angular momentum. For example, the z-direction might correspond to the direction of an external magnetic field. The relationship between L_z and L is given in [\[link\]](#).



The z-component of angular momentum is quantized with its own quantum number m .

Orbital Quantum Number l	Angular Momentum	State	Spectroscopic Name
0	0	s	Sharp
1	$\sqrt{2}h$	p	Principal
2	$\sqrt{6}h$	d	Diffuse
3	$\sqrt{12}h$	f	Fundamental
4	$\sqrt{20}h$	g	
5	$\sqrt{30}h$	h	

Spectroscopic Notation and Orbital Angular Momentum

	$l = 0$	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$
$n = 1$	1s					
$n = 2$	2s	2p				
$n = 3$	3s	3p	3d			
$n = 4$	4s	4p	4d	4f		
$n = 5$	5s	5p	5d	5f	5g	
$n = 6$	6s	6p	6d	6f	6g	6h

Spectroscopic Description of Quantum States

The quantization of L_z is equivalent to the quantization of θ . Substituting $\sqrt{l(l+1)}\hbar$ for L and m for L_z into this equation, we find

Equation:

$$m\hbar = \sqrt{l(l+1)}\hbar \cos \theta.$$

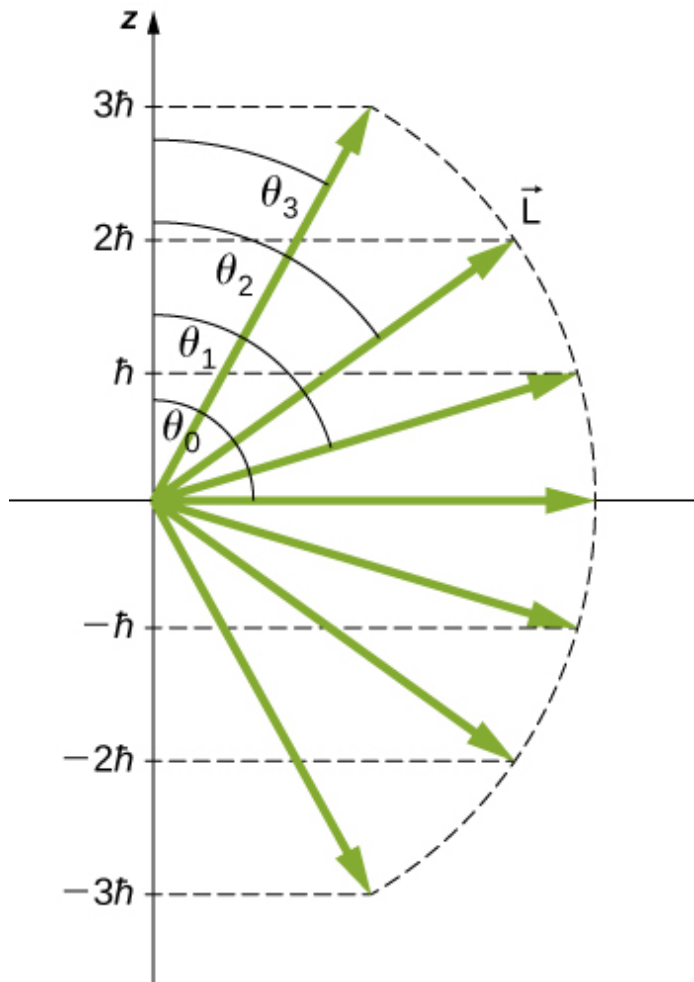
Thus, the angle θ is quantized with the particular values

Equation:

$$\theta = \cos^{-1} \left(\frac{m}{\sqrt{l(l+1)}} \right).$$

Notice that both the polar angle (θ) and the projection of the angular momentum vector onto an arbitrary z-axis (L_z) are quantized.

The quantization of the polar angle for the $l = 3$ state is shown in [\[link\]](#). The orbital angular momentum vector lies somewhere on the surface of a cone with an opening angle θ relative to the z-axis (unless $m = 0$, in which case $\theta = 90^\circ$ and the vector points are perpendicular to the z-axis).

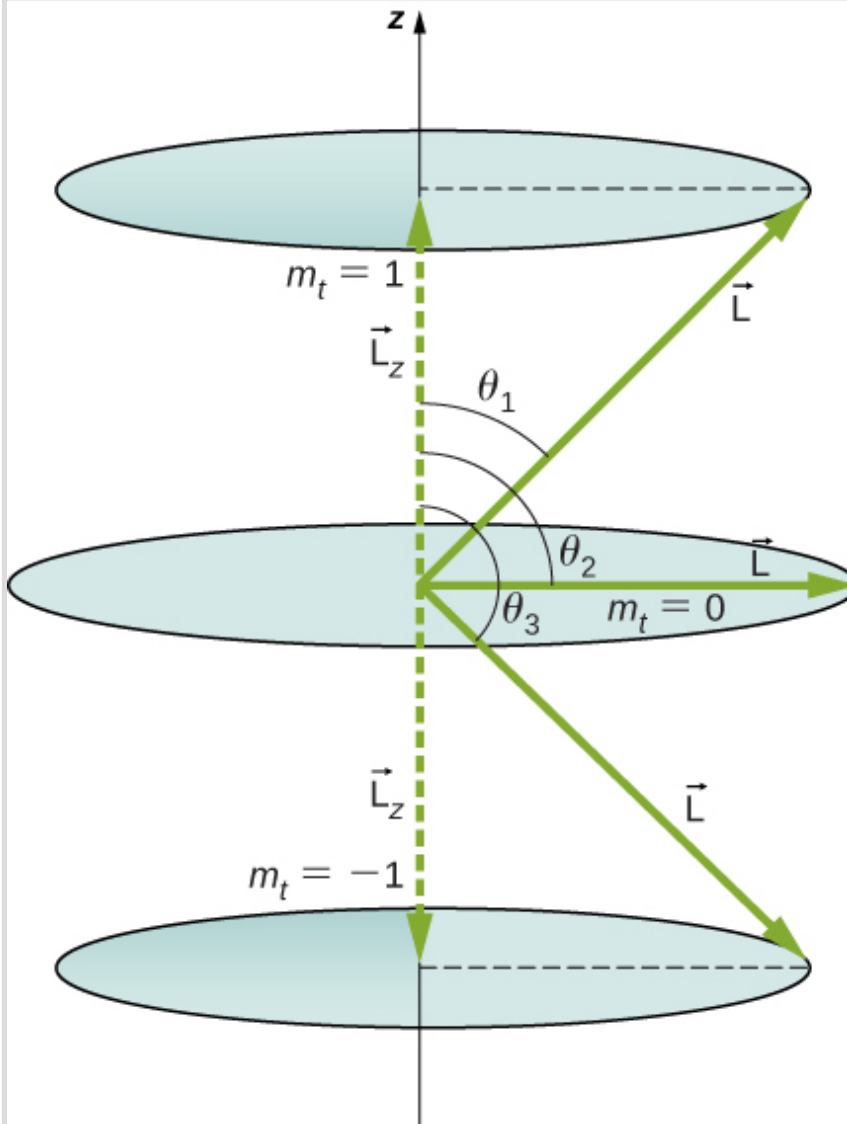


The quantization of orbital angular momentum. Each vector lies on the surface of a cone with axis along the z -axis.

A detailed study of angular momentum reveals that we cannot know all three components simultaneously. In the previous section, the z -component of orbital angular momentum has definite values that depend on the quantum number m . This implies that we cannot know both x - and y -components of angular momentum, L_x and L_y , with certainty. As a result, the precise direction of the orbital angular momentum vector is unknown.

Example:**What Are the Allowed Directions?**

Calculate the angles that the angular momentum vector \vec{L} can make with the z-axis for $l = 1$, as shown in [\[link\]](#).



The component of a given angular momentum along the z-axis (defined by the direction of a magnetic field) can have only certain values.

These are shown here for $l = 1$, for which $m = -1, 0$, and $+1$. The direction of \vec{L} is quantized in the sense that it can have only certain angles relative to the z-axis.

Strategy

The vectors $\vec{\mathbf{L}}$ and $\vec{\mathbf{L}}_z$ (in the z-direction) form a right triangle, where $\vec{\mathbf{L}}$ is the hypotenuse and $\vec{\mathbf{L}}_z$ is the adjacent side. The ratio of L_z to $|\vec{\mathbf{L}}|$ is the cosine of the angle of interest. The magnitudes $L = |\vec{\mathbf{L}}|$ and L_z are given by

Equation:

$$L = \sqrt{l(l+1)}\hbar \text{ and } L_z = m\hbar.$$

Solution

We are given $l = 1$, so m_l can be $+1$, 0 , or -1 . Thus, L has the value given by

Equation:

$$L = \sqrt{l(l+1)}\hbar = \sqrt{2}\hbar.$$

The quantity L_z can have three values, given by $L_z = m_l\hbar$.

Equation:

$$L_z = m_l\hbar = \begin{cases} \hbar, & m_l = +1 \\ 0, & m_l = 0 \\ -\hbar, & m_l = -1 \end{cases}$$

As you can see in [\[link\]](#), $\cos \theta = L_z/L$, so for $m = +1$, we have

Equation:

$$\cos \theta_1 = \frac{L_z}{L} = \frac{\hbar}{\sqrt{2}\hbar} = \frac{1}{\sqrt{2}} = 0.707.$$

Thus,

Equation:

$$\theta_1 = \cos^{-1}0.707 = 45.0^\circ.$$

Similarly, for $m = 0$, we find $\cos \theta_2 = 0$; this gives

Equation:

$$\theta_2 = \cos^{-1} 0 = 90.0^\circ.$$

Then for $m_l = -1$:

Equation:

$$\cos \theta_3 = \frac{L_z}{L} = \frac{-\hbar}{\sqrt{2}\hbar} = -\frac{1}{\sqrt{2}} = -0.707,$$

so that

Equation:

$$\theta_3 = \cos^{-1}(-0.707) = 135.0^\circ.$$

Significance

The angles are consistent with the figure. Only the angle relative to the z -axis is quantized. L can point in any direction as long as it makes the proper angle with the z -axis. Thus, the angular momentum vectors lie on cones, as illustrated. To see how the correspondence principle holds here, consider that the smallest angle (θ_1 in the example) is for the maximum value of m_l , namely $m_l = l$. For that smallest angle,

Equation:

$$\cos \theta = \frac{L_z}{L} = \frac{l}{\sqrt{l(l+1)}},$$

which approaches 1 as l becomes very large. If $\cos \theta = 1$, then $\theta = 0^\circ$. Furthermore, for large l , there are many values of m_l , so that all angles become possible as l gets very large.

Note:

Exercise:

Problem:

Check Your Understanding Can the magnitude of L_z ever be equal to L ?

Solution:

No. The quantum number $m = -l, -l + 1, \dots, 0, \dots, l - 1, l$. Thus, the magnitude of L_z is always less than L because $< \sqrt{l(l+1)}$

Using the Wave Function to Make Predictions

As we saw earlier, we can use quantum mechanics to make predictions about physical events by the use of probability statements. It is therefore proper to state, “An electron is located within this volume with this probability at this time,” but not, “An electron is located at the position (x, y, z) at this time.” To determine the probability of finding an electron in a hydrogen atom in a particular region of space, it is necessary to integrate the probability density $|\psi_{nlm}|^2$ over that region:

Equation:

$$\text{Probability} = \int_{\text{volume}} |\psi_{nlm}|^2 dV,$$

where dV is an infinitesimal volume element. If this integral is computed for all space, the result is 1, because the probability of the particle to be located *somewhere* is 100% (the normalization condition). In a more advanced course on modern physics, you will find that $|\psi_{nlm}|^2 = \psi_{nlm}^* \psi_{nlm}$, where ψ_{nlm}^* is the complex conjugate. This eliminates the occurrences of $i = \sqrt{-1}$ in the above calculation.

Consider an electron in a state of zero angular momentum ($l = 0$). In this case, the electron’s wave function depends only on the radial coordinate r .

(Refer to the states ψ_{100} and ψ_{200} in [\[link\]](#).) The infinitesimal volume element corresponds to a spherical shell of radius r and infinitesimal thickness dr , written as

Equation:

$$dV = 4\pi r^2 dr.$$

The probability of finding the electron in the region r to $r + dr$ (“at approximately r ”) is

Note:

Equation:

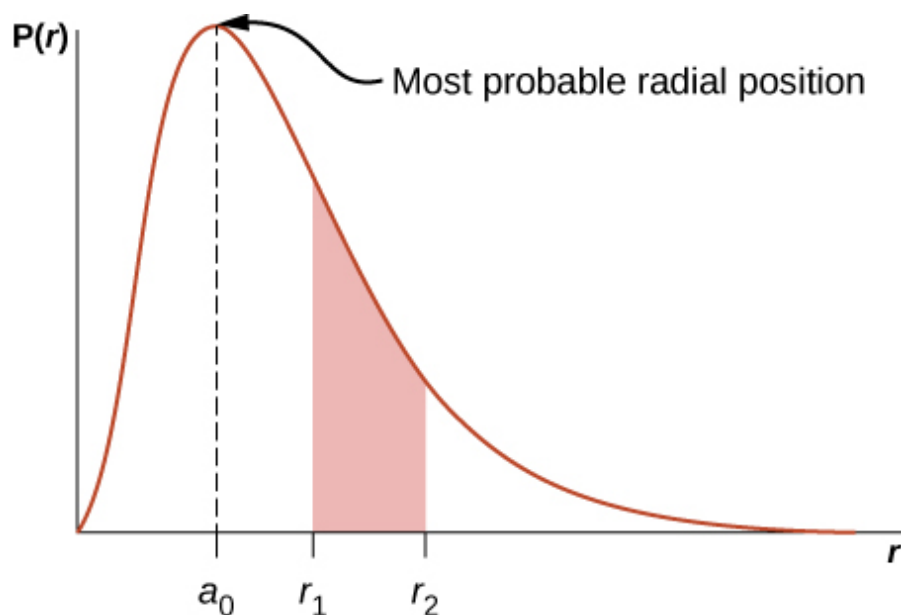
$$P(r)dr = |\psi_{n00}|^2 4\pi r^2 dr.$$

Here $P(r)$ is called the **radial probability density function** (a probability per unit length). For an electron in the ground state of hydrogen, the probability of finding an electron in the region r to $r + dr$ is

Equation:

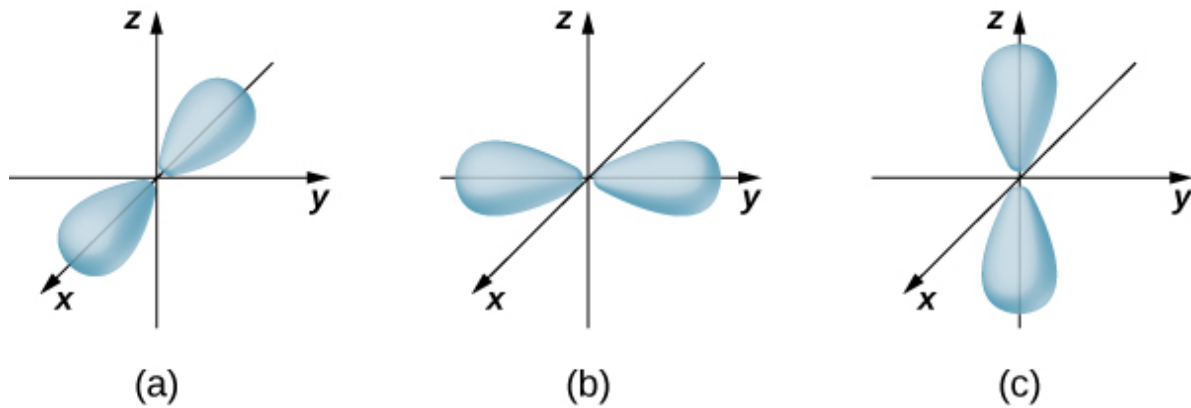
$$|\psi_{n00}|^2 4\pi r^2 dr = (4/a_0^3) r^2 \exp(-2r/a_0) dr,$$

where $a_0 = 0.5$ angstroms. The radial probability density function $P(r)$ is plotted in [\[link\]](#). The area under the curve between any two radial positions, say r_1 and r_2 , gives the probability of finding the electron in that radial range. To find the most probable radial position, we set the first derivative of this function to zero ($dP/dr = 0$) and solve for r . The most probable radial position is not equal to the average or expectation value of the radial position because $|\psi_{n00}|^2$ is not symmetrical about its peak value.



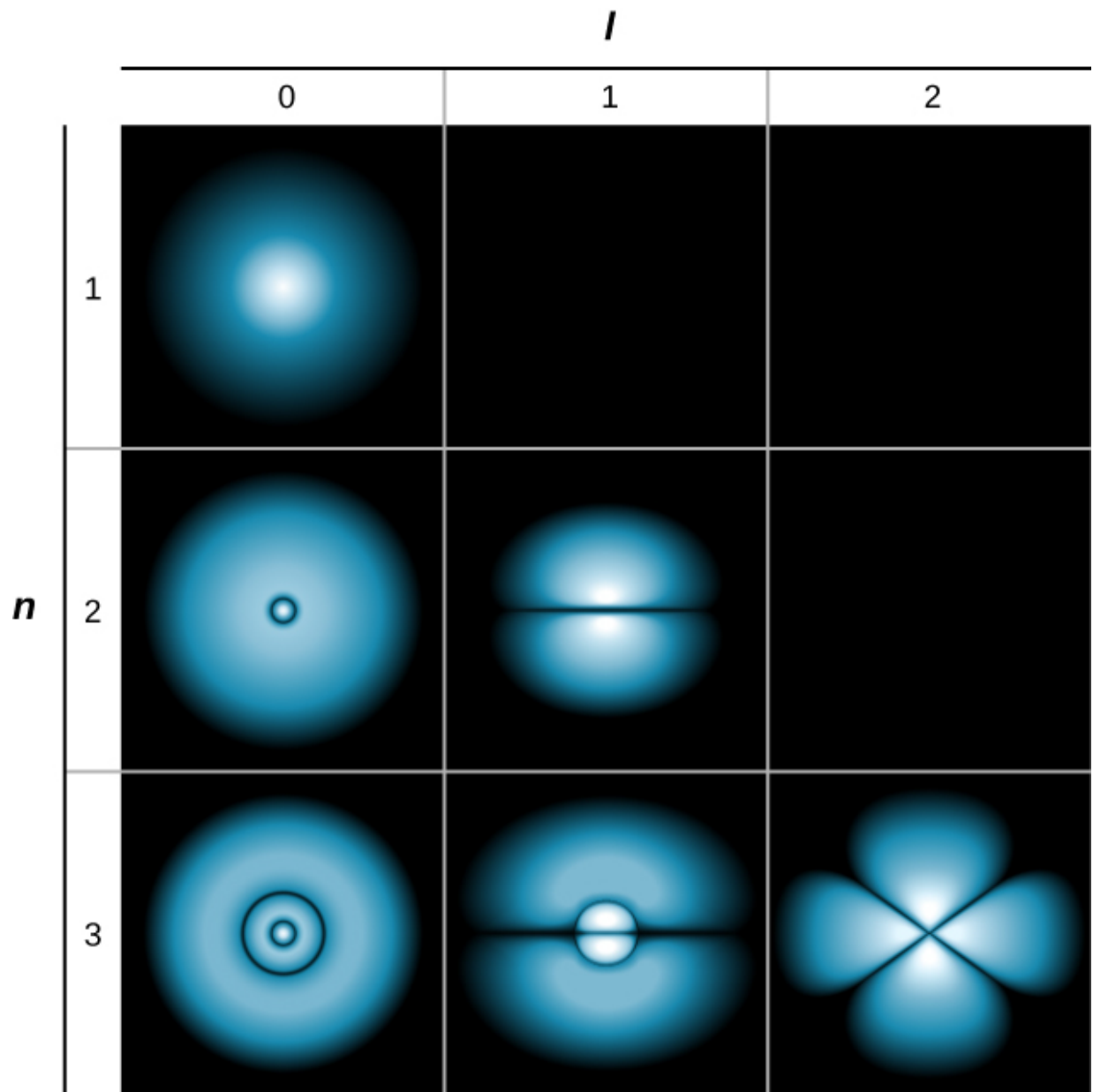
The radial probability density function for the ground state of hydrogen.

If the electron has orbital angular momentum ($l \neq 0$), then the wave functions representing the electron depend on the angles θ and ϕ ; that is, $\psi_{nlm} = \psi_{nlm}(r, \theta, \phi)$. Atomic orbitals for three states with $n = 2$ and $l = 1$ are shown in [\[link\]](#). An **atomic orbital** is a region in space that encloses a certain percentage (usually 90%) of the electron probability. (Sometimes atomic orbitals are referred to as “clouds” of probability.) Notice that these distributions are pronounced in certain directions. This directionality is important to chemists when they analyze how atoms are bound together to form molecules.



The probability density distributions for three states with $n = 2$ and $l = 1$. The distributions are directed along the (a) x -axis, (b) y -axis, and (c) z -axis.

A slightly different representation of the wave function is given in [\[link\]](#). In this case, light and dark regions indicate locations of relatively high and low probability, respectively. In contrast to the Bohr model of the hydrogen atom, the electron does not move around the proton nucleus in a well-defined path. Indeed, the uncertainty principle makes it impossible to know how the electron gets from one place to another.



Probability clouds for the electron in the ground state and several excited states of hydrogen. The probability of finding the electron is indicated by the shade of color; the lighter the coloring, the greater the chance of finding the electron.

Summary

- A hydrogen atom can be described in terms of its wave function, probability density, total energy, and orbital angular momentum.
- The state of an electron in a hydrogen atom is specified by its quantum numbers (n, l, m).
- In contrast to the Bohr model of the atom, the Schrödinger model makes predictions based on probability statements.
- The quantum numbers of a hydrogen atom can be used to calculate important information about the atom.

Conceptual Questions

Exercise:

Problem:

Identify the physical significance of each of the quantum numbers of the hydrogen atom.

Solution:

n (principal quantum number) \rightarrow total energy
 l (orbital angular quantum number) \rightarrow total absolute magnitude of the orbital angular momentum
 m (orbital angular projection quantum number) \rightarrow z-component of the orbital angular momentum

Exercise:

Problem:

Describe the ground state of hydrogen in terms of wave function, probability density, and atomic orbitals.

Exercise:

Problem:

Distinguish between Bohr's and Schrödinger's model of the hydrogen atom. In particular, compare the energy and orbital angular momentum of the ground states.

Solution:

The Bohr model describes the electron as a particle that moves around the proton in well-defined orbits. Schrödinger's model describes the electron as a wave, and knowledge about the position of the electron is restricted to probability statements. The total energy of the electron in the ground state (and all excited states) is the same for both models. However, the orbital angular momentum of the ground state is different for these models. In Bohr's model, L (ground state) = 1, and in Schrödinger's model, L (ground state) = 0.

Problems

Exercise:

Problem:

The wave function is evaluated at rectangular coordinates $(x, y, z) = (2, 1, 1)$ in arbitrary units. What are the spherical coordinates of this position?

Solution:

$$(r, \theta, \phi) = (\sqrt{6}, 66^\circ, 27^\circ).$$

Exercise:

Problem:

If an atom has an electron in the $n = 5$ state with $m = 3$, what are the possible values of l ?

Exercise:

Problem:

What are the possible values of m for an electron in the $n = 4$ state?

Solution:

$\pm 3, \pm 2, \pm 1, 0$ are possible

Exercise:**Problem:**

What, if any, constraints does a value of $m = 1$ place on the other quantum numbers for an electron in an atom?

Exercise:

Problem: How many possible states are there for the $l = 4$ state?

Solution:

18

Exercise:**Problem:**

- (a) How many angles can L make with the z -axis for an $l = 2$ electron?
- (b) Calculate the value of the smallest angle.

Exercise:**Problem:**

The force on an electron is “negative the gradient of the potential energy function.” Use this knowledge and [\[link\]](#) to show that the force on the electron in a hydrogen atom is given by Coulomb’s force law.

Solution:

$$F = -k \frac{Qq}{r^2}$$

Exercise:**Problem:**

What is the total number of states with orbital angular momentum $l = 0$? (Ignore electron spin.)

Exercise:

Problem:

The wave function is evaluated at spherical coordinates $(r, \theta, \phi) = (\sqrt{3}, 45^\circ, 45^\circ)$, where the value of the radial coordinate is given in arbitrary units. What are the rectangular coordinates of this position?

Solution:

(1, 1, 1)

Exercise:**Problem:**

Coulomb's force law states that the force between two charged particles is:

$F = k \frac{Qq}{r^2}$. Use this expression to determine the potential energy function.

Exercise:**Problem:**

Write an expression for the total number of states with orbital angular momentum l .

Solution:

For the orbital angular momentum quantum number, l , the allowed values of:

$$m = -l, -l + 1, \dots, 0, \dots, l - 1, l.$$

With the exception of $m = 0$, the total number is just $2l$ because the number of states on either side of $m = 0$ is just l . Including $m = 0$, the total number of orbital angular momentum states for the orbital angular momentum quantum number, l , is: $2l + 1$. Later, when we consider electron spin, the total number of angular momentum states will be

found to twice this value because each orbital angular momentum states is associated with two states of electron spin: spin up and spin down).

Exercise:

Problem:

Consider hydrogen in the ground state, ψ_{100} . (a) Use the derivative to determine the radial position for which the probability density, $P(r)$, is a maximum.

(b) Use the integral concept to determine the average radial position. (This is called the expectation value of the electron's radial position.) Express your answers into terms of the Bohr radius, a_0 . Hint: The expectation value is the just average value. (c) Why are these values different?

Exercise:

Problem:

What is the probability that the 1s electron of a hydrogen atom is found outside the Bohr radius?

Solution:

The probability that the 1s electron of a hydrogen atom is found outside of the Bohr radius is $\int_{a_0}^{\infty} P(r) dr \approx 0.68$

Exercise:

Problem:

How many polar angles are possible for an electron in the $l = 5$ state?

Exercise:

Problem:

What is the maximum number of orbital angular momentum electron states in the $n = 2$ shell of a hydrogen atom? (Ignore electron spin.)

Solution:

For $n = 2$, $l = 0$ (1 state), and $l = 1$ (3 states). The total is 4.

Exercise:**Problem:**

What is the maximum number of orbital angular momentum electron states in the $n = 3$ shell of a hydrogen atom? (Ignore electron spin.)

Glossary

angular momentum orbital quantum number (l)

quantum number associated with the orbital angular momentum of an electron in a hydrogen atom

angular momentum projection quantum number (m)

quantum number associated with the z-component of the orbital angular momentum of an electron in a hydrogen atom

atomic orbital

region in space that encloses a certain percentage (usually 90%) of the electron probability

principal quantum number (n)

quantum number associated with the total energy of an electron in a hydrogen atom

radial probability density function

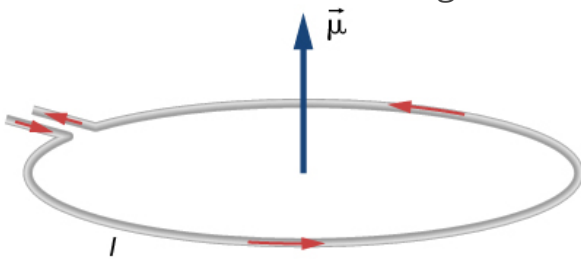
function use to determine the probability of a electron to be found in a spatial interval in r

Orbital Magnetic Dipole Moment of the Electron

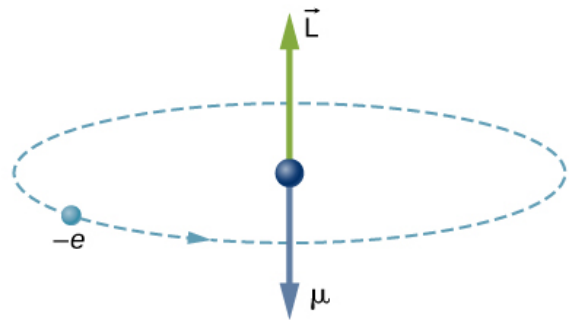
By the end of this section, you will be able to:

- Explain why the hydrogen atom has magnetic properties
- Explain why the energy levels of a hydrogen atom associated with orbital angular momentum are split by an external magnetic field
- Use quantum numbers to calculate the magnitude and direction of the orbital magnetic dipole moment of a hydrogen atom

In Bohr's model of the hydrogen atom, the electron moves in a circular orbit around the proton. The electron passes by a particular point on the loop in a certain time, so we can calculate a current $I = Q/t$. An electron that orbits a proton in a hydrogen atom is therefore analogous to current flowing through a circular wire ([link](#)). In the study of magnetism, we saw that a current-carrying wire produces magnetic fields. It is therefore reasonable to conclude that the hydrogen atom produces a magnetic field and interacts with other magnetic fields.



(a) Current-carrying loop



(b) Hydrogen atom

(a) Current flowing through a circular wire is analogous to (b) an electron that orbits a proton in a hydrogen atom.

The **orbital magnetic dipole moment** is a measure of the strength of the magnetic field produced by the orbital angular momentum of an electron. From [Force and Torque on a Current Loop](#), the magnitude of the orbital magnetic dipole moment for a current loop is

Equation:

$$\mu = IA,$$

where I is the current and A is the area of the loop. (For brevity, we refer to this as the magnetic moment.) The current I associated with an electron in orbit about a proton in a hydrogen atom is

Equation:

$$I = \frac{e}{T},$$

where e is the magnitude of the electron charge and T is its orbital period. If we assume that the electron travels in a perfectly circular orbit, the orbital period is

Equation:

$$T = \frac{2\pi r}{v},$$

where r is the radius of the orbit and v is the speed of the electron in its orbit. Given that the area of a circle is πr^2 , the absolute magnetic moment is

Equation:

$$\mu = IA = \frac{e}{\left(\frac{2\pi r}{v}\right)} \pi r^2 = \frac{evr}{2}.$$

It is helpful to express the magnetic momentum μ in terms of the orbital angular momentum ($\vec{L} = \vec{r} \times \vec{p}$). Because the electron orbits in a circle, the position vector \vec{r} and the momentum vector \vec{p} form a right angle. Thus, the magnitude of the orbital angular momentum is

Equation:

$$L = |\vec{L}| = |\vec{r} \times \vec{p}| = rp \sin \theta = rp = rmv.$$

Combining these two equations, we have

Equation:

$$\mu = \left(\frac{e}{2m_e} \right) L.$$

In full vector form, this expression is written as

Note:

Equation:

$$\vec{\mu} = - \left(\frac{e}{2m_e} \right) \vec{L}.$$

The negative sign appears because the electron has a negative charge. Notice that the direction of the magnetic moment of the electron is antiparallel to the orbital angular momentum, as shown in [\[link\]](#)(b). In the Bohr model of the atom, the relationship between $\vec{\mu}$ and \vec{L} in [\[link\]](#) is independent of the radius of the orbit.

The magnetic moment μ can also be expressed in terms of the orbital angular quantum number l . Combining [\[link\]](#) and [\[link\]](#), the magnitude of the magnetic moment is

Equation:

$$\mu = \left(\frac{e}{2m_e} \right) L = \left(\frac{e}{2m_e} \right) \sqrt{l(l+1)} \hbar = \mu_B \sqrt{l(l+1)}.$$

The z-component of the magnetic moment is

Equation:

$$\mu_z = - \left(\frac{e}{2m_e} \right) L_z = - \left(\frac{e}{2m_e} \right) m \hbar = -\mu_B m.$$

The quantity μ_B is a fundamental unit of magnetism called the **Bohr magneton**, which has the value 9.3×10^{-24} joule/tesla (J/T) or 5.8×10^{-5} eV/T. Quantization of the magnetic moment is the result of quantization of the orbital angular momentum.

As we will see in the next section, the total magnetic dipole moment of the hydrogen atom is due to both the orbital motion of the electron and its intrinsic spin. For now, we ignore the effect of electron spin.

Example:

Orbital Magnetic Dipole Moment

What is the magnitude of the orbital dipole magnetic moment μ of an electron in the hydrogen atom in the (a) *s* state, (b) *p* state, and (c) *d* state? (Assume that the spin of the electron is zero.)

Strategy

The magnetic momentum of the electron is related to its orbital angular momentum L . For the hydrogen atom, this quantity is related to the orbital angular quantum number l . The states are given in spectroscopic notation, which relates a letter (*s*, *p*, *d*, etc.) to a quantum number.

Solution

The magnitude of the magnetic moment is given in [\[link\]](#):

Equation:

$$\mu = \left(\frac{e}{2m_e} \right) L = \left(\frac{e}{2m_e} \right) \sqrt{l(l+1)} \hbar = \mu_B \sqrt{l(l+1)}.$$

- a. For the *s* state, $l = 0$ so we have $\mu = 0$ and $\mu_z = 0$.
- b. For the *p* state, $l = 1$ and we have

Equation:

$$\begin{aligned}\mu &= \mu_B \sqrt{1(1+1)} = \sqrt{2}\mu_B \\ \mu_z &= -\mu_B m, \text{ where } m = (-1, 0, 1), \text{ so} \\ \mu_z &= \mu_B, 0, -\mu_B.\end{aligned}$$

c. For the d state, $l = 2$ and we obtain

Equation:

$$\begin{aligned}\mu &= \mu_B \sqrt{2(2+1)} = \sqrt{6}\mu_B \\ \mu_z &= -\mu_B m, \text{ where } m = (-2, -1, 0, 1, 2), \text{ so} \\ \mu_z &= 2\mu_B, \mu_B, 0, -\mu_B, -2\mu_B.\end{aligned}$$

Significance

In the s state, there is no orbital angular momentum and therefore no magnetic moment. This does not mean that the electron is at rest, just that the overall motion of the electron does not produce a magnetic field. In the p state, the electron has a magnetic moment with three possible values for the z -component of this magnetic moment; this means that magnetic moment can point in three different polar directions—each antiparallel to the orbital angular momentum vector. In the d state, the electron has a magnetic moment with five possible values for the z -component of this magnetic moment. In this case, the magnetic moment can point in five different polar directions.

A hydrogen atom has a magnetic field, so we expect the hydrogen atom to interact with an external magnetic field—such as the push and pull between two bar magnets. From [Force and Torque on a Current Loop](#), we know that when a current loop interacts with an external magnetic field \vec{B} , it experiences a torque given by

Equation:

$$\vec{\tau} = I (\vec{A} \times \vec{B}) = \vec{\mu} \times \vec{B},$$

where I is the current, \vec{A} is the area of the loop, $\vec{\mu}$ is the magnetic moment, and \vec{B} is the external magnetic field. This torque acts to rotate the magnetic moment vector of the hydrogen atom to align with the external magnetic field. Because mechanical work is done by the external magnetic field on the hydrogen atom, we can talk about energy transformations in the atom. The potential energy of the hydrogen atom associated with this magnetic interaction is given by [\[link\]](#):

Equation:

$$U = -\vec{\mu} \cdot \vec{B}.$$

If the magnetic moment is antiparallel to the external magnetic field, the potential energy is large, but if the magnetic moment is parallel to the field, the potential energy is small. Work done on the hydrogen atom to rotate the atom's magnetic moment vector in the direction of the external magnetic field is therefore associated with a drop in potential energy. The energy of the system is conserved, however, because a drop in potential energy produces radiation (the emission of a photon). These energy transitions are quantized because the magnetic moment can point in only certain directions.

If the external magnetic field points in the positive z -direction, the potential energy associated with the orbital magnetic dipole moment is

Note:

Equation:

$$U(\theta) = -\mu B \cos \theta = -\mu_z B = -(-\mu_B m)B = m\mu_B B,$$

where μ_B is the Bohr magneton and m is the angular momentum projection quantum number (or **magnetic orbital quantum number**), which has the

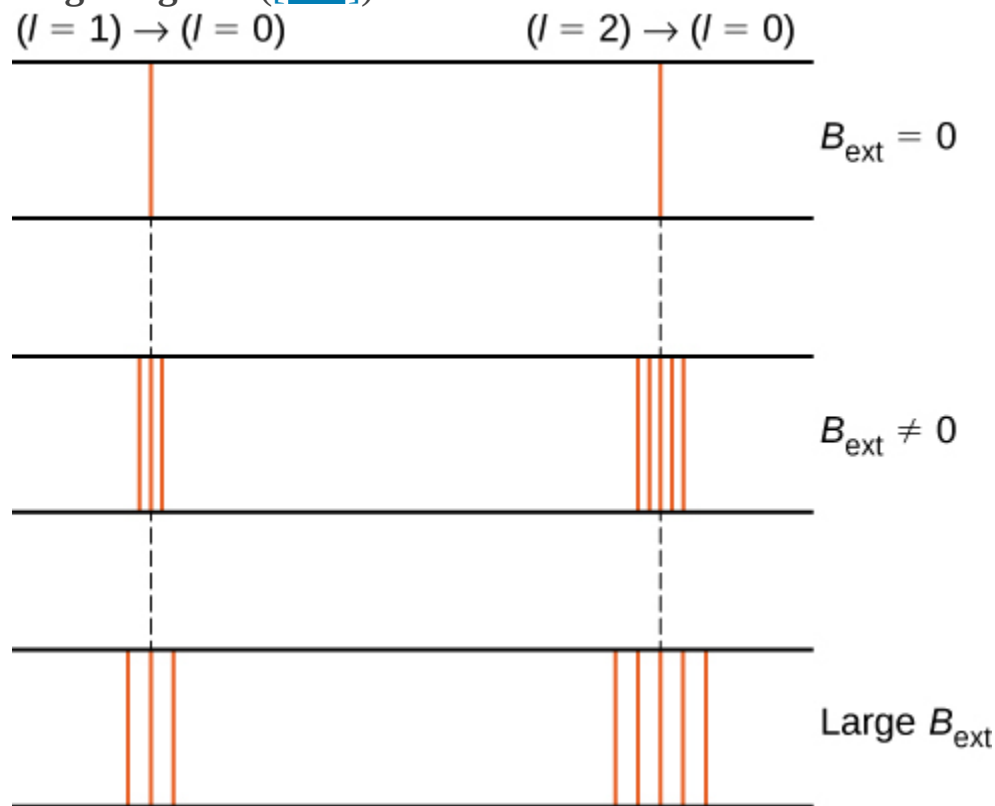
values

Equation:

$$m = -l, -l + 1, \dots, 0, \dots, l - 1, l.$$

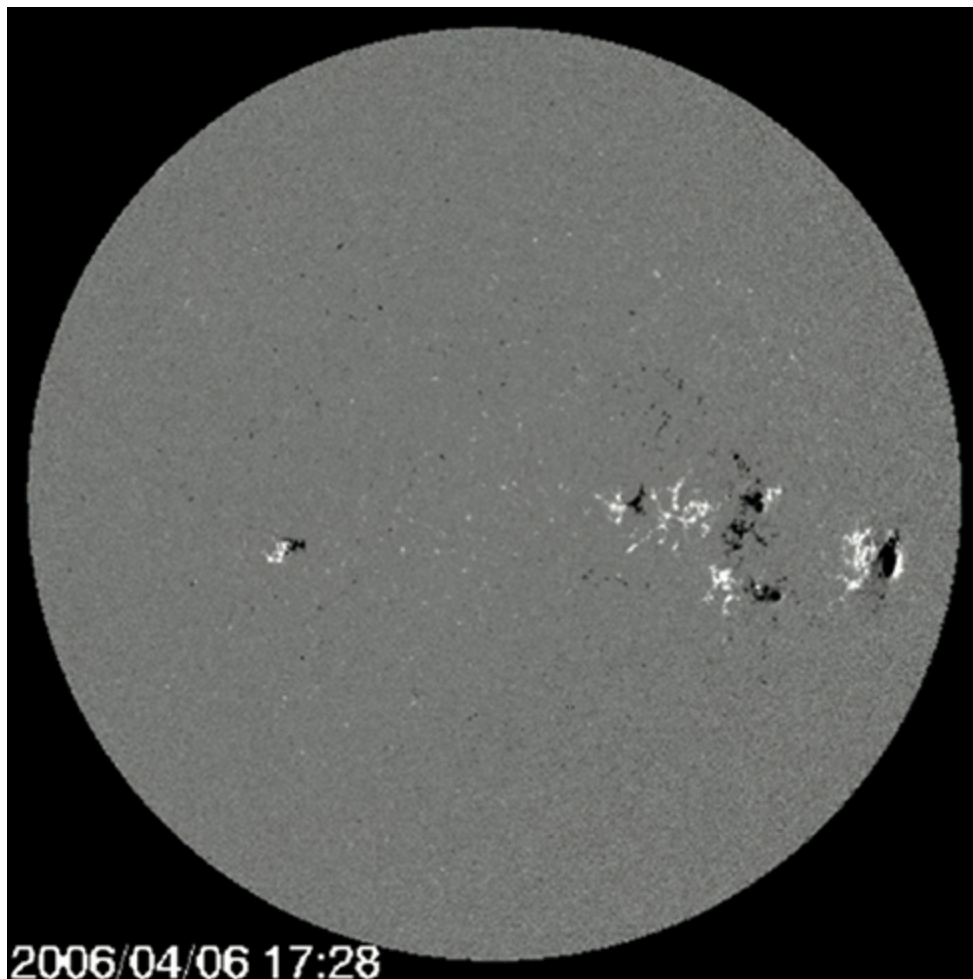
For example, in the $l = 1$ electron state, the total energy of the electron is split into three distinct energy levels corresponding to $U = -\mu_B B, 0, \mu_B B$.

The splitting of energy levels by an external magnetic field is called the **Zeeman effect**. Ignoring the effects of electron spin, transitions from the $l = 1$ state to a common lower energy state produce three closely spaced spectral lines ([link](#), left column). Likewise, transitions from the $l = 2$ state produce five closely spaced spectral lines (right column). The separation of these lines is proportional to the strength of the external magnetic field. This effect has many applications. For example, the splitting of lines in the hydrogen spectrum of the Sun is used to determine the strength of the Sun's magnetic field. Many such magnetic field measurements can be used to make a map of the magnetic activity at the Sun's surface called a **magnetogram** ([link](#)).



The Zeeman effect refers to the splitting of spectral lines by an external magnetic field. In the left column, the energy splitting occurs due to transitions from the state ($n = 2, l = 1$) to a lower energy state; and in the right column, energy splitting occurs due to transitions from the state ($n = 2, l = 2$) to a lower-energy state.

The separation of these lines is proportional to the strength of the external magnetic field.



A magnetogram of the Sun. The bright and dark spots show significant magnetic activity at the surface of the Sun. (credit: NASA, SDO)

Summary

- A hydrogen atom has magnetic properties because the motion of the electron acts as a current loop.
- The energy levels of a hydrogen atom associated with orbital angular momentum are split by an external magnetic field because the orbital angular magnetic moment interacts with the field.
- The quantum numbers of an electron in a hydrogen atom can be used to calculate the magnitude and direction of the orbital magnetic dipole moment of the atom.

Conceptual Questions

Exercise:

Problem:

Explain why spectral lines of the hydrogen atom are split by an external magnetic field. What determines the number and spacing of these lines?

Exercise:

Problem:

A hydrogen atom is placed in a magnetic field. Which of the following quantities are affected? (a) total energy; (b) angular momentum; (c) z-component of angular momentum; (d) polar angle.

Solution:

a, c, d; The total energy is changed (Zeeman splitting). The work done on the hydrogen atom rotates the atom, so the z-component of angular momentum and polar angle are affected. However, the angular momentum is not affected.

Exercise:

Problem:

On what factors does the orbital magnetic dipole moment of an electron depend?

Problems**Exercise:****Problem:**

Find the magnitude of the orbital magnetic dipole moment of the electron in the $3p$ state. (Express your answer in terms of μ_B .)

Solution:

The $3p$ state corresponds to $n = 3, l = 2$. Therefore, $\mu = \mu_B \sqrt{6}$

Exercise:**Problem:**

A current of $I = 2\text{A}$ flows through a square-shaped wire with 2-cm side lengths. What is the magnetic moment of the wire?

Exercise:**Problem:**

Estimate the ratio of the electron magnetic moment to the *muon* magnetic moment for the same state of orbital angular momentum. (Hint: $m_\mu = 105.7 \text{ MeV}/c^2$)

Solution:

The ratio of their masses is $1/207$, so the ratio of their magnetic moments is 207. The electron's magnetic moment is more than 200 times larger than the muon.

Exercise:

Problem:

Find the magnitude of the orbital magnetic dipole moment of the electron in the $4d$ state. (Express your answer in terms of μ_B .)

Exercise:**Problem:**

For a $3d$ electron in an external magnetic field of $2.50 \times 10^{-3} \text{ T}$, find (a) the current associated with the orbital angular momentum, and (b) the maximum torque.

Solution:

a. The $3d$ state corresponds to $n = 3$, $l = 2$. So,

$$I = 4.43 \times 10^{-7} \text{ A}.$$

b. The maximum torque occurs when the magnetic moment and external magnetic field vectors are at right angles ($\sin \theta = 1$). In this case:

$$|\vec{\tau}| = \mu B.$$

$$\tau = 5.70 \times 10^{-26} \text{ N} \cdot \text{m}.$$

Exercise:**Problem:**

An electron in a hydrogen atom is in the $n = 5$, $l = 4$ state. Find the smallest angle the magnetic moment makes with the z -axis. (Express your answer in terms of μ_B .)

Exercise:**Problem:**

Find the minimum torque magnitude $|\vec{\tau}|$ that acts on the orbital magnetic dipole of a $3p$ electron in an external magnetic field of $2.50 \times 10^{-3} \text{ T}$.

Solution:

A $3p$ electron is in the state $n = 3$ and $l = 1$. The minimum torque magnitude occurs when the magnetic moment and external magnetic field vectors are most parallel (antiparallel). This occurs when

$m = \pm 1$. The torque magnitude is given by

$$|\vec{\tau}| = \mu B \sin \theta,$$

Where

$$\mu = (1.31 \times 10^{-24} \text{ J/T}).$$

For $m = \pm 1$, we have:

$$|\vec{\tau}| = 2.32 \times 10^{-21} \text{ N} \cdot \text{m}.$$

Exercise:

Problem:

An electron in a hydrogen atom is in $3p$ state. Find the smallest angle the magnetic moment makes with the z -axis. (Express your answer in terms of μ_B .)

Exercise:

Problem: Show that $U = -\vec{\mu} \cdot \vec{B}$.

(Hint: An infinitesimal amount of work is done to align the magnetic moment with the external field. This work rotates the magnetic moment vector through an angle $-d\theta$ (toward the positive z -direction), where $d\theta$ is a positive angle change.)

Solution:

An infinitesimal work dW done by a magnetic torque τ to rotate the magnetic moment through an angle $-d\theta$:

$$dW = \tau (-d\theta),$$

where $\tau = |\vec{\mu} \times \vec{B}|$. Work done is interpreted as a drop in potential energy U , so

$$dW = -dU.$$

The total energy change is determined by summing over infinitesimal changes in the potential energy:

$$U = -\mu B \cos \theta$$

$$U = -\vec{\mu} \cdot \vec{B}.$$

Glossary

Bohr magneton

magnetic moment of an electron, equal to $9.3 \times 10^{-24} \text{ J/T}$ or $5.8 \times 10^{-5} \text{ eV/T}$

magnetic orbital quantum number

another term for the angular momentum projection quantum number

magnetogram

pictorial representation, or map, of the magnetic activity at the Sun's surface

orbital magnetic dipole moment

measure of the strength of the magnetic field produced by the orbital angular momentum of the electron

Zeeman effect

splitting of energy levels by an external magnetic field

Electron Spin

By the end of this section, you will be able to:

- Express the state of an electron in a hydrogen atom in terms of five quantum numbers
- Use quantum numbers to calculate the magnitude and direction of the spin and magnetic moment of an electron
- Explain the fine and hyperfine structure of the hydrogen spectrum in terms of magnetic interactions inside the hydrogen atom

In this section, we consider the effects of electron spin. Spin introduces two additional quantum numbers to our model of the hydrogen atom. Both were discovered by looking at the fine structure of atomic spectra. Spin is a fundamental characteristic of all particles, not just electrons, and is analogous to the intrinsic spin of extended bodies about their own axes, such as the daily rotation of Earth.

Spin is quantized in the same manner as orbital angular momentum. It has been found that the magnitude of the intrinsic spin angular momentum S of an electron is given by

Note:

Equation:

$$S = \sqrt{s(s + 1)}\hbar,$$

where s is defined to be the **spin quantum number**. This is similar to the quantization of L given in [\[link\]](#), except that the only value allowed for s for an electron is $s = 1/2$. The electron is said to be a “spin-half particle.” The **spin projection quantum number** m_s is associated with the z-components of spin, expressed by

Note:

Equation:

$$S_z = m_s \hbar.$$

In general, the allowed quantum numbers are

Equation:

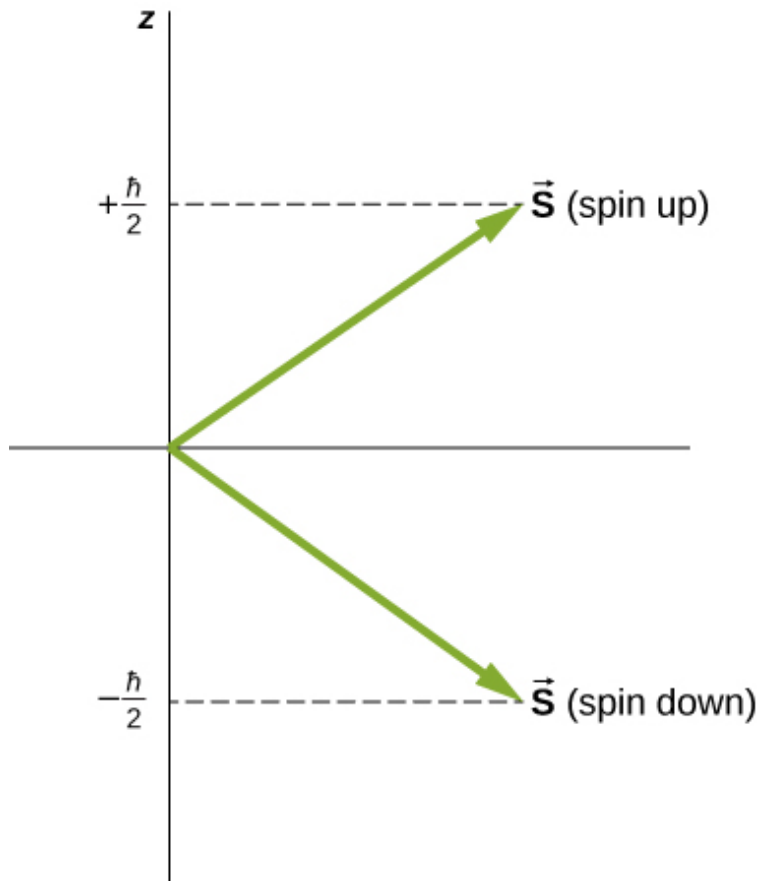
$$m_s = -s, -s + 1, \dots, 0, \dots, +s - 1, s.$$

For the special case of an electron ($s = 1/2$),

Equation:

$$m_s = -\frac{1}{2}, \frac{1}{2}.$$

Directions of intrinsic spin are quantized, just as they were for orbital angular momentum. The $m_s = -1/2$ state is called the “spin-down” state and has a z-component of spin, $s_z = -1/2$; the $m_s = +1/2$ state is called the “spin-up” state and has a z-component of spin, $s_z = +1/2$. These states are shown in [\[link\]](#).



The two possible states of electron spin.

The intrinsic magnetic dipole moment of an electron μ_e can also be expressed in terms of the spin quantum number. In analogy to the orbital angular momentum, the magnitude of the electron magnetic moment is

Equation:

$$\mu_s = \left(\frac{e}{2m_e} \right) S.$$

According to the special theory of relativity, this value is low by a factor of 2. Thus, in vector form, the spin magnetic moment is

Note:

Equation:

$$\vec{\mu} = \left(\frac{e}{m_e} \right) \vec{S}.$$

The z-component of the magnetic moment is

Equation:

$$\mu_z = - \left(\frac{e}{m_e} \right) S_z = - \left(\frac{e}{m_e} \right) m_s \hbar.$$

The spin projection quantum number has just two values ($m_s = \pm 1/2$), so the z-component of the magnetic moment also has just two values:

Equation:

$$\mu_z = \pm \left(\frac{e}{2m_e} \right) \hbar = \pm \mu_B \hbar,$$

where μ_B is one Bohr magneton. An electron is magnetic, so we expect the electron to interact with other magnetic fields. We consider two special cases: the interaction of a free electron with an external (nonuniform) magnetic field, and an electron in a hydrogen atom with a magnetic field produced by the orbital angular momentum of the electron.

Example:

Electron Spin and Radiation

A hydrogen atom in the ground state is placed in an external uniform magnetic field ($B = 1.5$ T). Determine the frequency of radiation produced in a transition between the spin-up and spin-down states of the electron.

Strategy

The spin projection quantum number is $m_s = \pm 1/2$, so the z-component of the magnetic moment is

Equation:

$$\mu_z = \pm \left(\frac{e}{2m_e} \right) = \pm \mu_B \hbar.$$

The potential energy associated with the interaction between the electron magnetic moment and the external magnetic field is

Equation:

$$U = -\mu_z B = \mp \mu_B B.$$

The frequency of light emitted is proportional to the energy (ΔE) difference between these two states.

Solution

The energy difference between these states is $\Delta E = 2\mu_B B$, so the frequency of radiation produced is

Equation:

$$f = \frac{\Delta E}{h} = \frac{2\mu_B B}{h} = \frac{2 \left(5.79 \times \frac{10^{-5} \text{ eV}}{\text{T}} \right) (1.5 \text{ T})}{4.136 \times 10^{-15} \text{ eV} \cdot \text{s}} = 4.2 \times 10^{10} \frac{\text{cycles}}{\text{s}}.$$

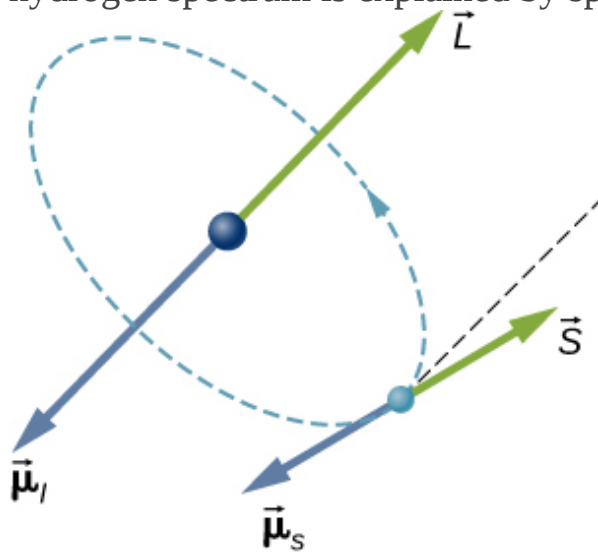
Significance

The electron magnetic moment couples with the external magnetic field. The energy of this system is different whether the electron is aligned or not with the proton. The frequency of radiation produced by a transition between these states is proportional to the energy difference. If we double the strength of the magnetic field, holding all other things constant, the frequency of the radiation doubles and its wavelength is cut in half.

In a hydrogen atom, the electron magnetic moment can interact with the magnetic field produced by the orbital angular momentum of the electron, a phenomenon called **spin-orbit coupling**. The orbital angular momentum (\vec{L}),

orbital magnetic moment ($\vec{\mu}_l$), spin angular momentum (\vec{S}), and spin magnetic moment ($\vec{\mu}_s$) vectors are shown together in [\[link\]](#).

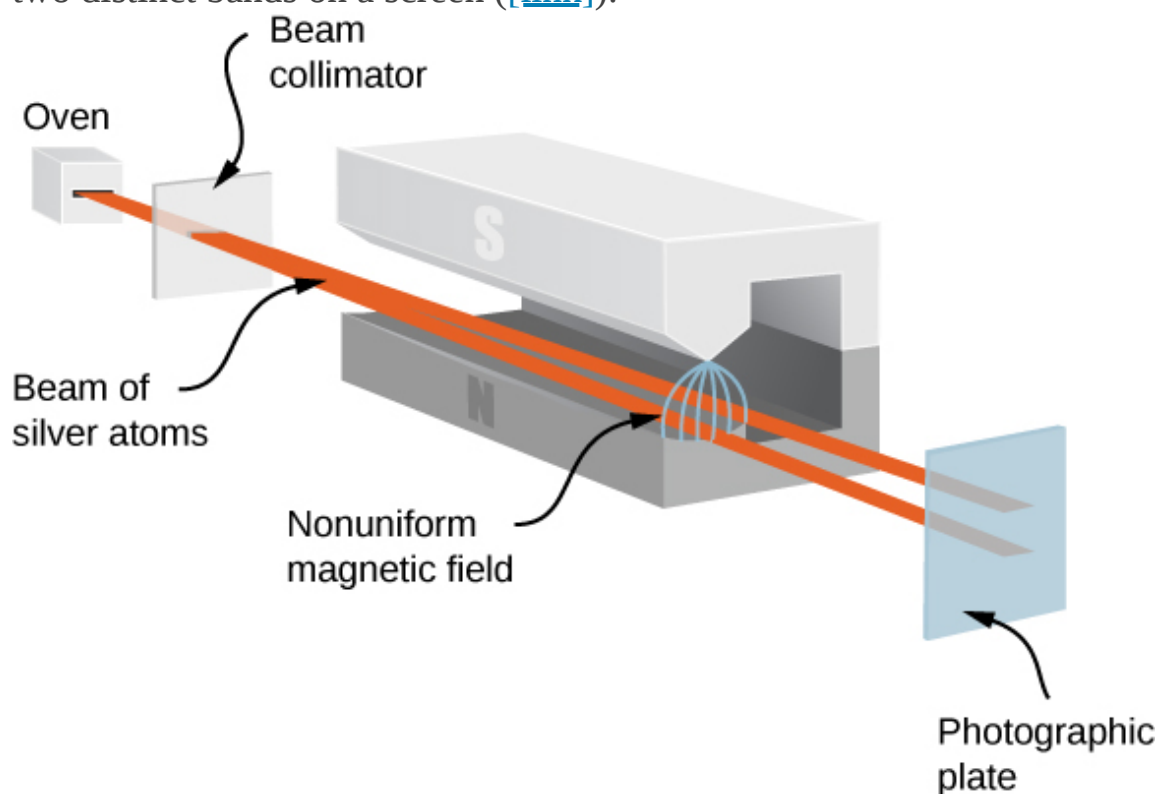
Just as the energy levels of a hydrogen atom can be split by an *external* magnetic field, so too are the energy levels of a hydrogen atom split by *internal* magnetic fields of the atom. If the magnetic moment of the electron and orbital magnetic moment of the electron are antiparallel, the potential energy from the magnetic interaction is relatively high, but when these moments are parallel, the potential energy is relatively small. Transition from each of these two states to a lower-energy level results in the emission of a photon of slightly different frequency. That is, the spin-orbit coupling “splits” the spectral line expected from a spin-less electron. The **fine structure** of the hydrogen spectrum is explained by spin-orbit coupling.



Spin-orbit coupling is the interaction of an electron's spin magnetic moment $\vec{\mu}_s$ with its orbital magnetic moment $\vec{\mu}_l$.

The Stern-Gerlach experiment provides experimental evidence that electrons have spin angular momentum. The experiment passes a stream of silver (Ag) atoms through an external, nonuniform magnetic field. The Ag atom has an orbital angular momentum of zero and contains a single unpaired electron in

the outer shell. Therefore, the total angular momentum of the Ag atom is due entirely to the spin of the outer electron ($s = 1/2$). Due to electron spin, the Ag atoms act as tiny magnets as they pass through the magnetic field. These “magnets” have two possible orientations, which correspond to the spin-up and -down states of the electron. The magnetic field diverts the spin up atoms in one direction and the spin-down atoms in another direction. This produces two distinct bands on a screen ([link](#)).



In the Stern-Gerlach experiment, an external, nonuniform magnetic field diverts a beam of electrons in two different directions. This result is due to the quantization of spin angular momentum.

According to classical predictions, the angular momentum (and, therefore, the magnetic moment) of the Ag atom can point in any direction, so one expects, instead, a continuous smudge on the screen. The resulting two bands of the Stern-Gerlach experiment provide startling support for the ideas of quantum mechanics.

Note:

Visit [PhET Explorations: Stern-Gerlach Experiment](#) to learn more about the Stern-Gerlach experiment.

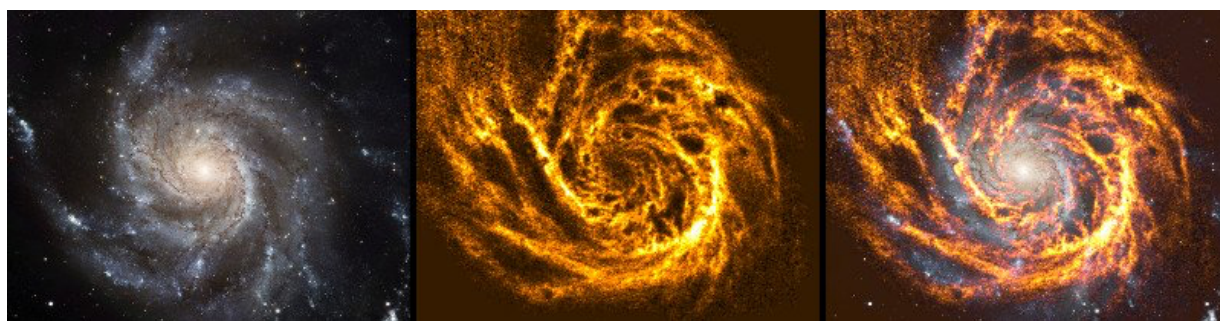
Note:**Exercise:****Problem:**

Check Your Understanding If the Stern-Gerlach experiment yielded four distinct bands instead of two, what might be concluded about the spin quantum number of the charged particle?

Solution:

$$s = 3/2 <$$

Just like an electron, a proton is spin $1/2$ and has a magnetic moment. (According to nuclear theory, this moment is due to the orbital motion of quarks within the proton.) The **hyperfine structure** of the hydrogen spectrum is explained by the interaction between the magnetic moment of the proton and the magnetic moment of the electron, an interaction known as spin-spin coupling. The energy of the electron-proton system is different depending on whether or not the moments are aligned. Transitions between these states (**spin-flip transitions**) result in the emission of a photon with a wavelength of $\lambda \approx 21 \text{ cm}$ (in the radio range). The 21-cm line in atomic spectroscopy is a “fingerprint” of hydrogen gas. Astronomers exploit this spectral line to map the spiral arms of galaxies, which are composed mostly of hydrogen ([\[link\]](#)).



(a)

(b)

(c)

The magnetic interaction between the electron and proton in the hydrogen atom is used to map the spiral arms of the Pinwheel Galaxy (NGC 5457). (a) The galaxy seen in visible light; (b) the galaxy seen in 21-cm hydrogen radiation; (c) the composite image of (a) and (b). Notice how the hydrogen emission penetrates dust in the galaxy to show the spiral arms very clearly, whereas the galactic nucleus shows up better in visible light (credit a: modification of work by ESA & NASA; credit b: modification of work by Fabian Walter).

A complete specification of the state of an electron in a hydrogen atom requires five quantum numbers: n , l , m , s , and m_s . The names, symbols, and allowed values of these quantum numbers are summarized in [\[link\]](#).

Name	Symbol	Allowed values
Principal quantum number	n	1, 2, 3, ...
Angular momentum	l	0, 1, 2, ... $n - 1$
Angular momentum projection	m	0, ± 1 , ± 2 , ... $\pm l$

Name	Symbol	Allowed values
Spin	s	$1/2$ (electrons)
Spin projection	m_s	$-1/2, +1/2$

Summary of Quantum Numbers of an Electron in a Hydrogen Atom

Note that the intrinsic quantum numbers introduced in this section (s and m_s) are valid for many particles, not just electrons. For example, quarks within an atomic nucleus are also spin-half particles. As we will see later, quantum numbers help to classify subatomic particles and enter into scientific models that attempt to explain how the universe works.

Summary

- The state of an electron in a hydrogen atom can be expressed in terms of five quantum numbers.
- The spin angular momentum quantum of an electron is $= +1/2$. The spin angular momentum projection quantum number is $m_s = +1/2$ or $-1/2$ (spin up or spin down).
- The fine and hyperfine structures of the hydrogen spectrum are explained by magnetic interactions within the atom.

Conceptual Questions

Exercise:

Problem:

Explain how a hydrogen atom in the ground state ($l = 0$) can interact magnetically with an external magnetic field.

Solution:

Even in the ground state ($l = 0$), a hydrogen atom has magnetic properties due the intrinsic (internal) electron spin. The magnetic

moment of an electron is proportional to its spin.

Exercise:

Problem:

Compare orbital angular momentum with spin angular momentum of an electron in the hydrogen atom.

Exercise:

Problem:

List all the possible values of s and m_s for an electron. Are there particles for which these values are different?

Solution:

For all electrons, $s = \frac{1}{2}$ and $m_s = \pm\frac{1}{2}$. As we will see, not all particles have the same spin quantum number. For example, a photon has a spin 1 ($s = 1$), and a Higgs boson has spin 0 ($s = 0$).

Exercise:

Problem:

Are the angular momentum vectors $\vec{\mathbf{L}}$ and $\vec{\mathbf{S}}$ necessarily aligned?

Exercise:

Problem: What is spin-orbit coupling?

Solution:

An electron has a magnetic moment associated with its intrinsic (internal) spin. Spin-orbit coupling occurs when this interacts with the magnetic field produced by the orbital angular momentum of the electron.

Problems

Exercise:**Problem:**

What is the magnitude of the spin momentum of an electron? (Express your answer in terms of \hbar .)

Exercise:**Problem:**

What are the possible polar orientations of the spin momentum vector for an electron?

Solution:

Spin up (relative to positive z-axis):

$$\theta = 55^\circ.$$

Spin down (relative to positive z-axis):

$$\theta = \cos^{-1} \left(\frac{S_z}{S} \right) = \cos^{-1} \left(\frac{-\frac{1}{2}}{\frac{\sqrt{3}}{2}} \right) = \cos^{-1} \left(\frac{-1}{\sqrt{3}} \right) = 125^\circ.$$

Exercise:**Problem:**

For $n = 1$, write all the possible sets of quantum numbers (n, l, m, m_s).

Exercise:**Problem:**

A hydrogen atom is placed in an external uniform magnetic field ($B = 200 \text{ T}$). Calculate the wavelength of light produced in a transition from a spin up to spin down state.

Solution:

The spin projection quantum number is $m_s = \pm\frac{1}{2}$, so the z-component of the magnetic moment is

$$\mu_z = \pm\mu_B.$$

The potential energy associated with the interaction between the

electron and the external magnetic field is

$$U = \mp \mu_B B.$$

The energy difference between these states is $\Delta E = 2\mu_B B$, so the wavelength of light produced is

$$\lambda = 8.38 \times 10^{-5} \text{ m} \approx 84 \mu\text{m}$$

Exercise:

Problem:

If the magnetic field in the preceding problem is quadrupled, what happens to the wavelength of light produced in a transition from a spin up to spin down state?

Exercise:

Problem:

If the magnetic moment in the preceding problem is doubled, what happens to the frequency of light produced in a transition from a spin-up to spin-down state?

Solution:

It is increased by a factor of 2.

Exercise:

Problem:

For $n = 2$, write all the possible sets of quantum numbers (n, l, m, m_s) .

Glossary

fine structure

detailed structure of atomic spectra produced by spin-orbit coupling

hyperfine structure

detailed structure of atomic spectra produced by spin-orbit coupling

spin-flip transitions

atomic transitions between states of an electron-proton system in which the magnetic moments are aligned and not aligned

spin-orbit coupling

interaction between the electron magnetic moment and the magnetic field produced by the orbital angular momentum of the electron

spin projection quantum number (m_s)

quantum number associated with the z-component of the spin angular momentum of an electron

spin quantum number (s)

quantum number associated with the spin angular momentum of an electron

The Exclusion Principle and the Periodic Table

By the end of this section, you will be able to:

- Explain the importance of Pauli's exclusion principle to an understanding of atomic structure and molecular bonding
- Explain the structure of the periodic table in terms of the total energy, orbital angular momentum, and spin of individual electrons in an atom
- Describe the electron configuration of atoms in the periodic table

So far, we have studied only hydrogen, the simplest chemical element. We have found that an electron in the hydrogen atom can be completely specified by five quantum numbers:

Equation:

n :	principal quantum number
l :	angular momentum quantum number
m :	angular momentum projection quantum number
s :	spin quantum number
m_s :	spin projection quantum number

To construct the ground state of a neutral multi-electron atom, imagine starting with a nucleus of charge Ze (that is, a nucleus of atomic number Z) and then adding Z electrons one by one. Assume that each electron moves in a spherically symmetrical electric field produced by the nucleus and all other electrons of the atom. The assumption is valid because the electrons are distributed randomly around the nucleus and produce an average electric field (and potential) that is spherically symmetrical. The electric potential $U(r)$ for each electron does not follow the simple $-1/r$ form because of interactions between electrons, but it turns out that we can still label each individual electron state by quantum numbers, (n, l, m, s, m_s) . (The spin quantum number s is the same for all electrons, so it will not be used in this section.)

The structure and chemical properties of atoms are explained in part by **Pauli's exclusion principle**: No two electrons in an atom can have the

same values for all four quantum numbers (n, l, m, m_s). This principle is related to two properties of electrons: All electrons are identical (“when you’ve seen one electron, you’ve seen them all”) and they have half-integral spin ($s = 1/2$). Sample sets of quantum numbers for the electrons in an atom are given in [\[link\]](#). Consistent with Pauli’s exclusion principle, no two rows of the table have the exact same set of quantum numbers.

n	l	m	m_s	Subshell symbol	No. of electrons: subshell	No. of electrons: shell
1	0	0	$\frac{1}{2}$	1s	2	2
1	0	0	$-\frac{1}{2}$			
2	0	0	$\frac{1}{2}$	2s	2	8
2	0	0	$-\frac{1}{2}$			
2	1	-1	$\frac{1}{2}$	2p	6	
2	1	-1	$-\frac{1}{2}$			
2	1	0	$\frac{1}{2}$			
2	1	0	$-\frac{1}{2}$			
2	1	1	$\frac{1}{2}$			
2	1	1	$-\frac{1}{2}$			

n	l	m	m_s	Subshell symbol	No. of electrons: subshell	No. of electrons: shell
3	0	0	$\frac{1}{2}$	3s	2	18
3	0	0	$-\frac{1}{2}$			
3	1	-1	$\frac{1}{2}$	3p	6	
3	1	-1	$-\frac{1}{2}$			
3	1	0	$\frac{1}{2}$			
3	1	0	$-\frac{1}{2}$			
3	1	1	$\frac{1}{2}$			
3	1	1	$-\frac{1}{2}$			
3	2	-2	$\frac{1}{2}$	3d	10	
3	2	-2	$-\frac{1}{2}$			
3	2	-1	$\frac{1}{2}$			
3	2	-1	$-\frac{1}{2}$			
3	2	0	$\frac{1}{2}$			

n	l	m	m_s	Subshell symbol	No. of electrons: subshell	No. of electrons: shell
3	2	0	$-\frac{1}{2}$			
3	2	1	$\frac{1}{2}$			
3	2	1	$-\frac{1}{2}$			
3	2	2	$\frac{1}{2}$			
3	2	2	$-\frac{1}{2}$			

Electron States of Atoms Because of Pauli's exclusion principle, no two electrons in an atom have the same set of four quantum numbers.

Electrons with the same principal quantum number n are said to be in the same shell, and those that have the same value of l are said to occupy the same subshell. An electron in the $n = 1$ state of a hydrogen atom is denoted 1s, where the first digit indicates the shell ($n = 1$) and the letter indicates the subshell (s, p, d, f, \dots correspond to $l = 0, 1, 2, 3, \dots$). Two electrons in the $n = 1$ state are denoted as $1s^2$, where the superscript indicates the number of electrons. An electron in the $n = 2$ state with $l = 1$ is denoted 2p. The combination of two electrons in the $n = 2$ and $l = 0$ state, and three electrons in the $n = 2$ and $l = 1$ state is written as $2s^2 2p^3$, and so on. This representation of the electron state is called the **electron configuration** of the atom. The electron configurations for several atoms are given in [\[link\]](#). Electrons in the outer shell of an atom are called **valence electrons**. Chemical bonding between atoms in a molecule are explained by the transfer and sharing of valence electrons.

Element	Electron Configuration	Spin Alignment
H	$1s^1$	(\uparrow)
He	$1s^2$	$(\uparrow\downarrow)$
Li	$1s^2 2s^1$	(\uparrow)
Be	$1s^2 2s^2$	$(\uparrow\downarrow)$
B	$1s^2 2s^2 2p^1$	$(\uparrow\downarrow)(\uparrow)$
C	$1s^2 2s^2 2p^2$	$(\uparrow\downarrow)(\uparrow)(\uparrow)$
N	$1s^2 2s^2 2p^3$	$(\uparrow\downarrow)(\uparrow)(\uparrow)(\uparrow)$
O	$1s^2 2s^2 2p^4$	$(\uparrow\downarrow)(\uparrow\downarrow)(\uparrow)(\uparrow)$
F	$1s^2 2s^2 2p^5$	$(\uparrow\downarrow)(\uparrow\downarrow)(\uparrow\downarrow)(\uparrow)$
Ne	$1s^2 2s^2 2p^6$	$(\uparrow\downarrow)(\uparrow\downarrow)(\uparrow\downarrow)(\uparrow\downarrow)$
Na	$1s^2 2s^2 2p^6 3s^1$	(\uparrow)
Mg	$1s^2 2s^2 2p^6 3s^2$	$(\uparrow\downarrow)$
Al	$1s^2 2s^2 2p^6 3s^2 3p^1$	$(\uparrow\downarrow)(\uparrow)$

Electron Configurations of Electrons in an Atom
 The symbol (\uparrow) indicates an unpaired electron in the outer shell, whereas the symbol $(\uparrow\downarrow)$ indicates a pair of spin-up and -down electrons in an outer shell.

The maximum number of electrons in a subshell depends on the value of the angular momentum quantum number, l . For a given a value l , there are $2l + 1$ orbital angular momentum states. However, each of these states can

be filled by two electrons (spin up and down, $\uparrow\downarrow$). Thus, the maximum number of electrons in a subshell is

Note:

Equation:

$$N = 2(2l + 1) = 4l + 2.$$

In the $2s$ ($l = 0$) subshell, the maximum number of electrons is 2. In the $2p$ ($l = 1$) subshell, the maximum number of electrons is 6. Therefore, the total maximum number of electrons in the $n = 2$ shell (including both the $l = 0$ and 1 subshells) is $2 + 6$ or 8. In general, the maximum number of electrons in the n th shell is $2n^2$.

Example:

Subshells and Totals for $n = 3$

How many subshells are in the $n = 3$ shell? Identify each subshell and calculate the maximum number of electrons that will fill each. Show that the maximum number of electrons that fill an atom is $2n^2$.

Strategy

Subshells are determined by the value of l ; thus, we first determine which values of l are allowed, and then we apply the equation “maximum number of electrons that can be in a subshell = $2(2l + 1)$ ” to find the number of electrons in each subshell.

Solution

Because $n = 3$, we know that l can be 0, 1, or 2; thus, there are three possible subshells. In standard notation, they are labeled the $3s$, $3p$, and $3d$ subshells. We have already seen that two electrons can be in an s state, and six in a p state, but let us use the equation “maximum number of electrons that can be in a subshell = $2(2l + 1)$ ” to calculate the maximum number in each:

Equation:

$3s$ has $l = 0$; thus, $2(2l + 1) = 2(0 + 1) = 2$

$3p$ has $l = 1$; thus, $2(2l + 1) = 2(2 + 1) = 6$

$3d$ has $l = 2$; thus, $2(2l + 1) = 2(4 + 1) = 10$

Total = 18

(in the $n = 3$ shell).

The equation “maximum number of electrons that can be in a shell = $2n^2$ ” gives the maximum number in the $n = 3$ shell to be

Equation:

Maximum number of electrons = $2n^2 = 2(3)^2 = 2(9) = 18$.

Significance

The total number of electrons in the three possible subshells is thus the same as the formula $2n^2$. In standard (spectroscopic) notation, a filled $n = 3$ shell is denoted as $3s^23p^63d^{10}$. Shells do not fill in a simple manner. Before the $n = 3$ shell is completely filled, for example, we begin to find electrons in the $n = 4$ shell.

The structure of the periodic table ([\[link\]](#)) can be understood in terms of shells and subshells, and, ultimately, the total energy, orbital angular momentum, and spin of the electrons in the atom. A detailed discussion of the periodic table is left to a chemistry course—we sketch only its basic features here. In this discussion, we assume that the atoms are electrically neutral; that is, they have the same number of electrons and protons. (Recall that the total number of protons in an atomic nucleus is called the atomic number, Z .)

First, the periodic table is arranged into columns and rows. The table is read left to right and top to bottom in the order of increasing atomic number Z . Atoms that belong to the same column or **chemical group** share many of the same chemical properties. For example, the Li and Na atoms (in the first

column) bond to other atoms in a similar way. The first row of the table corresponds to the $1s$ ($l = 0$) shell of an atom.

Consider the hypothetical procedure of adding electrons, one by one, to an atom. For hydrogen (H) (upper left), the $1s$ shell is filled with either a spin up or down electron (\uparrow or \downarrow). This lone electron is easily shared with other atoms, so hydrogen is chemically active. For helium (He) (upper right), the $1s$ shell is filled with both a spin up and a spin down ($\uparrow\downarrow$) electron. This “fills” the $1s$ shell, so a helium atom tends not to share electrons with other atoms. The helium atom is said to be chemically inactive, inert, or noble; likewise, helium gas is said to be an inert gas or noble gas.

Note:

Build an atom by adding and subtracting protons, neutrons, and electrons. How does the element, charge, and mass change? Visit [PhET Explorations: Build an Atom](#) to explore the answers to these questions.

Next, we look at the right side of the table. For boron (B), the 1s and 2s shells are filled and the 2p ($l = 1$) shell contains either a spin up or down electron (\uparrow or \downarrow). From carbon (C) to neon (N), we fill the 2p shell. The maximum number of electrons in the 2p shells is $4l + 2 = 4(2) + 2 = 6$. For neon (Ne), the 1s shell is filled with a spin-up and spin-down electron ($\uparrow\downarrow$), and the 2p shell is filled with six electrons ($\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow$). This “fills” the 1s, 2s, and 2p subshells, so like helium, the neon atom tends not to share electrons with other atoms.

The process of electron filling repeats in the third row. However, beginning in the fourth row, the pattern is broken. The actual order of order of electron filling is given by

1s, 2s, 2p, 3s, 3p, 4s, **3d**, 4p, 5s, **4d**, 5p, 6s, **4f**, **5d**, 6p, 7s,...

Notice that the 3d, 4d, 4f, and 5d subshells (in bold) are filled out of order; this occurs because of interactions between electrons in the atom, which so far we have neglected. The **transition metals** are elements in the gap between the first two columns and the last six columns that contain electrons that fill the d ($l = 1$) subshell. As expected, these atoms are arranged in $4l + 2 = 4(2) + 2 = 10$ columns. The structure of the periodic table can be understood in terms of the quantization of the total energy (n), orbital angular momentum (l), and spin (s). The first two columns correspond to the s ($l = 0$) subshell, the next six columns correspond to the p ($l = 1$) subshell, and the gap between these columns corresponds to the d ($l = 2$) subshell.

The periodic table also gives information on molecular bonding. To see this, consider atoms in the left-most column (the so-called alkali metals including: Li, Na, and K). These atoms contain a single electron in the 2s subshell, which is easily donated to other atoms. In contrast, atoms in the second-to-right column (the halogens: for example, Cl, F, and Br) are relatively stingy in sharing electrons. These atoms would much rather accept an electron, because they are just one electron shy of a filled shell (“of being noble”).

Therefore, if a Na atom is placed in close proximity to a Cl atom, the Na atom freely donates its 2s electron and the Cl atom eagerly accepts it. In the

process, the Na atom (originally a neutral charge) becomes positively charged and the Cl (originally a neutral charge) becomes negatively charged. Charged atoms are called ions. In this case, the ions are Na^+ and Cl^- , where the superscript indicates charge of the ion. The electric (Coulomb) attraction between these atoms forms a NaCl (salt) molecule. A chemical bond between two ions is called an **ionic bond**. There are many kinds of chemical bonds. For example, in an oxygen molecule O_2 electrons are equally shared between the atoms. The bonding of oxygen atoms is an example of a **covalent bond**.

Summary

- Pauli's exclusion principle states that no two electrons in an atom can have all the same quantum numbers.
- The structure of the periodic table of elements can be explained in terms of the total energy, orbital angular momentum, and spin of electrons in an atom.
- The state of an atom can be expressed by its electron configuration, which describes the shells and subshells that are filled in the atom.

Conceptual Questions

Exercise:

Problem:

What is Pauli's exclusion principle? Explain the importance of this principle for the understanding of atomic structure and molecular bonding.

Exercise:

Problem:

Compare the electron configurations of the elements in the same column of the periodic table.

Solution:

Elements that belong in the same column in the periodic table of elements have the same fillings of their outer shells, and therefore the same number of valence electrons. For example:

Li: $1s^2 2s^1$ (one valence electron in the $n = 2$ shell)

Na: $1s^2 2s^2 2p^6 3s^1$ (one valence electron in the $n = 3$ shell)

Both, Li and Na belong to first column.

Exercise:

Problem:

Compare the electron configurations of the elements that belong in the same row of the periodic table of elements.

Problems

Exercise:

Problem: (a) How many electrons can be in the $n = 4$ shell?

(b) What are its subshells, and how many electrons can be in each?

Solution:

a. 32; b.

ℓ		$2(2\ell + 1)$	
0	<i>s</i>	$2(0 + 1)$	= 2
1	<i>p</i>	$2(2 + 1)$	= 6
2	<i>d</i>	$2(4 + 1)$	= 10
3	<i>f</i>	$2(6 + 1)$	= 14
<hr/>			32

Exercise:

Problem:

(a) What is the minimum value of l for a subshell that contains 11 electrons?

(b) If this subshell is in the $n = 5$ shell, what is the spectroscopic notation for this atom?

Exercise:**Problem:**

Unreasonable result. Which of the following spectroscopic notations are not allowed? (a) $5s^1$ (b) $1d^1$ (c) $4s^3$ (d) $3p^7$ (e) $5g^{15}$. State which rule is violated for each notation that is not allowed.

Solution:

a. and e. are allowed; the others are not allowed.

b. $l = 3$ not allowed for $n = 1$, $l \leq (n - 1)$.

c. Cannot have three electrons in s subshell because $3 > 2(2l + 1) = 2$.

d. Cannot have seven electrons in p subshell (max of 6) $2(2l + 1) = 2(2 + 1) = 6$.

Exercise:

Problem: Write the electron configuration for potassium.

Exercise:

Problem: Write the electron configuration for iron.

Solution:

$[\text{Ar}] 4s^2 3d^6$

Exercise:

Problem:

The valence electron of potassium is excited to a $5d$ state. (a) What is the magnitude of the electron's orbital angular momentum? (b) How many states are possible along a chosen direction?

Exercise:**Problem:**

(a) If one subshell of an atom has nine electrons in it, what is the minimum value of l ? (b) What is the spectroscopic notation for this atom, if this subshell is part of the $n = 3$ shell?

Solution:

- a. The minimum value of ℓ is $l = 2$ to have nine electrons in it.
- b. $3d^9$.

Exercise:

Problem: Write the electron configuration for magnesium.

Exercise:

Problem: Write the electron configuration for carbon.

Solution:

$[\text{He}] 2s^2 2p^2$

Exercise:**Problem:**

The magnitudes of the resultant spins of the electrons of the elements B through Ne when in the ground state are: $\sqrt{3}\hbar/2$, $\sqrt{2}\hbar$, $\sqrt{15}\hbar/2$, $\sqrt{2}\hbar$, $\sqrt{3}\hbar/2$, and 0, respectively. Argue that these spins are consistent with Hund's rule.

Glossary

chemical group

group of elements in the same column of the periodic table that possess similar chemical properties

covalent bond

chemical bond formed by the sharing of electrons between two atoms

electron configuration

representation of the state of electrons in an atom, such as $1s^2 2s^1$ for lithium

ionic bond

chemical bond formed by the electric attraction between two oppositely charged ions

Pauli's exclusion principle

no two electrons in an atom can have the same values for all four quantum numbers (n, l, m, m_s)

transition metal

element that is located in the gap between the first two columns and the last six columns of the table of elements that contains electrons that fill the d subshell

valence electron

electron in the outer shell of an atom that participates in chemical bonding

Atomic Spectra and X-rays

By the end of this section, you will be able to:

- Describe the absorption and emission of radiation in terms of atomic energy levels and energy differences
- Use quantum numbers to estimate the energy, frequency, and wavelength of photons produced by atomic transitions in multi-electron atoms
- Explain radiation concepts in the context of atomic fluorescence and X-rays

The study of atomic spectra provides most of our knowledge about atoms. In modern science, atomic spectra are used to identify species of atoms in a range of objects, from distant galaxies to blood samples at a crime scene.

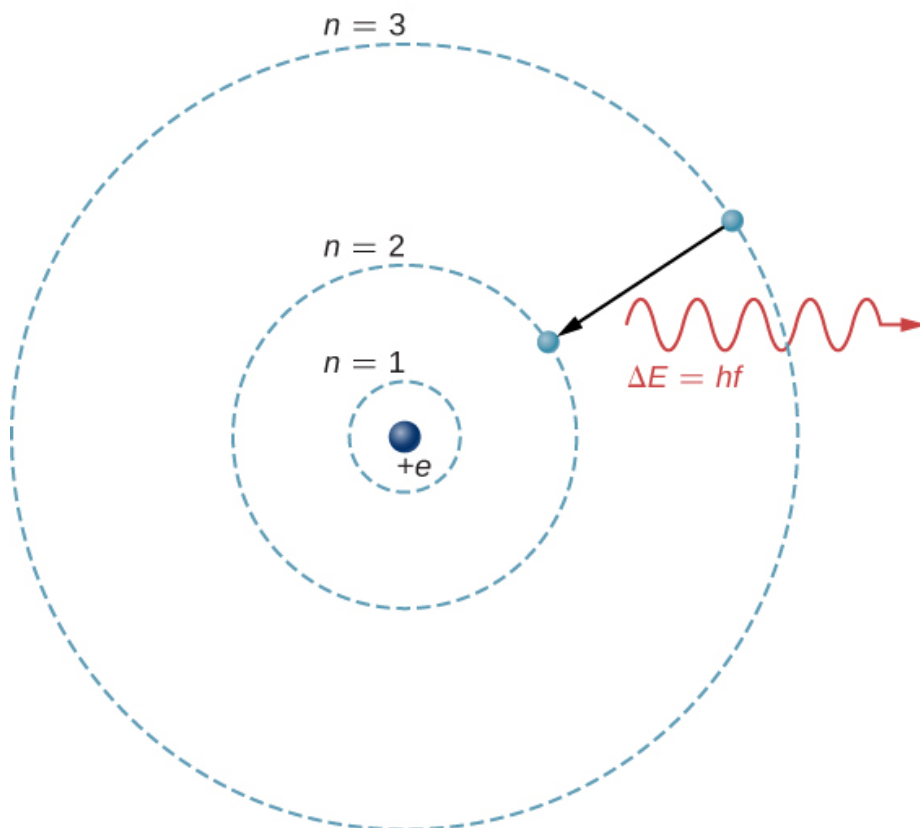
The theoretical basis of atomic spectroscopy is the transition of electrons between energy levels in atoms. For example, if an electron in a hydrogen atom makes a transition from the $n = 3$ to the $n = 2$ shell, the atom emits a photon with a wavelength

Equation:

$$\lambda = \frac{c}{f} = \frac{h \cdot c}{h \cdot f} = \frac{hc}{\Delta E} = \frac{hc}{E_3 - E_2},$$

where $\Delta E = E_3 - E_2$ is energy carried away by the photon and $hc = 1240 \text{ eV} \cdot \text{nm}$. After this radiation passes through a spectrometer, it appears as a sharp spectral line on a screen.

The Bohr model of this process is shown in [\[link\]](#). If the electron later absorbs a photon with energy ΔE , the electron returns to the $n = 3$ shell. (We examined the Bohr model earlier, in [Photons and Matter Waves](#).)



An electron transition from the $n = 3$ to the $n = 2$ shell of a hydrogen atom.

To understand atomic transitions in multi-electron atoms, it is necessary to consider many effects, including the Coulomb repulsion between electrons and internal magnetic interactions (spin-orbit and spin-spin couplings). Fortunately, many properties of these systems can be understood by neglecting interactions between electrons and representing each electron by its own single-particle wave function ψ_{nlm} .

Atomic transitions must obey **selection rules**. These rules follow from principles of quantum mechanics and symmetry. Selection rules classify transitions as either allowed or forbidden. (Forbidden transitions do occur, but the probability of the typical forbidden transition is very small.) For a hydrogen-like atom, atomic transitions that involve electromagnetic interactions (the emission and absorption of photons) obey the following selection rule:

Note:
Equation:

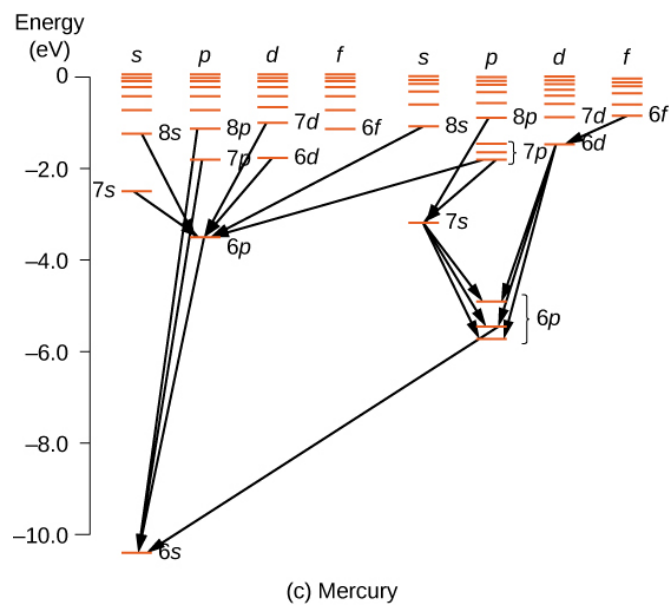
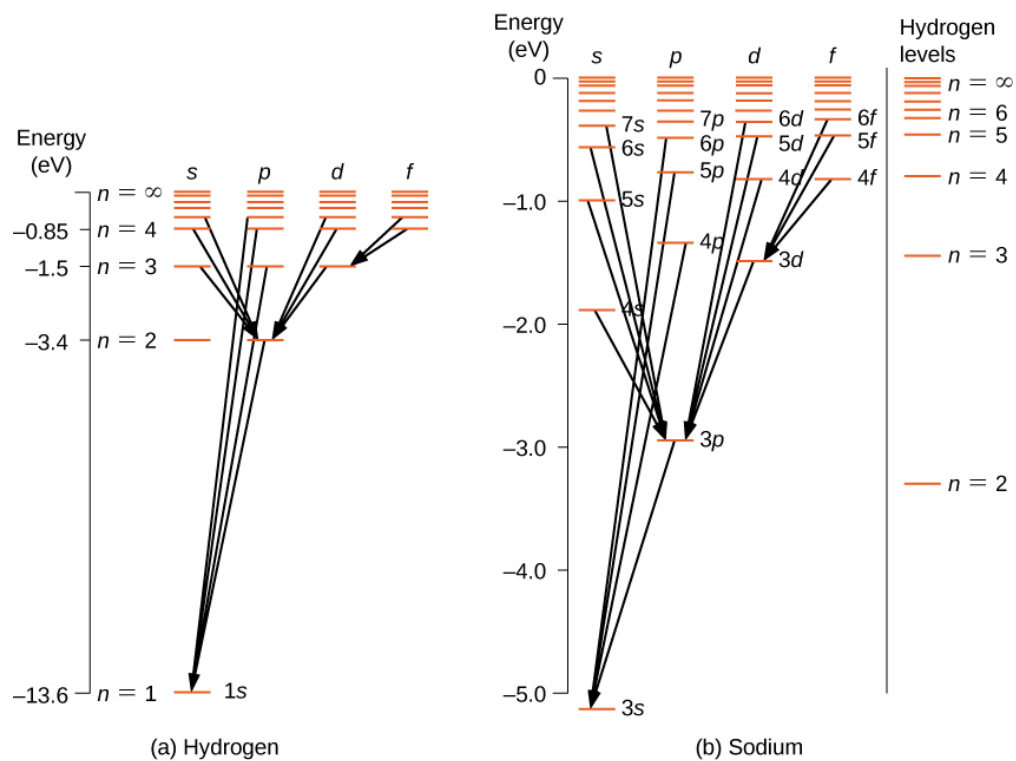
$$\Delta l = \pm 1,$$

where l is associated with the magnitude of orbital angular momentum,

Equation:

$$L = \sqrt{l(l+1)}\hbar.$$

For multi-electron atoms, similar rules apply. To illustrate this rule, consider the observed atomic transitions in hydrogen (H), sodium (Na), and mercury (Hg) ([link](#)). The horizontal lines in this diagram correspond to atomic energy levels, and the transitions allowed by this selection rule are shown by lines drawn between these levels. The energies of these states are on the order of a few electron volts, and photons emitted in transitions are in the visible range. Technically, atomic transitions can violate the selection rule, but such transitions are uncommon.



Energy-level diagrams for (a) hydrogen, (b) sodium, and (c) mercury. For comparison, hydrogen energy levels are shown in the sodium diagram.

The hydrogen atom has the simplest energy-level diagram. If we neglect electron spin, all states with the same value of n have the same total energy. However, spin-orbit coupling splits the $n = 2$ states into two angular momentum states (s and p) of slightly different energies. (These levels are not vertically displaced, because the energy splitting is too small to show up in this diagram.) Likewise, spin-orbit coupling splits the $n = 3$ states into three angular momentum states (s , p , and d).

The energy-level diagram for hydrogen is similar to sodium, because both atoms have one electron in the outer shell. The valence electron of sodium moves in the electric field of a nucleus shielded by electrons in the inner shells, so it does not experience a simple $1/r$ Coulomb potential and its total energy depends on both n and l . Interestingly, mercury has two separate energy-level diagrams; these diagrams correspond to two net spin states of its $6s$ (valence) electrons.

Example:**The Sodium Doublet**

The spectrum of sodium is analyzed with a spectrometer. Two closely spaced lines with wavelengths 589.00 nm and 589.59 nm are observed. (a) If the doublet corresponds to the excited (valence) electron that transitions from some excited state down to the $3s$ state, what was the original electron angular momentum? (b) What is the energy difference between these two excited states?

Strategy

Sodium and hydrogen belong to the same column or chemical group of the periodic table, so sodium is “hydrogen-like.” The outermost electron in sodium is in the $3s$ ($l = 0$) subshell and can be excited to higher energy levels. As for hydrogen, subsequent transitions to lower energy levels must obey the selection rule:

Equation:

$$\Delta l = \pm 1.$$

We must first determine the quantum number of the initial state that satisfies the selection rule. Then, we can use this number to determine the magnitude of orbital angular momentum of the initial state.

Solution

- a. Allowed transitions must obey the selection rule. If the quantum number of the initial state is $l = 0$, the transition is forbidden because $\Delta l = 0$. If the quantum number of the initial state is $l = 2, 3, 4, \dots$ the transition is forbidden because $\Delta l > 1$. Therefore, the quantum of the initial state must be $l = 1$. The orbital angular momentum of the initial state is

Equation:

$$L = \sqrt{l(l+1)}\hbar = 1.41\hbar.$$

- b. Because the final state for both transitions is the same (3s), the difference in energies of the photons is equal to the difference in energies of the two excited states. Using the equation

Equation:

$$\Delta E = hf = h \left(\frac{c}{\lambda} \right),$$

we have

Equation:

$$\begin{aligned} \Delta E &= hc \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \\ &= (4.14 \times 10^{-15} \text{ eVs}) (3.00 \times 10^8 \text{ m/s}) \times \left(\frac{1}{589.00 \times 10^{-9} \text{ m}} - \frac{1}{589.59 \times 10^{-9} \text{ m}} \right) \\ &= 2.11 \times 10^{-3} \text{ eV}. \end{aligned}$$

Significance

To understand the difficulty of measuring this energy difference, we compare this difference with the average energy of the two photons emitted in the transition. Given an average wavelength of 589.30 nm, the average energy of the photons is

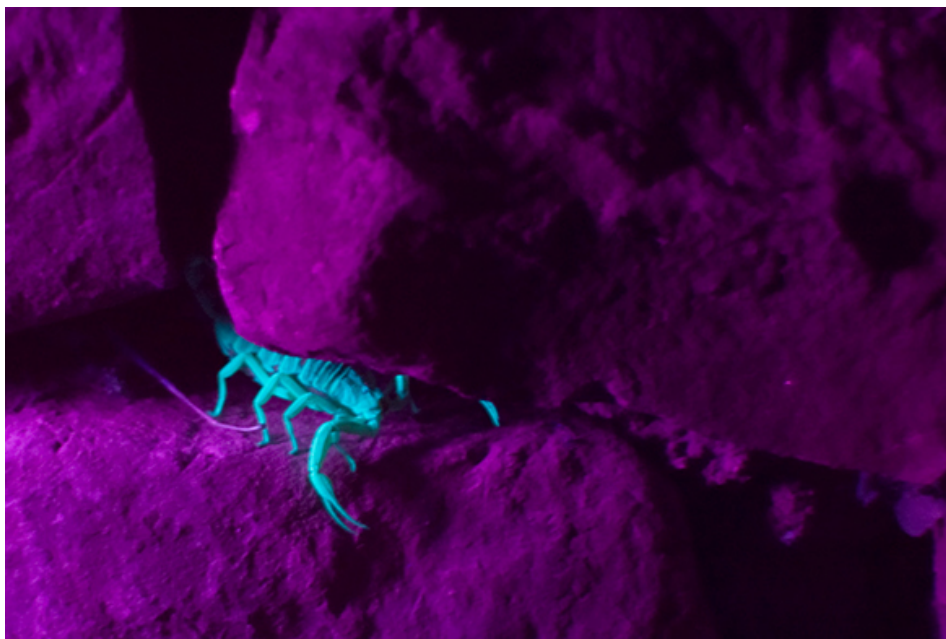
Equation:

$$E = \frac{hc}{\lambda} = \frac{(4.14 \times 10^{-15} \text{ eVs})(3.00 \times 10^8 \text{ m/s})}{589.30 \times 10^{-9} \text{ m}} = 2.11 \text{ eV}.$$

The energy difference ΔE is about 0.1% (1 part in 1000) of this average energy. However, a sensitive spectrometer can measure the difference.

Atomic Fluorescence

Fluorescence occurs when an electron in an atom is excited several steps above the ground state by the absorption of a high-energy ultraviolet (UV) photon. Once excited, the electron “de-excites” in two ways. The electron can drop back to the ground state, emitting a photon of the same energy that excited it, or it can drop in a series of smaller steps, emitting several low-energy photons. Some of these photons may be in the visible range. Fluorescent dye in clothes can make colors seem brighter in sunlight by converting UV radiation into visible light. Fluorescent lights are more efficient in converting electrical energy into visible light than incandescent filaments (about four times as efficient). [\[link\]](#) shows a scorpion illuminated by a UV lamp. Proteins near the surface of the skin emit a characteristic blue light.



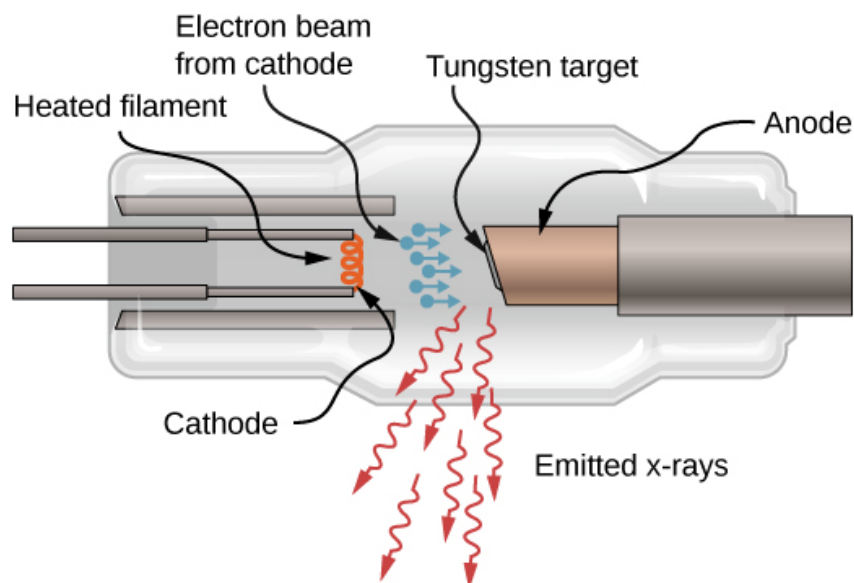
A scorpion glows blue under a UV lamp. (credit: Ken Bosma)

X-rays

The study of atomic energy transitions enables us to understand X-rays and X-ray technology. Like all electromagnetic radiation, X-rays are made of photons. X-ray photons are produced when electrons in the outermost shells of an atom drop to the inner shells. (Hydrogen atoms do not emit X-rays, because the electron energy levels are too closely spaced together to permit the emission of high-frequency radiation.) Transitions of this kind are normally forbidden because the lower states are already filled. However, if an inner shell has a vacancy (an inner electron is missing, perhaps from being knocked away by a high-speed electron), an electron from one of the outer shells can drop in energy to fill the vacancy. The energy gap for such a transition is relatively large, so wavelength of the radiated X-ray photon is relatively short.

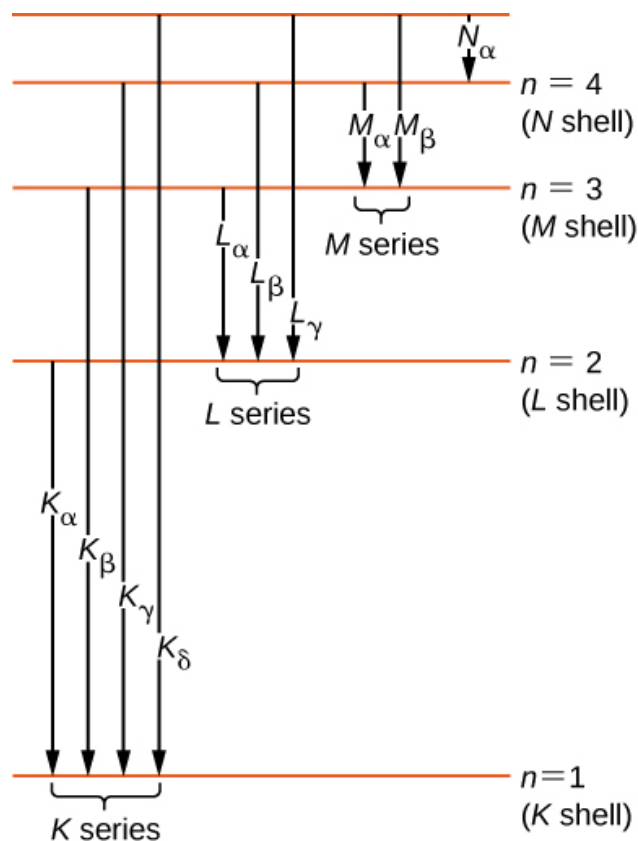
X-rays can also be produced by bombarding a metal target with high-energy electrons, as shown in [\[link\]](#). In the figure, electrons are boiled off a filament and accelerated by an electric field into a tungsten target. According to the classical theory of electromagnetism, *any* charged particle that accelerates emits radiation. Thus, when the electron strikes the tungsten target, and suddenly slows down, the electron emits **braking radiation**. (Braking radiation refers to radiation produced by any charged particle that is slowed by a medium.) In this case, braking radiation contains a continuous range of frequencies, because the electrons will collide with the target atoms in slightly different ways.

Braking radiation is not the only type of radiation produced in this interaction. In some cases, an electron collides with another inner-shell electron of a target atom, and knocks the electron out of the atom—billiard ball style. The empty state is filled when an electron in a higher shell drops into the state (drop in energy level) and emits an X-ray photon.



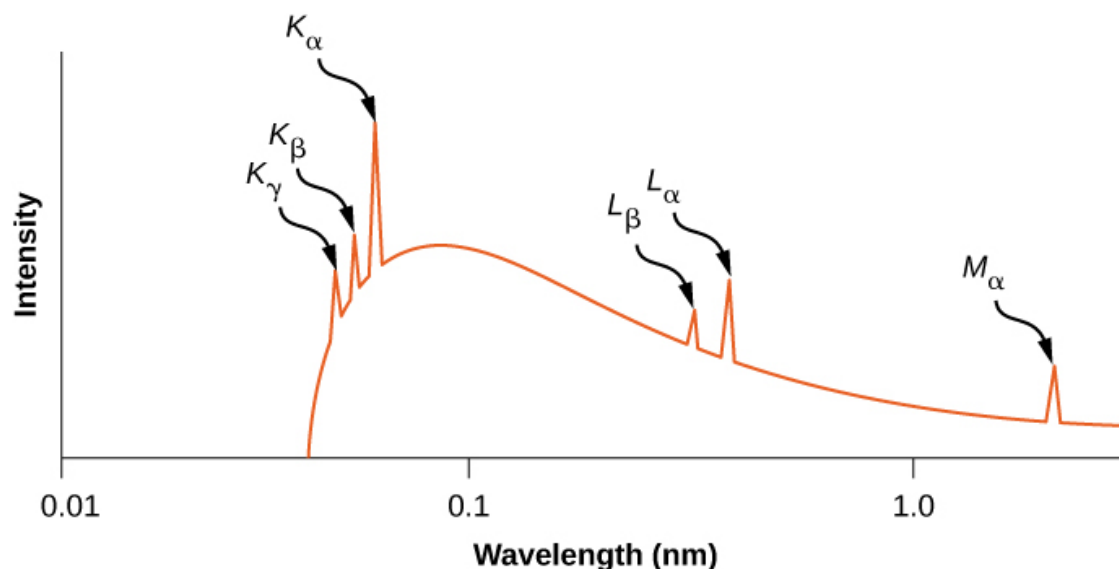
A sketch of an X-ray tube. X-rays are emitted from the tungsten target.

Historically, X-ray spectral lines were labeled with letters (K , L , M , N , ...). These letters correspond to the atomic shells ($n = 1, 2, 3, 4, \dots$). X-rays produced by a transition from any higher shell to the K ($n = 1$) shell are labeled as K X-rays. X-rays produced in a transition from the L ($n = 2$) shell are called K_α X-rays; X-rays produced in a transition from the M ($n = 3$) shell are called K_β X-rays; X-rays produced in a transition from the N ($n = 4$) shell are called K_γ X-rays; and so forth. Transitions from higher shells to L and M shells are labeled similarly. These transitions are represented by an energy-level diagram in [\[link\]](#).



X-ray transitions in an atom.

The distribution of X-ray wavelengths produced by striking metal with a beam of electrons is given in [\[link\]](#). X-ray transitions in the target metal appear as peaks on top of the braking radiation curve. Photon frequencies corresponding to the spikes in the X-ray distribution are called characteristic frequencies, because they can be used to identify the target metal. The sharp cutoff wavelength (just below the K_γ peak) corresponds to an electron that loses all of its energy to a single photon. Radiation of shorter wavelengths is forbidden by the conservation of energy.



X-ray spectrum from a silver target. The peaks correspond to characteristic frequencies of X-rays emitted by silver when struck by an electron beam.

Example:

X-Rays from Aluminum

Estimate the characteristic energy and frequency of the K_α X-ray for aluminum ($Z = 13$).

Strategy

A K_α X-ray is produced by the transition of an electron in the L ($n = 2$) shell to the K ($n = 1$) shell. An electron in the L shell “sees” a charge $Z = 13 - 1 = 12$, because one electron in the K shell shields the nuclear charge. (Recall, two electrons are not in the K shell because the other electron state is vacant.) The frequency of the emitted photon can be estimated from the energy difference between the L and K shells.

Solution

The energy difference between the L and K shells in a hydrogen atom is 10.2 eV. Assuming that other electrons in the L shell or in higher-energy shells do not shield the nuclear charge, the energy difference between the L and K shells in an atom with $Z = 13$ is approximately

Equation:

$$\Delta E_{L \rightarrow K} \approx (Z - 1)^2 (10.2 \text{ eV}) = (13 - 1)^2 (10.2 \text{ eV}) = 1.47 \times 10^3 \text{ eV}.$$

Based on the relationship $f = (\Delta E_{L \rightarrow K}) / h$, the frequency of the X-ray is

Equation:

$$f = \frac{1.47 \times 10^3 \text{ eV}}{4.14 \times 10^{-15} \text{ eV} \cdot \text{s}} = 3.55 \times 10^{17} \text{ Hz}.$$

Significance

The wavelength of the typical X-ray is 0.1–10 nm. In this case, the wavelength is:

Equation:

$$\lambda = \frac{c}{f} = \frac{3.0 \times 10^8 \text{ m/s}}{3.55 \times 10^{17} \text{ Hz}} = 8.5 \times 10^{-10} = 0.85 \text{ nm}.$$

Hence, the transition $L \rightarrow K$ in aluminum produces X-ray radiation.

X-ray production provides an important test of quantum mechanics. According to the Bohr model, the energy of a K_α X-ray depends on the nuclear charge or atomic number, Z . If Z is large, Coulomb forces in the atom are large, energy differences (ΔE) are large, and, therefore, the energy of radiated photons is large. To illustrate, consider a single electron in a multi-electron atom. Neglecting interactions between the electrons, the allowed energy levels are

Equation:

$$E_n = -\frac{Z^2 (13.6 \text{ eV})}{n^2},$$

where $n = 1, 2, \dots$ and Z is the atomic number of the nucleus. However, an electron in the L ($n = 2$) shell “sees” a charge $Z - 1$, because one electron in the K shell shields the nuclear charge. (Recall that there is only one electron in the K shell because the other electron was “knocked out.”) Therefore, the approximate energies of the electron in the L and K shells are

Equation:

$$\begin{aligned} E_L &\approx -\frac{(Z-1)^2 (13.6 \text{ eV})}{2^2} \\ E_K &\approx -\frac{(Z-1)^2 (13.6 \text{ eV})}{1^2}. \end{aligned}$$

The energy carried away by a photon in a transition from the L shell to the K shell is therefore

Equation:

$$\begin{aligned} \Delta E_{L \rightarrow K} &= (Z-1)^2 (13.6 \text{ eV}) \left(\frac{1}{1^2} - \frac{1}{2^2} \right) \\ &= (Z-1)^2 (10.2 \text{ eV}), \end{aligned}$$

where Z is the atomic number. In general, the X-ray photon energy for a transition from an outer shell to the K shell is

Equation:

$$\Delta E_{L \rightarrow K} = hf = \text{constant} \times (Z-1)^2,$$

or

Note:

Equation:

$$(Z - 1) = \text{constant} \sqrt{f},$$

where f is the frequency of a K_α X-ray. This equation is **Moseley's law**. For large values of Z , we have approximately

Equation:

$$Z \approx \text{constant} \sqrt{f}.$$

This prediction can be checked by measuring f for a variety of metal targets. This model is supported if a plot of Z versus \sqrt{f} data (called a **Moseley plot**) is linear. Comparison of model predictions and experimental results, for both the K and L series, is shown in [\[link\]](#). The data support the model that X-rays are produced when an outer shell electron drops in energy to fill a vacancy in an inner shell.

Note:

Exercise:

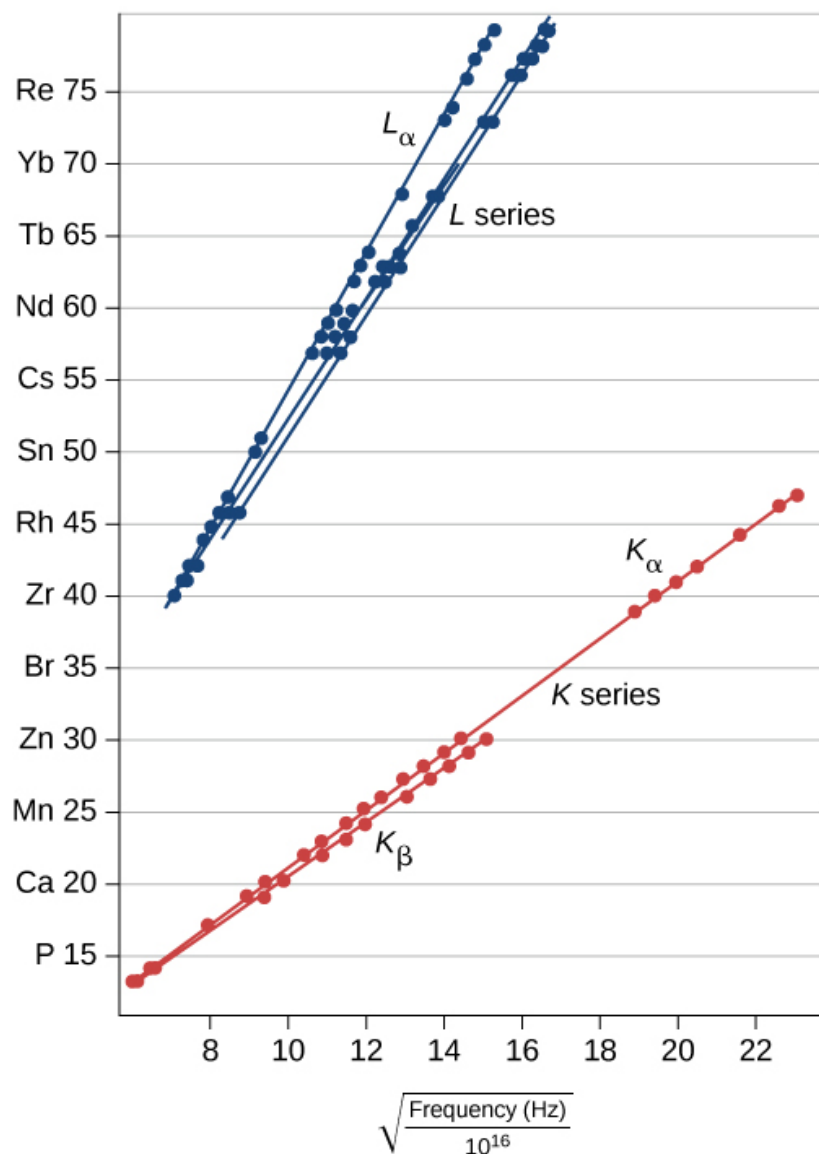
Problem:

Check Your Understanding X-rays are produced by bombarding a metal target with high-energy electrons. If the target is replaced by another with two times the atomic number, what happens to the frequency of X-rays?

Solution:

frequency quadruples

Moseley Plot of Characteristic X-Rays



A Moseley plot. These data were adapted from Moseley's original data (H. G. J. Moseley, *Philos. Mag.* (6) 77:703, 1914).

Example:

Characteristic X-Ray Energy

Calculate the approximate energy of a K_α X-ray from a tungsten anode in an X-ray tube.

Strategy

Two electrons occupy a filled K shell. A vacancy in this shell would leave one electron, so the effective charge for an electron in the L shell would be $Z - 1$ rather than Z . For tungsten, $Z = 74$, so the effective charge is 73. This number can be used to calculate the energy-level difference between the L and K shells, and, therefore, the energy carried away by a photon in the transition $L \rightarrow K$.

Solution

The effective Z is 73, so the K_α X-ray energy is given by

Equation:

$$E_{K_\alpha} = \Delta E = E_i - E_f = E_2 - E_1,$$

where

Equation:

$$E_1 = -\frac{Z^2}{1^2} E_0 = -\frac{73^2}{1} (13.6 \text{ eV}) = -72.5 \text{ keV}$$

and

Equation:

$$E_2 = -\frac{Z^2}{2^2} E_0 = -\frac{73^2}{4} (13.6 \text{ eV}) = -18.1 \text{ keV}.$$

Thus,

Equation:

$$E_{K_\alpha} = -18.1 \text{ keV} - (-72.5 \text{ keV}) = 54.4 \text{ keV}.$$

Significance

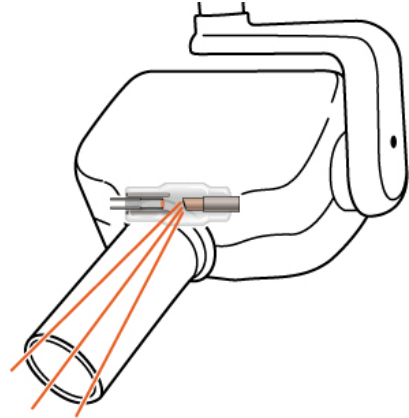
This large photon energy is typical of X-rays. X-ray energies become progressively larger for heavier elements because their energy increases approximately as Z^2 . An acceleration voltage of more than 50,000 volts is needed to “knock out” an inner electron from a tungsten atom.

X-ray Technology

X-rays have many applications, such as in medical diagnostics ([link](#)), inspection of luggage at airports ([link](#)), and even detection of cracks in crucial aircraft components. The most common X-ray images are due to shadows. Because X-ray photons have high energy, they penetrate materials that are opaque to visible light. The more energy an X-ray photon has, the more material it penetrates. The depth of penetration is related to the density of the material, as well as to the energy of the photon. The denser the material, the fewer X-ray photons get through and the darker the shadow. X-rays are effective at identifying bone breaks and tumors; however, overexposure to X-rays can damage cells in biological organisms.



(a)



(b)

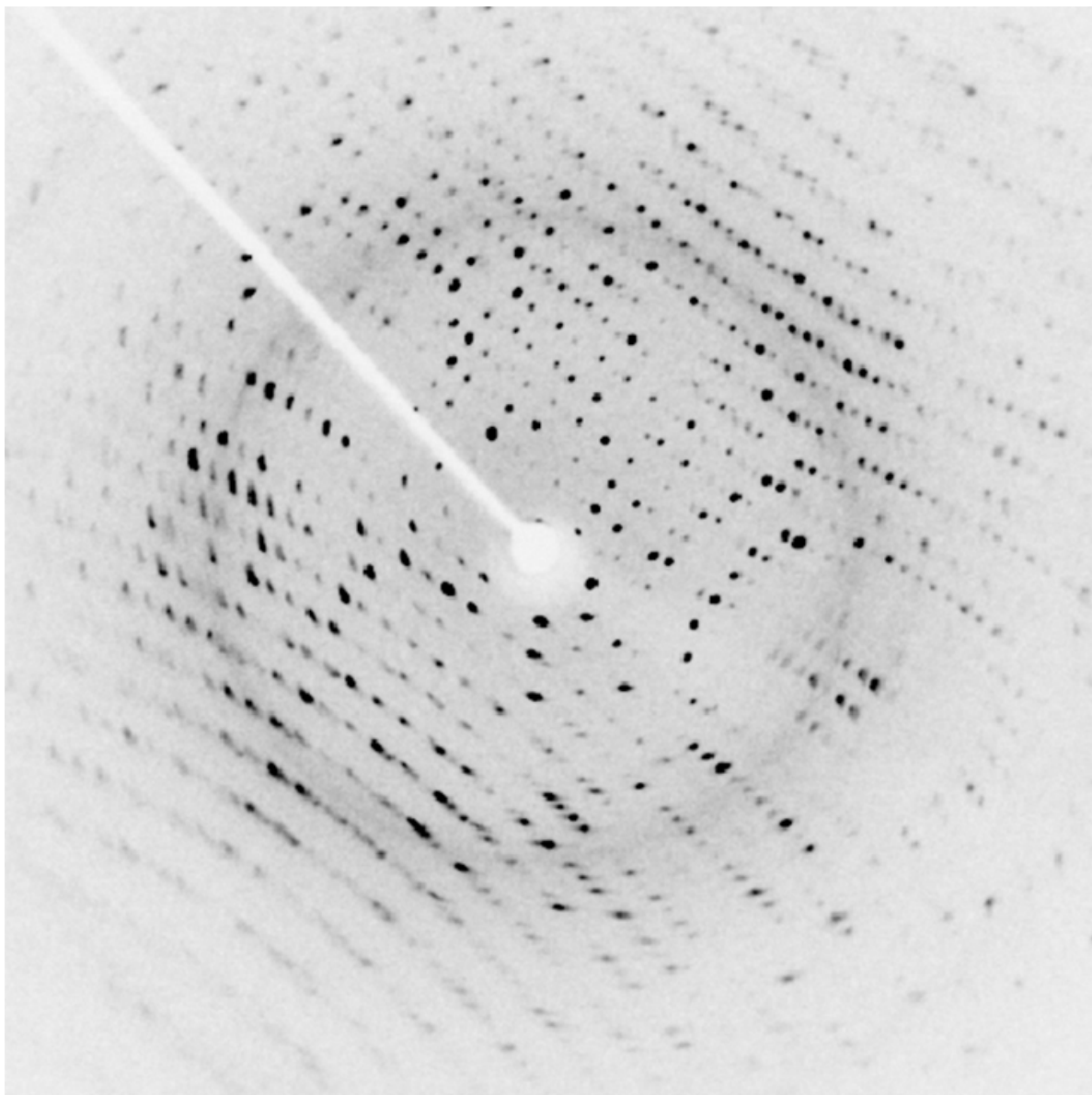
(a) An X-ray image of a person's teeth. (b) A typical X-ray machine in a dentist's office produces relatively low-energy radiation to minimize patient exposure. (credit a: modification of work by "Dmitry G"/Wikimedia Commons)



An X-ray image of a piece of luggage. The denser the material, the darker the shadow. Object colors relate to material composition—metallic objects show up as blue in this image. (credit: "IDuke"/Wikimedia Commons)

A standard X-ray image provides a two-dimensional view of the object. However, in medical applications, this view does not often provide enough information to draw firm conclusions. For example, in a two-dimensional X-ray image of the body, bones can easily hide soft tissues or organs. The CAT (computed axial tomography) scanner addresses this problem by collecting numerous X-ray images in “slices” throughout the body. Complex computer-image processing of the relative absorption of the X-rays, in different directions, can produce a highly detailed three-dimensional X-ray image of the body.

X-rays can also be used to probe the structures of atoms and molecules. Consider X-rays incident on the surface of a crystalline solid. Some X-ray photons reflect at the surface, and others reflect off the “plane” of atoms just below the surface. Interference between these photons, for different angles of incidence, produces a beautiful image on a screen ([\[link\]](#)). The interaction of X-rays with a solid is called X-ray diffraction. The most famous example using X-ray diffraction is the discovery of the double-helix structure of DNA.



X-ray diffraction from the crystal of a protein (hen egg lysozyme) produced this interference pattern. Analysis of the pattern yields information about the structure of the protein. (credit: "Del45"/Wikimedia Commons)

Summary

- Radiation is absorbed and emitted by atomic energy-level transitions.
- Quantum numbers can be used to estimate the energy, frequency, and wavelength of photons produced by atomic transitions.

- Atomic fluorescence occurs when an electron in an atom is excited several steps above the ground state by the absorption of a high-energy ultraviolet (UV) photon.
- X-ray photons are produced when a vacancy in an inner shell of an atom is filled by an electron from the outer shell of the atom.
- The frequency of X-ray radiation is related to the atomic number Z of an atom.

Conceptual Questions

Exercise:

Problem:

Atomic and molecular spectra are discrete. What does discrete mean, and how are discrete spectra related to the quantization of energy and electron orbits in atoms and molecules?

Solution:

Atomic and molecular spectra are said to be “discrete,” because only certain spectral lines are observed. In contrast, spectra from a white light source (consisting of many photon frequencies) are continuous because a continuous “rainbow” of colors is observed.

Exercise:

Problem:

Discuss the process of the absorption of light by matter in terms of the atomic structure of the absorbing medium.

Exercise:

Problem:

NGC1763 is an emission nebula in the Large Magellanic Cloud just outside our Milky Way Galaxy. Ultraviolet light from hot stars ionize the hydrogen atoms in the nebula. As protons and electrons recombine, light in the visible range is emitted. Compare the energies of the photons involved in these two transitions.

Solution:

UV light consists of relatively high frequency (short wavelength) photons. So the energy of the absorbed photon and the energy transition (ΔE) in the atom is relatively large. In comparison, visible light consists of relatively lower-frequency photons. Therefore, the energy transition in the atom and the energy of the emitted photon is relatively small.

Exercise:

Problem:

Why are X-rays emitted only for electron transitions to inner shells? What type of photon is emitted for transitions between outer shells?

Exercise:**Problem:**

How do the allowed orbits for electrons in atoms differ from the allowed orbits for planets around the sun?

Solution:

For macroscopic systems, the quantum numbers are very large, so the energy difference (ΔE) between adjacent energy levels (orbits) is very small. The energy released in transitions between these closely spaced energy levels is much too small to be detected.

Problems**Exercise:****Problem:**

What is the minimum frequency of a photon required to ionize: (a) a He^+ ion in its ground state? (b) A Li^{2+} ion in its first excited state?

Solution:

For He^+ , one electron “orbits” a nucleus with two protons and two neutrons ($Z = 2$). Ionization energy refers to the energy required to remove the electron from the atom. The energy needed to remove the electron in the ground state of He^+ ion to infinity is negative the value of the ground state energy, written:

$$E = -54.4 \text{ eV}.$$

Thus, the energy to ionize the electron is $+54.4 \text{ eV}$.

Similarly, the energy needed to remove an electron in the first excited state of Li^{2+} ion to infinity is negative the value of the first excited state energy, written:

$$E = -30.6 \text{ eV}.$$

The energy to ionize the electron is 30.6 eV .

Exercise:**Problem:**

The ion Li^{2+} makes an atomic transition from an $n = 4$ state to an $n = 2$ state. (a) What is the energy of the photon emitted during the transition? (b) What is the wavelength of the photon?

Exercise:

Problem:

The red light emitted by a ruby laser has a wavelength of 694.3 nm. What is the difference in energy between the initial state and final state corresponding to the emission of the light?

Solution:

The wavelength of the laser is given by:

$$\lambda = \frac{hc}{-\Delta E},$$

where E_γ is the energy of the photon and ΔE is the magnitude of the energy difference.

Solving for the latter, we get:

$$\Delta E = -2.795 \text{ eV}.$$

The negative sign indicates that the electron lost energy in the transition.

Exercise:**Problem:**

The yellow light from a sodium-vapor street lamp is produced by a transition of sodium atoms from a $3p$ state to a $3s$ state. If the difference in energies of those two states is 2.10 eV, what is the wavelength of the yellow light?

Exercise:

Problem: Estimate the wavelength of the K_α X-ray from calcium.

Solution:

$$\Delta E_{L \rightarrow K} \approx (Z - 1)^2 (10.2 \text{ eV}) = 3.68 \times 10^3 \text{ eV}.$$

Exercise:

Problem: Estimate the frequency of the K_α X-ray from cesium.

Exercise:**Problem:**

X-rays are produced by striking a target with a beam of electrons. Prior to striking the target, the electrons are accelerated by an electric field through a potential energy difference:

$$\Delta U = -e\Delta V,$$

where e is the charge of an electron and ΔV is the voltage difference. If $\Delta V = 15,000$ volts, what is the minimum wavelength of the emitted radiation?

Solution:

According to the conservation of the energy, the potential energy of the electron is converted completely into kinetic energy. The initial kinetic energy of the electron is zero (the electron begins at rest). So, the kinetic energy of the electron just before it strikes the target is:

$$K = e\Delta V.$$

If *all* of this energy is converted into braking radiation, the frequency of the emitted radiation is a maximum, therefore:

$$f_{\max} = \frac{e\Delta V}{h}.$$

When the emitted frequency is a maximum, then the emitted wavelength is a minimum, so:

$$\lambda_{\min} = 0.1293 \text{ nm}.$$

Exercise:**Problem:**

For the preceding problem, what happens to the minimum wavelength if the voltage across the X-ray tube is doubled?

Exercise:**Problem:**

Suppose the experiment in the preceding problem is conducted with muons. What happens to the minimum wavelength?

Solution:

A muon is 200 times heavier than an electron, but the minimum wavelength does not depend on mass, so the result is unchanged.

Exercise:**Problem:**

An X-ray tube accelerates an electron with an applied voltage of 50 kV toward a metal target. (a) What is the shortest-wavelength X-ray radiation generated at the target? (b) Calculate the photon energy in eV. (c) Explain the relationship of the photon energy to the applied voltage.

Exercise:**Problem:**

A color television tube generates some X-rays when its electron beam strikes the screen. What is the shortest wavelength of these X-rays, if a 30.0-kV potential is used to accelerate the electrons? (Note that TVs have shielding to prevent these X-rays from exposing viewers.)

Solution:

$$4.13 \times 10^{-11} \text{ m}$$

Exercise:**Problem:**

An X-ray tube has an applied voltage of 100 kV. (a) What is the most energetic X-ray photon it can produce? Express your answer in electron volts and joules. (b) Find the wavelength of such an X-ray.

Exercise:**Problem:**

The maximum characteristic X-ray photon energy comes from the capture of a free electron into a K shell vacancy. What is this photon energy in keV for tungsten, assuming that the free electron has no initial kinetic energy?

Solution:

$$72.5 \text{ keV}$$

Exercise:

Problem: What are the approximate energies of the K_α and K_β X-rays for copper?

Exercise:

Problem: Compare the X-ray photon wavelengths for copper and gold.

Solution:

The atomic numbers for Cu and Au are $Z = 29$ and 79 , respectively. The X-ray photon frequency for gold is greater than copper by a factor:

$$\left(\frac{f_{\text{Au}}}{f_{\text{Cu}}}\right)^2 = \left(\frac{79-1}{29-1}\right)^2 \approx 8.$$

Therefore, the X-ray wavelength of Au is about eight times shorter than for copper.

Exercise:**Problem:**

The approximate energies of the K_α and K_β X-rays for copper are $E_{K_\alpha} = 8.00 \text{ keV}$ and $E_{K_\beta} = 9.48 \text{ keV}$, respectively. Determine the ratio of X-ray frequencies of gold to copper, then use this value to estimate the corresponding energies of K_α and K_β X-rays for gold.

Glossary

braking radiation

radiation produced by targeting metal with a high-energy electron beam (or radiation produced by the acceleration of any charged particle in a material)

fluorescence

radiation produced by the excitation and subsequent, gradual de-excitation of an electron in an atom

Moseley's law

relationship between the atomic number and X-ray photon frequency for X-ray production

Moseley plot

plot of the atomic number versus the square root of X-ray frequency

selection rules

rules that determine whether atomic transitions are allowed or forbidden (rare)

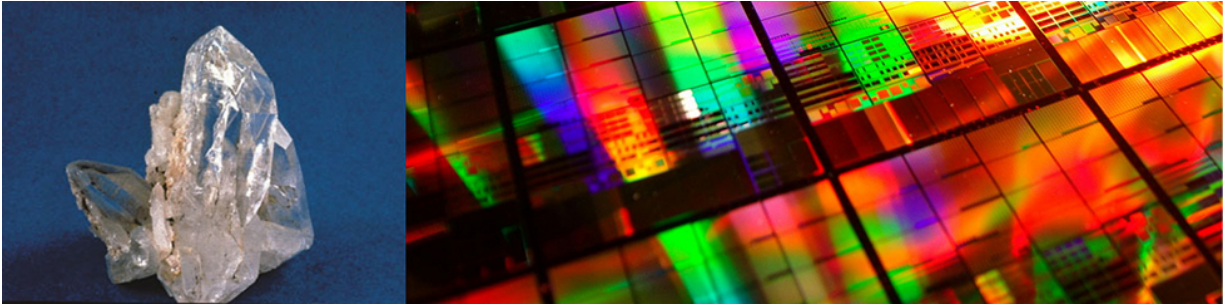
Introduction

class="introduction"

The
crystalline
structure of
quartz allows
it to cleave
into smooth
planes that
refract light,
making it
suitable for
jewelry.

Silicon, the
main element
in quartz, also
forms crystals
in its pure
form, and
these crystals
form the basis
for the
worldwide
semiconducto
r electronics
industry.

(credit left:
modification
of work by
the United
States
Geological
Survey)



In this chapter, we examine applications of quantum mechanics to more complex systems, such as molecules, metals, semiconductors, and superconductors. We review and develop concepts of the previous chapters, including wave functions, orbitals, and quantum states. We also introduce many new concepts, including covalent bonding, rotational energy levels, Fermi energy, energy bands, doping, and Cooper pairs.

The main topic in this chapter is the crystal structure of solids. For centuries, crystalline solids have been prized for their beauty, including gems like diamonds and emeralds, as well as geological crystals of quartz and metallic ores. But the crystalline structures of semiconductors such as silicon have also made possible the electronics industry of today. In this chapter, we study how the structures of solids give them properties from strength and transparency to electrical conductivity.

Types of Molecular Bonds

By the end of this section, you will be able to:

- Distinguish between the different types of molecular bonds
- Determine the dissociation energy of a molecule using the concepts ionization energy, electron affinity, and Coulomb force
- Describe covalent bonding in terms of exchange symmetry
- Explain the physical structure of a molecule in terms of the concept of hybridization

Quantum mechanics has been extraordinarily successful at explaining the structure and bonding in molecules, and is therefore the foundation for all of chemistry. Quantum chemistry, as it is sometimes called, explains such basic questions as why H_2O molecules exist, why the bonding angle between hydrogen atoms in this molecule is precisely 104.5° , and why these molecules bind together to form liquid water at room temperature. Applying quantum mechanics to molecules can be very difficult mathematically, so our discussion will be qualitative only.

As we study molecules and then solids, we will use many different scientific models. In some cases, we look at a molecule or crystal as a set of point nuclei with electrons whizzing around the outside in well-defined trajectories, as in the Bohr model. In other cases, we employ our full knowledge of quantum mechanics to study these systems using wave functions and the concept of electron spin. It is important to remember that we study modern physics with models, and that different models are useful for different purposes. We do not always use the most powerful model, when a less-powerful, easier-to-use model will do the job.

Types of Bonds

Chemical units form by many different kinds of chemical bonds. An **ionic bond** forms when an electron transfers from one atom to another. A **covalent bond** occurs when two or more atoms share electrons. A **van der Waals bond** occurs due to the attraction of charge-polarized molecules and is considerably weaker than ionic or covalent bonds. Many other types of bonding exist as well. Often, bonding occurs via more than one mechanism. The focus of this section is ionic and covalent bonding.

Ionic bonds

The ionic bond is perhaps the easiest type of bonding to understand. It explains the formation of salt compounds, such as sodium chloride, NaCl. The sodium atom (symbol Na) has the same electron arrangement as a neon atom plus one 3s electron. Only 5.14 eV of energy is required to remove this one electron from the sodium atom. Therefore, Na can easily give up or donate this electron to an adjacent (nearby) atom, attaining a more stable arrangement of electrons. Chlorine (symbol Cl) requires just one electron to complete its valence shell, so it readily accepts this electron if it is near the sodium atom. We therefore say that chlorine has a large **electron affinity**, which is the energy associated with an accepted electron. The energy given up by the chlorine atom in this process is 3.62 eV. After the electron transfers from the sodium atom to the chlorine atom, the sodium atom becomes a positive ion and the chlorine atom becomes a negative ion. The total energy required for this transfer is given by

Equation:

$$E_{\text{transfer}} = 5.14 \text{ eV} - 3.62 \text{ eV} = 1.52 \text{ eV}.$$

The positive sodium ion and negative chloride ion experience an attractive Coulomb force. The potential energy associated with this force is given by

Note:

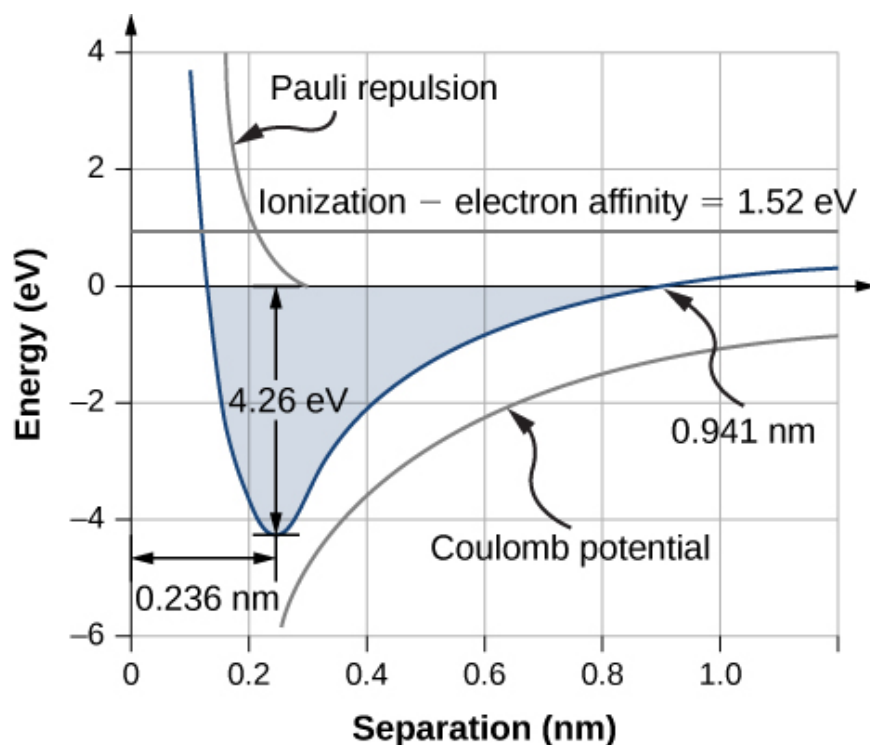
Equation:

$$U_{\text{coul}} = -\frac{ke^2}{r_0},$$

where $ke^2 = 1.440 \text{ eV}\cdot\text{nm}$ and r_0 is the distance between the ions.

As the sodium and chloride ions move together (“descend the potential energy hill”), the force of attraction between the ions becomes stronger. However, if the ions become too close, core-electron wave functions in the two ions begin to overlap. Due to the exclusion principle, this action promotes the core electrons—and therefore the entire molecule—into a higher energy state. The **equilibrium**

separation distance (or bond length) between the ions occurs when the molecule is in its lowest energy state. For diatomic NaCl, this distance is 0.236 nm. [\[link\]](#) shows the total energy of NaCl as a function of the distance of separation between ions.



Graph of energy versus ionic separation for sodium chloride. Equilibrium separation occurs when the total energy is a minimum (-4.36 eV).

The total energy required to form a single salt unit is

Note:

Equation:

$$U_{\text{form}} = E_{\text{transfer}} + U_{\text{coul}} + U_{\text{ex}},$$

where U_{ex} is the energy associated with the repulsion between core electrons due to Pauli's exclusion principle. The value of U_{form} must be negative for the bond to form spontaneously. The **dissociation energy** is defined as the energy required to separate the unit into its constituent ions, written

Equation:

$$U_{\text{diss}} = -U_{\text{form}}$$

Every diatomic formula unit has its own characteristic dissociation energy and equilibrium separation length. Sample values are given in [\[link\]](#).

Molecule	Dissociation Energy(eV)	Equilibrium Separation (nm) (Bond length)
NaCl	4.26	0.236
NaF	4.99	0.193
NaBr	3.8	0.250
NaI	3.1	0.271
NaH	2.08	0.189
LiCl	4.86	0.202
LiH	2.47	0.239
LiI	3.67	0.238
KCl	4.43	0.267
KBr	3.97	0.282

Molecule	Dissociation Energy(eV)	Equilibrium Separation (nm) (Bond length)
RbF	5.12	0.227
RbCl	4.64	0.279
CsI	3.57	0.337
H-H	4.5	0.075
N-N	9.8	0.11
O-O	5.2	0.12
F-F	1.6	0.14
Cl-Cl	2.5	0.20

Bond Length

Example:

The Energy of Salt

What is the dissociation energy of a salt formula unit (NaCl)?

Strategy

Sodium chloride (NaCl) is a salt formed by ionic bonds. The energy change associated with this bond depends on three main processes: the ionization of Na; the acceptance of the electron from a Na atom by a Cl atom; and Coulomb attraction of the resulting ions (Na^+ and Cl^-). If the ions get too close, they repel due to the exclusion principle (0.32 eV). The equilibrium separation distance is $r_0 = 0.236 \text{ nm}$.

Solution

The energy change associated with the transfer of an electron from Na to Cl is 1.52 eV, as discussed earlier in this section. At equilibrium separation, the atoms are $r_0 = 0.236 \text{ nm}$ apart. The electrostatic potential energy of the atoms is

Equation:

$$U_{\text{coul}} = -\frac{ke^2}{r_0} = -\frac{1.44 \text{ eV} \cdot \text{nm}}{0.236 \text{ nm}} = -6.10 \text{ eV}.$$

The total energy difference associated with the formation of a NaCl formula unit is

Equation:

$$E_{\text{form}} = E_{\text{xfr}} + U_{\text{coul}} + U_{\text{ex}} = 1.52 \text{ eV} + (-6.10 \text{ eV}) + 0.32 \text{ eV} = -4.26 \text{ eV}.$$

Therefore, the dissociated energy of NaCl is 4.26 eV.

Significance

The formation of a NaCl formula unit by ionic bonding is energetically favorable. The dissociation energy, or energy required to separate the NaCl unit into Na^+ and Cl^- ions is 4.26 eV, consistent with [\[link\]](#).

Note:

Exercise:

Problem:

Check Your Understanding Why is the potential energy associated with the exclusion principle positive in [\[link\]](#)?

Solution:

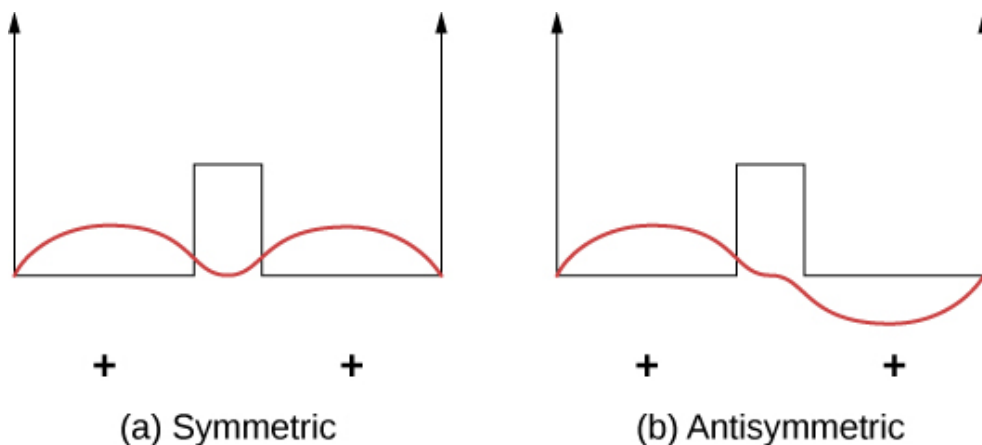
It corresponds to a repulsive force between core electrons in the ions.

For a sodium ion in an ionic NaCl crystal, the expression for Coulomb potential energy U_{coul} must be modified by a factor known as the **Madelung constant**. This factor takes into account the interaction of the sodium ion with all nearby chloride and sodium ions. The Madelung constant for a NaCl crystal is about 1.75. This value implies an equilibrium separation distance between Na^+ and Cl^- ions of 0.280 nm—slightly larger than for diatomic NaCl. We will return to this point again later.

Covalent bonds

In an ionic bond, an electron transfers from one atom to another. However, in a covalent bond, an electron is shared between two atoms. The ionic bonding mechanism cannot explain the existence of such molecules as H_2 , O_2 , and CO , since no separation distance exists for which the negative potential energy of attraction is greater in magnitude than the energy needed to create ions. Understanding precisely how such molecules are covalently bonded relies on a deeper understanding of quantum mechanics that goes beyond the coverage of this book, but we will qualitatively describe the mechanisms in the following section.

Covalent bonds can be understood using the simple example of a H_2^+ molecule, which consists of one electron in the electric field of two protons. This system can be modeled by an electron in a double square well ([link](#)). The electron is equally likely to be found in each well, so the wave function is either symmetric or antisymmetric about a point midway between the wells.



A one-dimensional model of covalent bonding in a H_2^+ molecule. (a) The symmetric wave function of the electron shared by the two positively charged protons (represented by the two finite square wells). (b) The corresponding antisymmetric wave function.

Now imagine that the two wells are separated by a large distance. In the ground state, the wave function exists in one of two possible states: either a single positive peak (a sine wave-like “hump”) in both wells (symmetric case), or a positive peak in one well and a negative peak in the other (antisymmetric case). These states have the same energy. However, when the wells are brought together, the symmetric wave function becomes the ground state and the antisymmetric state becomes the first excited state—in other words, the energy level of the electron is split. Notice, the space-symmetric state becomes the energetically favorable (lower energy) state.

The same analysis is appropriate for an electron bound to two hydrogen atoms. Here, the shapes of the ground-state wave functions have the form e^{-r/a_0} or $e^{(-|x|/a_0)}$ in one dimension. The energetically favorable, space-symmetric state implies a high charge density midway between the protons where the electrons are likely to pull the positively charged protons together.

If a second electron is added to this system to form a H_2 molecule, the wave function must describe both particles, including their spatial relationship and relative spins. This wave function must also respect the indistinguishability of electrons. (“If you’ve seen one electron, you’ve seen them all.”) In particular, switching or exchanging the electrons should *not* produce an observable effect, a property called **exchange symmetry**. Exchange symmetry can be *symmetric*, producing no change in the wave function, or *antisymmetric*, producing an overall change in the sign of the wave function—neither of which is observable.

As we discuss later, the total wave function of two electrons must be antisymmetric on exchange. For example, two electrons bound to a hydrogen molecule can be in a space-symmetric state with antiparallel spins ($\uparrow\downarrow$) or space-antisymmetric state with parallel spins ($\uparrow\uparrow$). The state with antiparallel spins is energetically favorable and therefore used in covalent bonding. If the protons are drawn too closely together, however, repulsion between the protons becomes important. (In other molecules, this effect is supplied by the exclusion principle.) As a result, H_2 reaches an equilibrium separation of about 0.074 nm with a binding energy is 4.52 eV.

Note:

Visit this [PBS Learning Media tutorial and interactive simulation](#) to explore the attractive and repulsive forces that act on atomic particles and covalent bonding

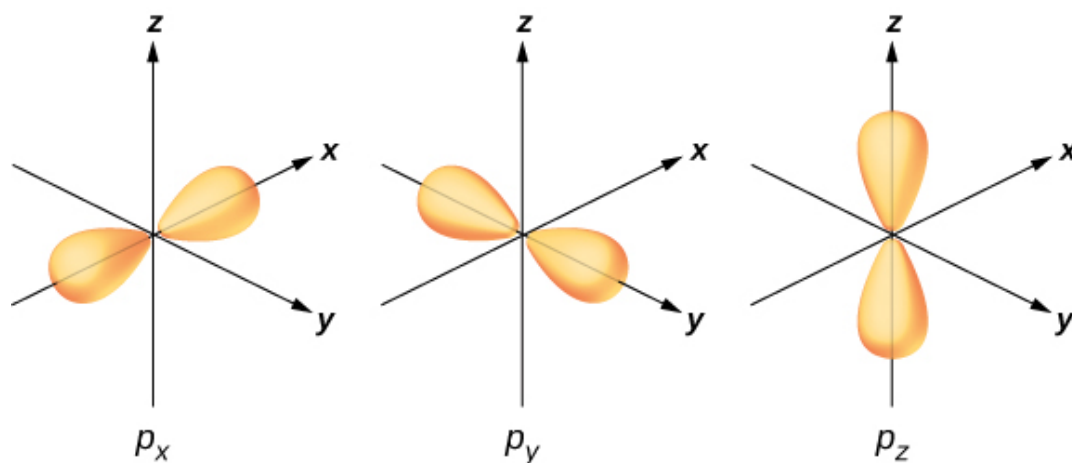
in a H_2 molecule.

Quantum mechanics excludes many types of molecules. For example, the molecule H_3 does not form, because if a third H atom approaches diatomic hydrogen, the wave function of the electron in this atom overlaps the electrons in the other two atoms. If all three electrons are in the ground states of their respective atoms, one pair of electrons shares all the same quantum numbers, which is forbidden by the exclusion principle. Instead, one of the electrons is forced into a higher energy state. No separation between three protons exists for which the total energy change of this process is negative—that is, where bonding occurs spontaneously. Similarly, He_2 is not covalently bonded under normal conditions, because these atoms have no valence electrons to share. As the atoms are brought together, the wave functions of the core electrons overlap, and due to the exclusion principle, the electrons are forced into a higher energy state. No separation exists for which such a molecule is energetically favorable.

Bonding in Polyatomic Molecules

A **polyatomic molecule** is a molecule made of more than two atoms. Examples range from a simple water molecule to a complex protein molecule. The structures of these molecules can often be understood in terms of covalent bonding and **hybridization**. Hybridization is a change in the energy structure of an atom in which mixed states (states that can be written as a linear superposition of others) participate in bonding.

To illustrate hybridization, consider the bonding in a simple water molecule, H_2O . The electron configuration of oxygen is $1s^2 2s^2 2p^4$. The 1s and 2s electrons are in “closed shells” and do not participate in bonding. The remaining four electrons are the valence electrons. These electrons can fill six possible states ($l = 1$, $m = 0, \pm 1$, plus spin up and down). The energies of these states are the same, so the oxygen atom can exploit any linear combination of these states in bonding with the hydrogen atoms. These linear combinations (which you learned about in the chapter on atomic structure) are called atomic orbitals, and they are denoted by p_x , p_y , and p_z . The electron charge distributions for these orbitals are given in [\[link\]](#).



Oxygen has four valence electrons. In the context of a water molecule, two valence electrons fill the p_z orbital and one electron fills each of the p_x and p_y orbitals. The p_x and p_y orbitals are used in bonding with hydrogen atoms to form H_2O . Without repulsion of H atoms, the bond angle between hydrogen atoms would be 90 degrees.

The transformation of the electron wave functions of oxygen to p_x , p_y , and p_z orbitals in the presence of the hydrogen atoms is an example of hybridization. Two electrons are found in the p_z orbital with paired spins (). One electron is found in each of the p_x and p_y orbitals, with unpaired spins. The latter orbitals participate in bonding with the hydrogen atoms. Based on [\[link\]](#), we expect the bonding angle for H—O—H to be 90° . However, if we include the effects of repulsion between atoms, the bond angle is 104.5° . The same arguments can be used to understand the tetrahedral shape of methane (CH_4) and other molecules.

Summary

- Molecules form by two main types of bonds: the ionic bond and the covalent bond. An ionic bond transfers an electron from one atom to another, and a covalent bond shares the electrons.
- The energy change associated with ionic bonding depends on three main processes: the ionization of an electron from one atom, the acceptance of the electron by the second atom, and the Coulomb attraction of the resulting ions.

- Covalent bonds involve space-symmetric wave functions.
- Atoms use a linear combination of wave functions in bonding with other molecules (hybridization).

Conceptual Questions

Exercise:

Problem:

What is the main difference between an *ionic bond*, a *covalent bond*, and a *van der Waals bond*?

Solution:

An ionic bond is formed by the attraction of a positive and negative ion. A covalent bond is formed by the sharing of one or more electrons between atoms. A van der Waals bond is formed by the attraction of two electrically polarized molecules.

Exercise:

Problem:

For the following cases, what type of bonding is expected? (a) KCl molecule; (b) N₂ molecule.

Exercise:

Problem: Describe three steps to ionic bonding.

Solution:

1. An electron is removed from one atom. The resulting atom is a positive ion. 2. An electron is absorbed from another atom. The result atom is a negative ion. 3. The positive and negative ions are attracted together until an equilibrium separation is reached.

Exercise:

Problem:

What prevents a positive and negative ion from having a zero separation?

Exercise:**Problem:**

For the H_2 molecule, why must the spins the electron spins be antiparallel?

Solution:

Bonding is associated with a spatial function that is symmetric under exchange of the two electrons. In this state, the electron density is largest between the atoms. The total function must be antisymmetric (since electrons are fermions), so the spin function must be antisymmetric. In this state, the spins of the electrons are antiparallel.

Problems**Exercise:****Problem:**

The electron configuration of carbon is $1s^2 2s^2 2p^2$. Given this electron configuration, what other element might exhibit the same type of hybridization as carbon?

Exercise:**Problem:**

Potassium chloride (KCl) is a molecule formed by an ionic bond. At equilibrium separation the atoms are $r_0 = 0.279$ nm apart. Determine the electrostatic potential energy of the atoms.

Solution:

$$U = -5.16 \text{ eV}$$

Exercise:**Problem:**

The electron affinity of Cl is 3.89 eV and the ionization energy of K is 4.34 eV. Use the preceding problem to find the dissociation energy. (Neglect the energy of repulsion.)

Exercise:**Problem:**

The measured energy dissociated energy of KCl is 4.43 eV. Use the results of the preceding problem to determine the energy of repulsion of the ions due to the exclusion principle.

Solution:

$$-4.43 \text{ eV} = -4.69 \text{ eV} + U_{\text{ex}}, U_{\text{ex}} = 0.26 \text{ eV}$$

Glossary

covalent bond

bond formed by the sharing of one or more electrons between atoms

dissociation energy

amount of energy needed to break apart a molecule into atoms; also, total energy per ion pair to separate the crystal into isolated ions

electron affinity

energy associated with an accepted (bound) electron

equilibrium separation distance

distance between atoms in a molecule

exchange symmetry

how a total wave function changes under the exchange of two electrons

hybridization

change in the energy structure of an atom in which energetically favorable mixed states participate in bonding

ionic bond

bond formed by the Coulomb attraction of a positive and negative ions

Madelung constant

constant that depends on the geometry of a crystal used to determine the total potential energy of an ion in a crystal

polyatomic molecule

molecule formed of more than one atom

van der Waals bond

bond formed by the attraction of two electrically polarized molecules

Molecular Spectra

By the end of this section, you will be able to:

- Use the concepts of vibrational and rotational energy to describe energy transitions in a diatomic molecule
- Explain key features of a vibrational-rotational energy spectrum of a diatomic molecule
- Estimate allowed energies of a rotating molecule
- Determine the equilibrium separation distance between atoms in a diatomic molecule from the vibrational-rotational absorption spectrum

Molecular energy levels are more complicated than atomic energy levels because molecules can also vibrate and rotate. The energies associated with such motions lie in different ranges and can therefore be studied separately. Electronic transitions are of order 1 eV, vibrational transitions are of order 10^{-2} eV, and rotational transitions are of order 10^{-3} eV. For complex molecules, these energy changes are difficult to characterize, so we begin with the simple case of a diatomic molecule.

According to classical mechanics, the energy of rotation of a diatomic molecule is given by **Equation:**

$$E_r = \frac{L^2}{2I},$$

where I is the moment of inertia and L is the angular momentum. According to quantum mechanics, the rotational angular momentum is quantized:

Equation:

$$L = \sqrt{l(l+1)}\hbar \quad (l = 0, 1, 2, 3, \dots),$$

where l is the orbital angular quantum number. The allowed **rotational energy level** of a diatomic molecule is therefore

Note:

Equation:

$$E_r = l(l+1)\frac{\hbar^2}{2I} = l(l+1)E_{0r} \quad (l = 0, 1, 2, 3, \dots),$$

where the characteristic rotational energy of a molecule is defined as

Note:
Equation:

$$E_{0r} = \frac{\hbar^2}{2I}.$$

For a diatomic molecule, the moment of inertia with reduced mass μ is

Note:
Equation:

$$I = \mu r_0^2,$$

where r_0 is the total distance between the atoms. The energy difference between rotational levels is therefore

Equation:

$$\Delta E_r = E_{l+1} - E_l = 2(l+1) E_{0r}.$$

A detailed study of transitions between rotational energy levels brought about by the absorption or emission of radiation (a so-called **electric dipole transition**) requires that

Note:
Equation:

$$\Delta l = \pm 1.$$

This rule, known as a **selection rule**, limits the possible transitions from one quantum state to another. [\[link\]](#) is the selection rule for rotational energy transitions. It applies only to diatomic molecules that have an electric dipole moment. For this reason, symmetric molecules such as H_2 and N_2 do not experience rotational energy transitions due to the absorption or emission of electromagnetic radiation.

Example:

The Rotational Energy of HCl

Determine the lowest three rotational energy levels of a hydrogen chloride (HCl) molecule.

Strategy

Hydrogen chloride (HCl) is a diatomic molecule with an equilibrium separation distance of 0.127 nm. Rotational energy levels depend only on the momentum of inertia I and the orbital angular momentum quantum number l (in this case, $l = 0, 1$, and 2). The momentum of inertia depends, in turn, on the equilibrium separation distance (which is given) and the reduced mass, which depends on the masses of the H and Cl atoms.

Solution

First, we compute the reduced mass. If Particle 1 is hydrogen and Particle 2 is chloride, we have

Equation:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} = \frac{(1.0 \text{ u})(35.4 \text{ u})}{1.0 \text{ u} + 35.4 \text{ u}} = 0.97 \text{ u} = 0.97 \text{ u} \left(\frac{931.5 \frac{\text{MeV}}{c^2}}{1 \text{ u}} \right) = 906 \frac{\text{MeV}}{c^2}.$$

The corresponding rest mass energy is therefore

Equation:

$$\mu c^2 = 9.06 \times 10^8 \text{ eV}.$$

This allows us to calculate the characteristic energy:

Equation:

$$E_{0r} = \frac{\hbar^2}{2I} = \frac{\hbar^2}{2(\mu r_0^2)} = \frac{(\hbar c)^2}{2(\mu c^2)r_0^2} = \frac{(197.3 \text{ eV} \cdot \text{nm})^2}{2(9.06 \times 10^8 \text{ eV})(0.127 \text{ nm})^2} = 1.33 \times 10^{-3} \text{ eV}.$$

(Notice how this expression is written in terms of the rest mass energy. This technique is common in modern physics calculations.) The rotational energy levels are given by

Equation:

$$E_r = l(l+1) \frac{\hbar^2}{2I} = l(l+1) E_{0r},$$

where l is the orbital quantum number. The three lowest rotational energy levels of an HCl molecule are therefore

Equation:

$$l = 0; E_r = 0 \text{ eV (no rotation)},$$

Equation:

$$l = 1; E_r = 2 E_{0r} = 2.66 \times 10^{-3} \text{ eV},$$

Equation:

$$l = 2; E_r = 6 E_{0r} = 7.99 \times 10^{-3} \text{ eV}.$$

Significance

The rotational spectrum is associated with weak transitions (1/1000 to 1/100 of an eV). By comparison, the energy of an electron in the ground state of hydrogen is -13.6 eV.

Note:

Exercise:

Problem:

Check Your Understanding What does the energy separation between absorption lines in a rotational spectrum of a diatomic molecule tell you?

Solution:

the moment of inertia

The **vibrational energy level**, which is the energy level associated with the vibrational energy of a molecule, is more difficult to estimate than the rotational energy level. However, we can estimate these levels by assuming that the two atoms in the diatomic molecule are connected by an ideal spring of spring constant k . The potential energy of this spring system is

Equation:

$$U_{\text{osc}} = \frac{1}{2}k \Delta r^2,$$

Where Δr is a change in the “natural length” of the molecule along a line that connects the atoms. Solving Schrödinger’s equation for this potential gives

Equation:

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \quad (n = 0, 1, 2, \dots),$$

Where ω is the natural angular frequency of vibration and n is the vibrational quantum number. The prediction that vibrational energy levels are evenly spaced ($\Delta E = \hbar\omega$) turns out to be good at lower energies.

A detailed study of transitions between vibrational energy levels induced by the absorption or emission of radiation (and the specifically so-called electric dipole transition) requires that

Note:

Equation:

$$\Delta n = \pm 1.$$

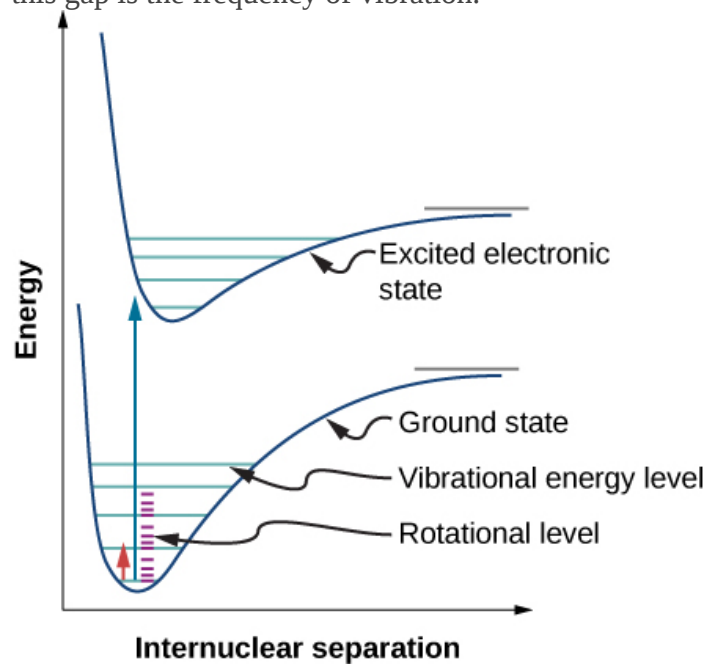
[\[link\]](#) represents the selection rule for vibrational energy transitions. As mentioned before, this rule applies only to diatomic molecules that have an electric dipole moment. Symmetric molecules do not experience such transitions.

Due to the selection rules, the absorption or emission of radiation by a diatomic molecule involves a transition in vibrational and rotational states. Specifically, if the vibrational quantum number (n) changes by one unit, then the rotational quantum number (l) changes by one unit. An energy-level diagram of a possible transition is given in [\[link\]](#). The absorption spectrum for such transitions in hydrogen chloride (HCl) is shown in [\[link\]](#). The absorption peaks are due to transitions from the $n = 0$ to $n = 1$ vibrational states. Energy differences for the band of peaks at the left and right are, respectively,

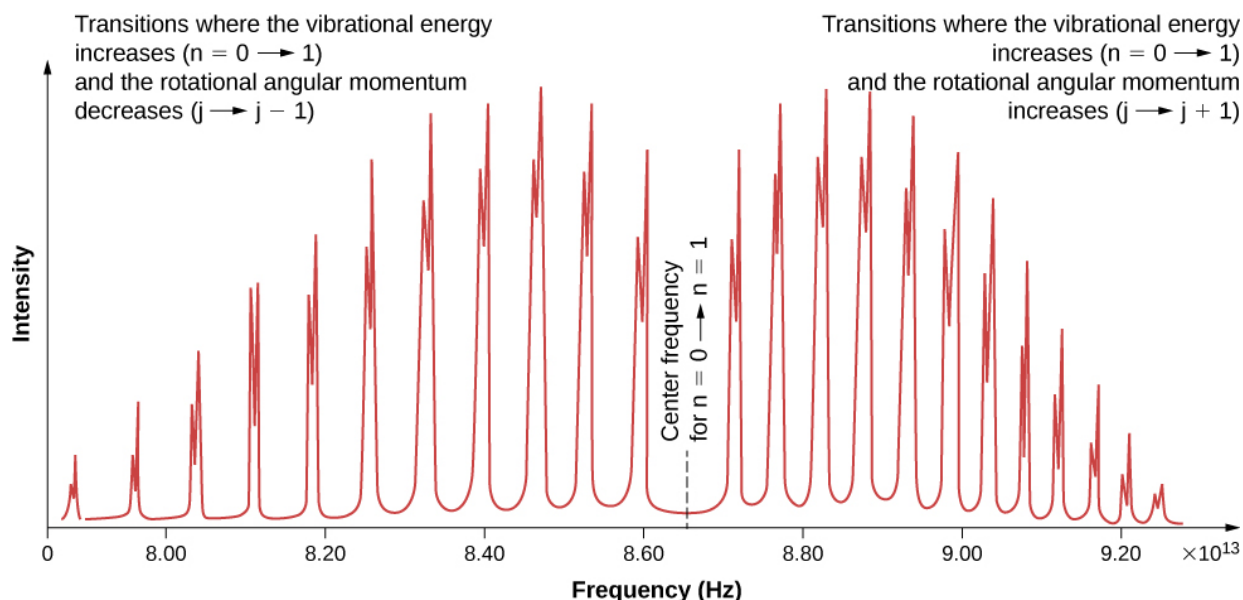
$$\Delta E_{l \rightarrow l+1} = \hbar\omega + 2(l+1)E_{0r} = \hbar\omega + 2E_{0r}, \hbar\omega + 4E_{0r}, \hbar\omega + 6E_{0r}, \dots \text{ (right band) and}$$

$$\Delta E_{l \rightarrow l-1} = \hbar\omega - 2lE_{0r} = \hbar\omega - 2E_{0r}, \hbar\omega - 4E_{0r}, \hbar\omega - 6E_{0r}, \dots \text{ (left band).}$$

The moment of inertia can then be determined from the energy spacing between individual peaks ($2E_{0r}$) or from the gap between the left and right bands ($4E_{0r}$). The frequency at the center of this gap is the frequency of vibration.



Three types of energy levels in a diatomic molecule: electronic, vibrational, and rotational. If the vibrational quantum number (n) changes by one unit, then the rotational quantum number (l) changes by one unit.



Absorption spectrum of hydrogen chloride (HCl) from the $n = 0$ to $n = 1$ vibrational levels. The discrete peaks indicate a quantization of the angular momentum of the molecule. The bands to the left indicate a decrease in angular momentum, whereas those to the right indicate an increase in angular momentum.

Summary

- Molecules possess vibrational and rotational energy.
- Energy differences between adjacent vibrational energy levels are larger than those between rotational energy levels.
- Separation between peaks in an absorption spectrum is inversely related to the moment of inertia.
- Transitions between vibrational and rotational energy levels follow selection rules.

Conceptual Questions

Exercise:

Problem:

Does the absorption spectrum of the diatomic molecule HCl depend on the isotope of chlorine contained in the molecule? Explain your reasoning.

Exercise:

Problem:

Rank the energy spacing (ΔE) of the following transitions from least to greatest: an electron energy transition in an atom (atomic energy), the rotational energy of a molecule, or the vibrational energy of a molecule?

Solution:

rotational energy, vibrational energy, and atomic energy

Exercise:**Problem:**

Explain key features of a vibrational-rotation energy spectrum of the diatomic molecule.

Problems**Exercise:****Problem:**

In a physics lab, you measure the vibrational-rotational spectrum of HCl. The estimated separation between absorption peaks is $\Delta f \approx 5.5 \times 10^{11}$ Hz. The central frequency of the band is $f_0 = 9.0 \times 10^{13}$ Hz. (a) What is the moment of inertia (I)? (b) What is the energy of vibration for the molecule?

Exercise:**Problem:**

For the preceding problem, find the equilibrium separation of the H and Cl atoms. Compare this with the actual value.

Solution:

The measured value is 0.484 nm, and the actual value is close to 0.127 nm. The laboratory results are the same order of magnitude, but a factor 4 high.

Exercise:**Problem:**

The separation between oxygen atoms in an O_2 molecule is about 0.121 nm. Determine the characteristic energy of rotation in eV.

Exercise:**Problem:**

The characteristic energy of the N_2 molecule is 2.48×10^{-4} eV. Determine the separation distance between the nitrogen atoms

Solution:

0.110 nm

Exercise:**Problem:**

The characteristic energy for KCl is 1.4×10^{-5} eV. (a) Determine μ for the KCl molecule. (b) Find the separation distance between the K and Cl atoms.

Exercise:**Problem:**

A diatomic F_2 molecule is in the $l = 1$ state. (a) What is the energy of the molecule? (b) How much energy is radiated in a transition from a $l = 2$ to a $l = 1$ state?

Solution:

a. $E = 2.2 \times 10^{-4}$ eV; b. $\Delta E = 4.4 \times 10^{-4}$ eV

Exercise:**Problem:**

In a physics lab, you measure the vibrational-rotational spectrum of potassium bromide (KBr). The estimated separation between absorption peaks is $\Delta f \approx 5.35 \times 10^{10}$ Hz. The central frequency of the band is $f_0 = 8.75 \times 10^{12}$ Hz. (a) What is the moment of inertia (I)? (b) What is the energy of vibration for the molecule?

Glossary

electric dipole transition

transition between energy levels brought by the absorption or emission of radiation

rotational energy level

energy level associated with the rotational energy of a molecule

selection rule

rule that limits the possible transitions from one quantum state to another

vibrational energy level

energy level associated with the vibrational energy of a molecule

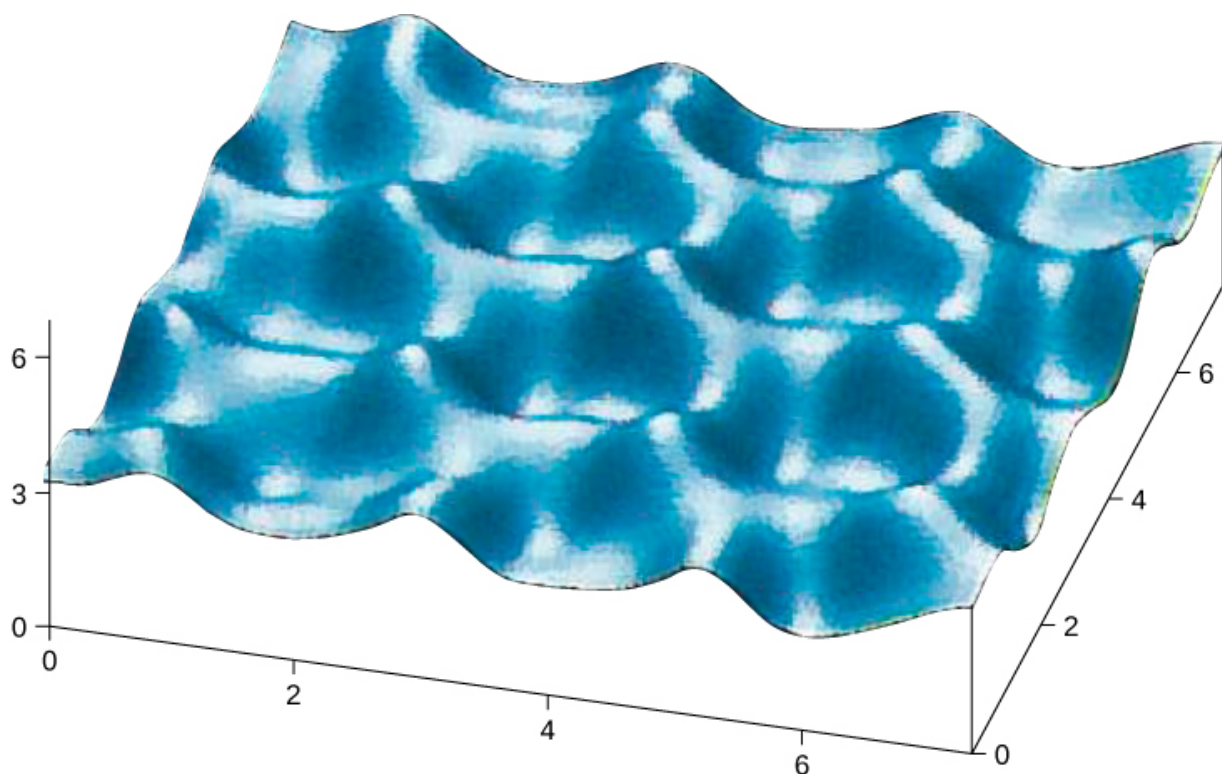
Bonding in Crystalline Solids

By the end of this section, you will be able to:

- Describe the packing structures of common solids
- Explain the difference between bonding in a solid and in a molecule
- Determine the equilibrium separation distance given crystal properties
- Determine the dissociation energy of a salt given crystal properties

Beginning in this section, we study crystalline solids, which consist of atoms arranged in an extended regular pattern called a **lattice**. Solids that do not or are unable to form crystals are classified as amorphous solids. Although amorphous solids (like glass) have a variety of interesting technological applications, the focus of this chapter will be on crystalline solids.

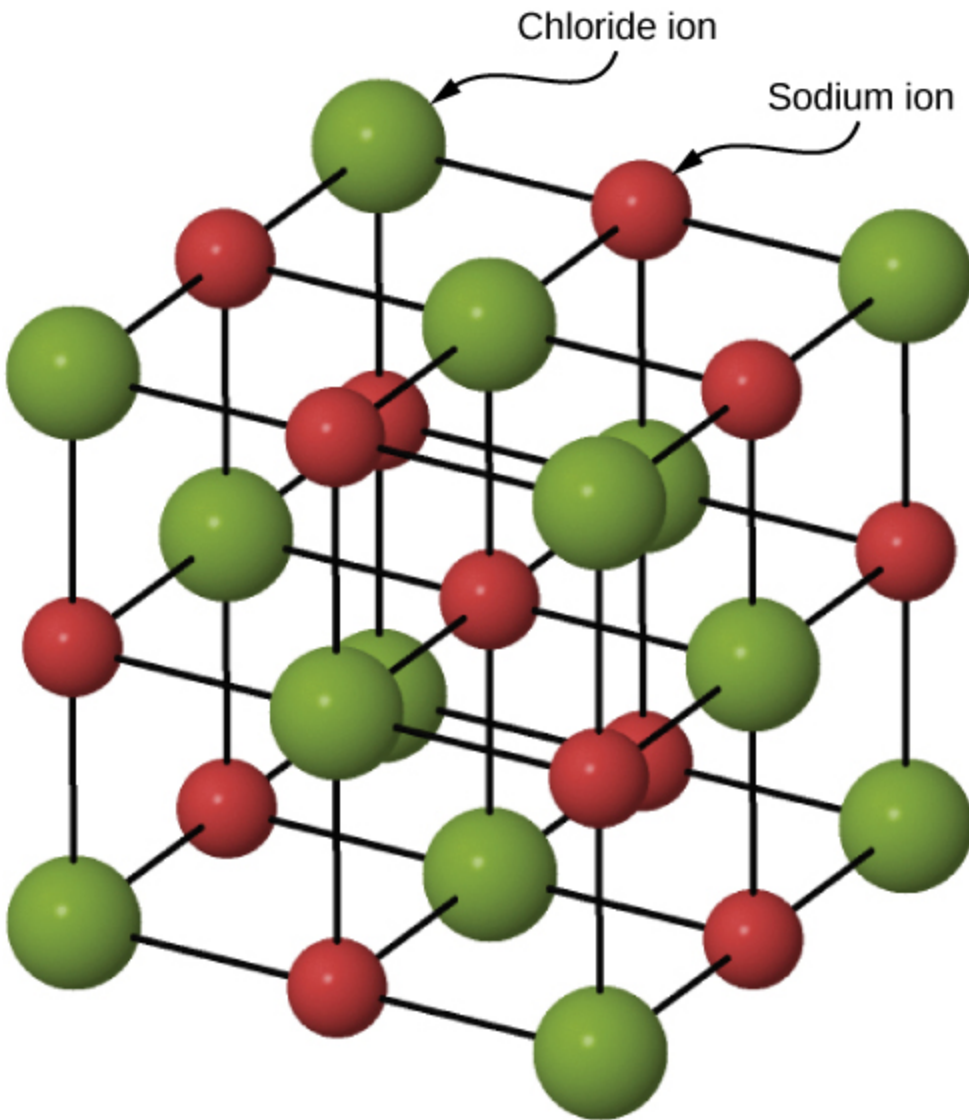
Atoms arrange themselves in a lattice to form a crystal because of a net attractive force between their constituent electrons and atomic nuclei. The crystals formed by the bonding of atoms belong to one of three categories, classified by their bonding: ionic, covalent, and metallic. Molecules can also bond together to form crystals; these bonds, not discussed here, are classified as molecular. Early in the twentieth century, the atomic model of a solid was speculative. We now have direct evidence of atoms in solids ([link](#)).



An image made with a scanning tunneling microscope of the surface of graphite. The peaks represent the atoms, which are arranged in hexagons. The scale is in angstroms.

Ionic Bonding in Solids

Many solids form by ionic bonding. A prototypical example is the sodium chloride crystal, as we discussed earlier. Electrons transfer from sodium atoms to adjacent chlorine atoms, since the valence electrons in sodium are loosely bound and chlorine has a large electron affinity. The positively charged sodium ions and negatively charged chlorine (chloride) ions organize into an extended regular array of atoms ([\[link\]](#)).



Structure of the sodium chloride crystal. The sodium and chloride ions are arranged in a face-centered cubic (FCC) structure.

The charge distributions of the sodium and chloride ions are spherically symmetric, and the chloride ion is about two times the diameter of the sodium ion. The lowest energy arrangement of these ions is called the **face-centered cubic (FCC)** structure. In this structure, each ion is closest to six ions of the other species. The unit cell is a cube—an atom occupies the

center and corners of each “face” of the cube. The attractive potential energy of the Na^+ ion due to the fields of these six Cl^- ions is written
Equation:

$$U_1 = -6 \frac{e^2}{4\pi\epsilon_0 r}$$

where the minus sign designates an attractive potential (and we identify $k = 1/4\pi\epsilon_0$). At a distance $\sqrt{2}r$ are its next-nearest neighbors: twelve Na^+ ions of the same charge. The total repulsive potential energy associated with these ions is

Equation:

$$U_2 = 12 \frac{e^2}{4\pi\epsilon_0 \sqrt{2}r}.$$

Next closest are eight Cl^- ions a distance $\sqrt{3}r$ from the Na^+ ion. The potential energy of the Na^+ ion in the field of these eight ions is

Equation:

$$U_3 = -8 \frac{e^2}{4\pi\epsilon_0 \sqrt{3}r}.$$

Continuing in the same manner with alternate sets of Cl^- and Na^+ ions, we find that the net attractive potential energy U_A of the single Na^+ ion can be written as

Equation:

$$U_{\text{coul}} = -\alpha \frac{e^2}{4\pi\epsilon_0 r}$$

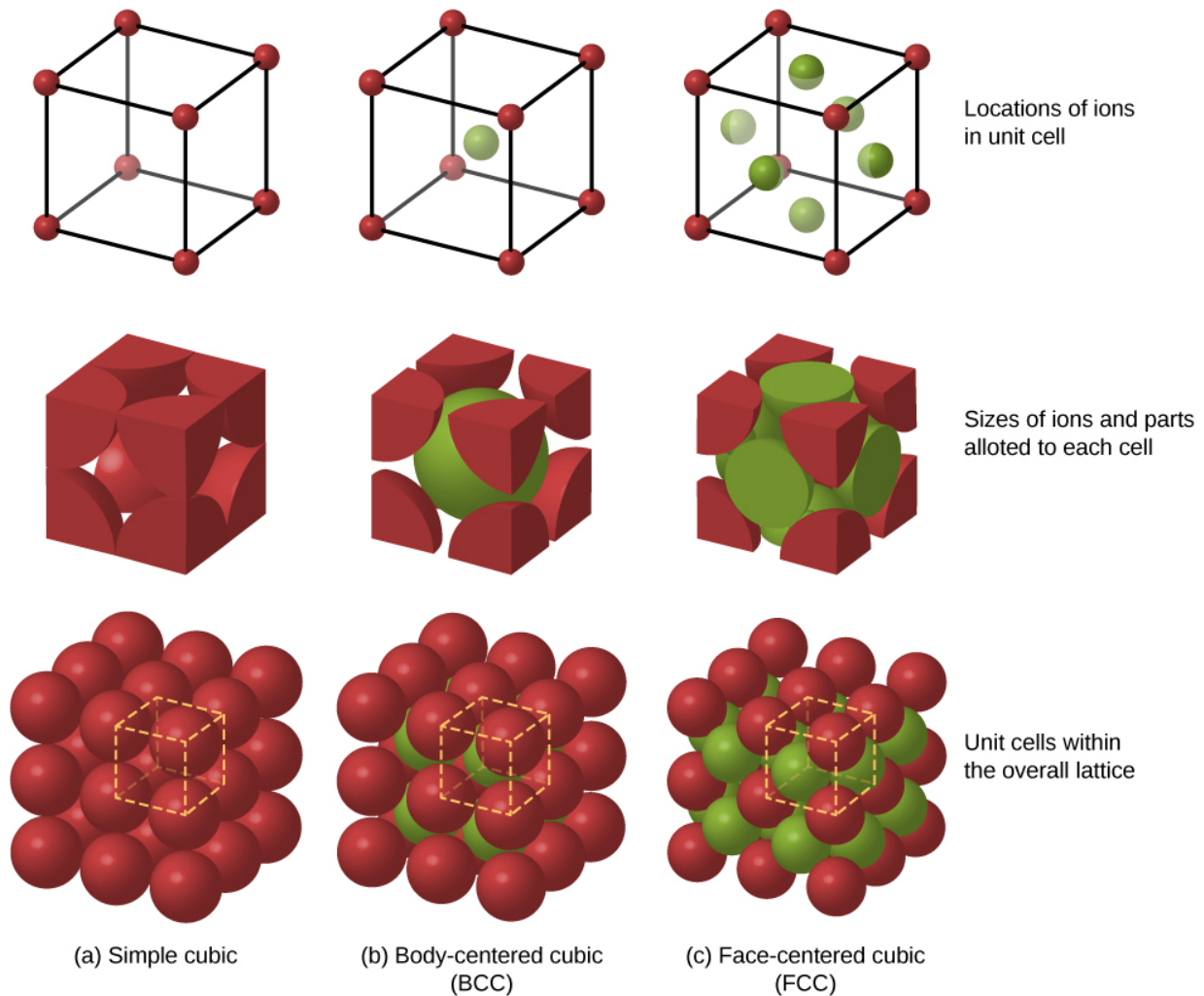
where α is the Madelung constant, introduced earlier. From this analysis, we can see that this constant is the infinite converging sum

Equation:

$$\alpha = 6 - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} + \dots$$

Distant ions make a significant contribution to this sum, so it converges slowly, and many terms must be used to calculate α accurately. For all FCC ionic solids, α is approximately 1.75.

Other possible packing arrangements of atoms in solids include **simple cubic** and **body-centered cubic (BCC)**. These three different packing structures of solids are compared in [\[link\]](#). The first row represents the location, but not the size, of the ions; the second row indicates the unit cells of each structure or lattice; and the third row represents the location and size of the ions. The BCC structure has eight nearest neighbors, with a Madelung constant of about 1.76—only slightly different from that for the FCC structure. Determining the Madelung constant for specific solids is difficult work and the subject of current research.



Packing structures for solids from left to right: (a) simple cubic, (b) body-centered cubic (BCC), and (c) face-centered cubic (FCC). Each crystal structure minimizes the energy of the system.

The energy of the sodium ions is not entirely due to attractive forces between oppositely charged ions. If the ions are brought too close together, the wave functions of core electrons of the ions overlap, and the electrons repel due to the exclusion principle. The total potential energy of the Na^+ ion is therefore the sum of the attractive Coulomb potential (U_{coul}) and the repulsive potential associated with the exclusion principle (U_{ex}). Calculating this repulsive potential requires powerful computers.

Fortunately, however, this energy can be described accurately by a simple formula that contains adjustable parameters:

Note:

Equation:

$$U_{\text{ex}} = \frac{A}{r^n}$$

where the parameters A and n are chosen to give predictions consistent with experimental data. For the problem at the end of this chapter, the parameter n is referred to as the **repulsion constant**. The total potential energy of the Na^+ ion is therefore

Equation:

$$U = -\alpha \frac{e^2}{4\pi\epsilon_0 r} + \frac{A}{r^n}.$$

At equilibrium, there is no net force on the ion, so the distance between neighboring Na^+ and Cl^- ions must be the value r_0 for which U is a minimum. Setting $\frac{dU}{dr} = 0$, we have

Equation:

$$0 = \frac{\alpha e^2}{4\pi\epsilon_0 r_0^2} - \frac{nA}{r_0^{n+1}}.$$

Thus,

Equation:

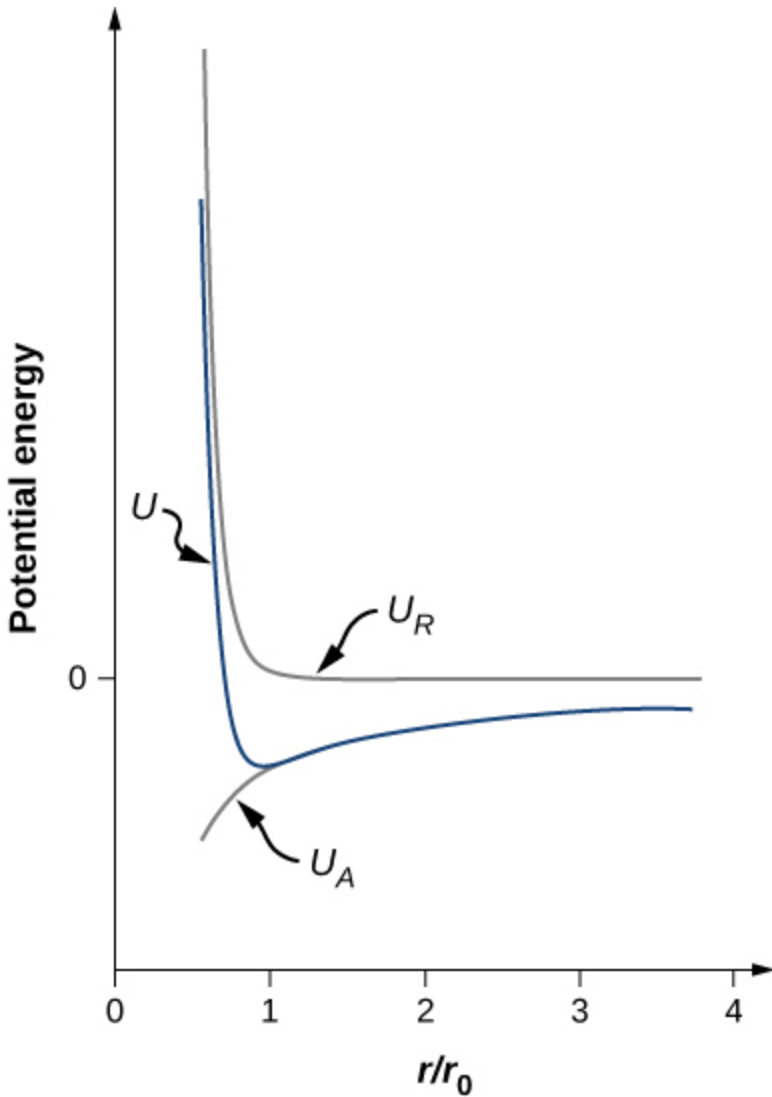
$$A = \frac{\alpha e^2 r_0^{n-1}}{4\pi\epsilon_0 n}.$$

Inserting this expression into the expression for the total potential energy, we have

Equation:

$$U = -\frac{\alpha e^2}{4\pi\epsilon_0 r_0} \left[\frac{r_0}{r} - \frac{1}{n} \left(\frac{r_0}{r} \right)^n \right].$$

Notice that the total potential energy now has only one adjustable parameter, n . The parameter A has been replaced by a function involving r_0 , the equilibrium separation distance, which can be measured by a diffraction experiment (you learned about diffraction in a previous chapter). The total potential energy is plotted in [\[link\]](#) for $n = 8$, the approximate value of n for NaCl.



The potential energy of a sodium ion in a NaCl crystal for $n = 8$. The equilibrium bond length occurs when the energy is a minimized.

As long as $n > 1$, the curve for U has the same general shape: U approaches infinity as $r \rightarrow 0$ and U approaches zero as $r \rightarrow \infty$. The minimum value of the potential energy is given by

Equation:

$$U_{\min}(r = r_0) = -\alpha \frac{ke^2}{r_0} \left(1 - \frac{1}{n}\right).$$

The energy per ion pair needed to separate the crystal into ions is therefore

Note:

Equation:

$$U_{\text{diss}} = \alpha \frac{ke^2}{r_0} \left(1 - \frac{1}{n}\right).$$

This is the **dissociation energy** of the solid. The dissociation energy can also be used to describe the total energy needed to break a mole of a solid into its constituent ions, often expressed in kJ/mole. The dissociation energy can be determined experimentally using the latent heat of vaporization. Sample values are given in the following table.

	F ⁻	Cl ⁻	Br ⁻	I ⁻
Li ⁺	1036	853	807	757
Na ⁺	923	787	747	704
K ⁺	821	715	682	649
Rb ⁺	785	689	660	630

Cs ⁺	740	659	631	604
-----------------	-----	-----	-----	-----

Lattice Energy for Alkali Metal Halides

Thus, we can determine the Madelung constant from the crystal structure and n from the lattice energy. For NaCl, we have $r_0 = 2.81 \text{ \AA}$, $n \approx 8$, and $U_{\text{diss}} = 7.84 \text{ eV/ion pair}$. This dissociation energy is relatively large. The most energetic photon from the visible spectrum, for example, has an energy of approximately

Equation:

$$hf = (4.14 \times 10^{-15} \text{ eV} \cdot \text{s})(7.5 \times 10^{14} \text{ Hz}) = 3.1 \text{ eV}.$$

Because the ions in crystals are so tightly bound, ionic crystals have the following general characteristics:

1. They are fairly hard and stable.
2. They vaporize at relatively high temperatures (1000 to 2000 K).
3. They are transparent to visible radiation, because photons in the visible portion of the spectrum are not energetic enough to excite an electron from its ground state to an excited state.
4. They are poor electrical conductors, because they contain effectively no free electrons.
5. They are usually soluble in water, because the water molecule has a large dipole moment whose electric field is strong enough to break the electrostatic bonds between the ions.

Example:

The Dissociation Energy of Salt

Determine the dissociation energy of sodium chloride (NaCl) in kJ/mol. (*Hint:* The repulsion constant n of NaCl is approximately 8.)

Strategy

A sodium chloride crystal has an equilibrium separation of 0.282 nm. (Compare this value with 0.236 nm for a free diatomic unit of NaCl.) The dissociation energy depends on the separation distance, repulsion constant,

and Madelung constant for an FCC structure. The separation distance depends in turn on the molar mass and measured density. We can determine the separation distance, and then use this value to determine the dissociation energy for one mole of the solid.

Solution

The atomic masses of Na and Cl are 23.0 u and 58.4 u, so the molar mass of NaCl is 58.4 g/mol. The density of NaCl is 2.16 g/cm³. The relationship between these quantities is

Equation:

$$\rho = \frac{M}{V} = \frac{M}{2N_A r_0^3},$$

where M is the mass of one mole of salt, N_A is Avogadro's number, and r_0 is the equilibrium separation distance. The factor 2 is needed since both the sodium and chloride ions represent a cubic volume r_0^3 . Solving for the distance, we get

Equation:

$$r_0^3 = \frac{M}{2N_A \rho} = \frac{58.4 \text{ g/mol}}{2 (6.03 \times 10^{23}) (2.160 \text{ g/cm}^3)} = 2.23 \times 10^{-23} \text{ cm}^3,$$

or

Equation:

$$r_0 = 2.80 \times 10^{-8} \text{ cm} = 0.280 \text{ nm}.$$

The potential energy of one ion pair ($\text{Na}^+ \text{Cl}^-$) is

Equation:

$$U = -\alpha \frac{ke^2}{r_0} \left(1 - \frac{1}{n} \right),$$

where α is the Madelung constant, r_0 is the equilibrium separation distance, and n is the repulsion constant. NaCl is FCC, so the Madelung constant is $\alpha = 1.7476$. Substituting these values, we get

Equation:

$$U = -1.75 \frac{1.44 \text{ eV} \cdot \text{nm}}{0.280 \text{ nm}} \left(1 - \frac{1}{8}\right) = -7.88 \frac{\text{eV}}{\text{ion pair}}.$$

The dissociation energy of one mole of sodium chloride is therefore

Equation:

$$D = \left(\frac{7.88 \text{ eV}}{\text{ion pair}} \right) \left(\frac{\frac{23.052 \text{ kcal}}{1 \text{ mol}}}{\frac{1 \text{ eV}}{\text{ion pair}}} \right) = 182 \text{ kcal/mol} = 760 \text{ kJ/mol}.$$

Significance

This theoretical value of the dissociation energy of 766 kJ/mol is close to the accepted experimental value of 787 kJ/mol. Notice that for larger density, the equilibrium separation distance between ion pairs is smaller, as expected. This small separation distance drives up the force between ions and therefore the dissociation energy. The conversion at the end of the equation took advantage of the conversion factor $1 \text{ kJ} = 0.239 \text{ kcal}$.

Note:

Exercise:

Problem:

Check Your Understanding If the dissociation energy were larger, would that make it easier or more difficult to break the solid apart?

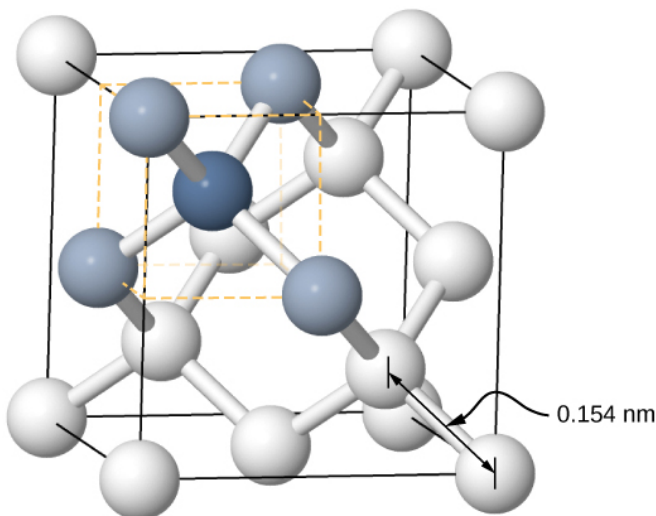
Solution:

more difficult

Covalent Bonding in Solids

Crystals can also be formed by covalent bonding. For example, covalent bonds are responsible for holding carbon atoms together in diamond

crystals. The electron configuration of the carbon atom is $1s^2 2s^2 2p^2$ —a He core plus four valence electrons. This electron configuration is four electrons short of a full shell, so by sharing these four electrons with other carbon atoms in a covalent bond, the shells of all carbon atoms are filled. Diamond has a more complicated structure than most ionic crystals ([link](#)). Each carbon atom is the center of a regular tetrahedron, and the angle between the bonds is 110° . This angle is a direct consequence of the directionality of the p orbitals of carbon atoms.



(a)



(b)

Structure of the diamond crystal. (a) The single carbon atom represented by the dark blue sphere is covalently bonded to the four carbon atoms represented by the light blue spheres. (b) Gem-quality diamonds can be cleaved along smooth planes, which gives a large number of angles that cause total internal reflection of incident light, and thus gives diamonds their prized brilliance.

Covalently bonded crystals are not as uniform as ionic crystals but are reasonably hard, difficult to melt, and are insoluble in water. For example, diamond has an extremely high melting temperature (4000 K) and is transparent to visible light. In comparison, covalently bonded tin (also

known as alpha-tin, which is nonmetallic) is relatively soft, melts at 600 K, and reflects visible light. Two other important examples of covalently bonded crystals are silicon and germanium. Both of these solids are used extensively in the manufacture of diodes, transistors, and integrated circuits. We will return to these materials later in our discussion of semiconductors.

Metallic Bonding in Solids

As the name implies, metallic bonding is responsible for the formation of metallic crystals. The valence electrons are essentially free of the atoms and are able to move relatively easily throughout the metallic crystal. Bonding is due to the attractive forces between the positive ions and the conduction electrons. Metallic bonds are weaker than ionic or covalent bonds, with dissociation energies in the range 1 – 3 eV.

Summary

- Packing structures of common ionic salts include FCC and BCC.
- The density of a crystal is inversely related to the equilibrium constant.
- The dissociation energy of a salt is large when the equilibrium separation distance is small.
- The densities and equilibrium radii for common salts (FCC) are nearly the same.

Conceptual Questions

Exercise:

Problem:

Why is the equilibrium separation distance between K^+ and Cl^- different for a diatomic molecule than for solid KCl?

Solution:

Each ion is in the field of multiple ions of the other opposite charge.

Exercise:**Problem:**

Describe the difference between a face-centered cubic structure (FCC) and a body-centered cubic structure (BCC).

Exercise:**Problem:**

In sodium chloride, how many Cl^- atoms are “nearest neighbors” of Na^+ ? How many Na^+ atoms are “nearest neighbors” of Cl^- ?

Solution:

6, 6

Exercise:**Problem:**

In cesium iodide, how many I^- atoms are “nearest neighbors” of Cs^+ ? How many Cs^+ atoms are “nearest neighbors” of I^- ?

Exercise:**Problem:**

The NaCl crystal structure is FCC. The equilibrium spacing is $r_0 = 0.282$ nm. If each ion occupies a cubic volume of r_0^3 , estimate the distance between “nearest neighbor” Na^+ ions (center-to-center)?

Solution:

0.399 nm

Problems**Exercise:**

Problem:

The CsI crystal structure is BCC. The equilibrium spacing is approximately $r_0 = 0.46$ nm. If Cs^+ ion occupies a cubic volume of r_0^3 , what is the distance of this ion to its “nearest neighbor” I^+ ion?

Solution:

0.65 nm

Exercise:**Problem:**

The potential energy of a crystal is -8.10 eV/ion pair. Find the dissociation energy for four moles of the crystal.

Exercise:**Problem:**

The measured density of a NaF crystal is 2.558 g/cm³. What is the equilibrium separate distance of Na^+ and F^- ions?

Solution:

$r_0 = 0.240$ nm

Exercise:**Problem:**

What value of the repulsion constant, n , gives the measured dissociation energy of 221 kcal/mole for NaF?

Exercise:**Problem:**

Determine the dissociation energy of 12 moles of sodium chloride (NaCl). (*Hint:* the repulsion constant n is approximately 8.)

Solution:

2196 kcal

Exercise:**Problem:**

The measured density of a KCl crystal is 1.984 g/cm^3 . What is the equilibrium separation distance of K^+ and Cl^- ions?

Exercise:**Problem:**

What value of the repulsion constant, n , gives the measured dissociation energy of 171 kcal/mol for KCl?

Solution:

11.5

Exercise:**Problem:**

The measured density of a CsCl crystal is 3.988 g/cm^3 . What is the equilibrium separate distance of Cs^+ and Cl^- ions?

Glossary

body-centered cubic (BCC)

crystal structure in which an ion is surrounded by eight nearest neighbors located at the corners of a unit cell

face-centered cubic (FCC)

crystal structure in which an ion is surrounded by six nearest neighbors located at the faces of a unit cell

lattice

regular array or arrangement of atoms into a crystal structure

repulsion constant

experimental parameter associated with a repulsive force between ions brought so close together that the exclusion principle is important

simple cubic

basic crystal structure in which each ion is located at the nodes of a three-dimensional grid

Free Electron Model of Metals

By the end of this section, you will be able to:

- Describe the classical free electron model of metals in terms of the concept electron number density
- Explain the quantum free-electron model of metals in terms of Pauli's exclusion principle
- Calculate the energy levels and energy-level spacing of a free electron in a metal

Metals, such as copper and aluminum, are held together by bonds that are very different from those of molecules. Rather than sharing and exchanging electrons, a metal is essentially held together by a system of free electrons that wander throughout the solid. The simplest model of a metal is the **free electron model**. This model views electrons as a gas. We first consider the simple one-dimensional case in which electrons move freely along a line, such as through a very thin metal rod. The potential function $U(x)$ for this case is a one-dimensional infinite square well where the walls of the well correspond to the edges of the rod. This model ignores the interactions between the electrons but respects the exclusion principle. For the special case of $T = 0$ K, N electrons fill up the energy levels, from lowest to highest, two at a time (spin up and spin down), until the highest energy level is filled. The highest energy filled is called the **Fermi energy**.

The one-dimensional free electron model can be improved by considering the three-dimensional case: electrons moving freely in a three-dimensional metal block. This system is modeled by a three-dimensional infinite square well. Determining the allowed energy states requires us to solve the time-independent Schrödinger equation

Equation:

$$-\frac{h^2}{2m_e} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi(x, y, z) = E \psi(x, y, z),$$

where we assume that the potential energy inside the box is zero and infinity otherwise. The allowed wave functions describing the electron's quantum states can be written as

Equation:

$$\psi(x, y, z) = \left(\sqrt{\frac{2}{L_x}} \sin \frac{n_x \pi x}{L_x} \right) \left(\sqrt{\frac{2}{L_y}} \sin \frac{n_y \pi y}{L_y} \right) \left(\sqrt{\frac{2}{L_z}} \sin \frac{n_z \pi z}{L_z} \right),$$

where n_x , n_y , and n_z are positive integers representing quantum numbers corresponding to the motion in the x -, y -, and z -directions, respectively, and L_x , L_y , and L_z are the dimensions of the box in those directions. [\[link\]](#) is simply the product of three one-dimensional wave functions. The allowed energies of an electron in a cube ($L = L_x = L_y = L_z$) are

Note:

Equation:

$$E = \frac{\pi^2 \hbar^2}{2mL^2} (n_1^2 + n_2^2 + n_3^2).$$

Associated with each set of quantum numbers (n_x, n_y, n_z) are two quantum states, spin up and spin down. In a real material, the number of filled states is enormous. For example, in a cubic centimeter of metal, this number is on the order of 10^{22} . Counting how many particles are in which state is difficult work, which often requires the help of a powerful computer. The effort is worthwhile, however, because this information is often an effective way to check the model.

Example:

Energy of a Metal Cube

Consider a solid metal cube of edge length 2.0 cm. (a) What is the lowest energy level for an electron within the metal? (b) What is the spacing between this level and the next energy level?

Strategy

An electron in a metal can be modeled as a wave. The lowest energy corresponds to the largest wavelength and smallest quantum number: $n_x, n_y, n_z = (1, 1, 1)$. [\[link\]](#) supplies this “ground state” energy value. Since

the energy of the electron increases with the quantum number, the next highest level involves the smallest increase in the quantum numbers, or $(n_x, n_y, n_z) = (2, 1, 1), (1, 2, 1),$ or $(1, 1, 2).$

Solution

The lowest energy level corresponds to the quantum numbers $n_x = n_y = n_z = 1.$ From [\[link\]](#), the energy of this level is

Equation:

$$\begin{aligned} E(1, 1, 1) &= \frac{\pi^2 \hbar^2}{2m_e L^2} (1^2 + 1^2 + 1^2) \\ &= \frac{3\pi^2 (1.05 \times 10^{-34} \text{ J}\cdot\text{s})^2}{2 (9.11 \times 10^{-31} \text{ kg}) (2.00 \times 10^{-2} \text{ m})^2} \\ &= 4.48 \times 10^{-34} \text{ J} = 2.80 \times 10^{-15} \text{ eV}. \end{aligned}$$

The next-higher energy level is reached by increasing any one of the three quantum numbers by 1. Hence, there are actually three quantum states with the same energy. Suppose we increase n_x by 1. Then the energy becomes

Equation:

$$\begin{aligned} E(2, 1, 1) &= \frac{\pi^2 \hbar^2}{2m_e L^2} (2^2 + 1^2 + 1^2) \\ &= \frac{6\pi^2 (1.05 \times 10^{-34} \text{ J}\cdot\text{s})^2}{2 (9.11 \times 10^{-31} \text{ kg}) (2.00 \times 10^{-2} \text{ m})^2} \\ &= 8.96 \times 10^{-34} \text{ J} = 5.60 \times 10^{-15} \text{ eV}. \end{aligned}$$

The energy spacing between the lowest energy state and the next-highest energy state is therefore

Equation:

$$E(2, 1, 1) - E(1, 1, 1) = 2.80 \times 10^{-15} \text{ eV}.$$

Significance

This is a very small energy difference. Compare this value to the average kinetic energy of a particle, $k_B T$, where k_B is Boltzmann's constant and T is the temperature. The product $k_B T$ is about 1000 times greater than the energy spacing.

Note:

Exercise:**Problem:**

Check Your Understanding What happens to the ground state energy of an electron if the dimensions of the solid increase?

Solution:

It decreases.

Often, we are not interested in the total number of particles in all states, but rather the number of particles dN with energies in a narrow energy interval. This value can be expressed by

Equation:

$$dN = n(E)dE = g(E)dE \cdot F$$

where $n(E)$ is the **electron number density**, or the number of electrons per unit volume; $g(E)$ is the **density of states**, or the number of allowed quantum states per unit energy; dE is the size of the energy interval; and F is the **Fermi factor**. The Fermi factor is the probability that the state will be filled. For example, if $g(E)dE$ is 100 available states, but F is only 5%, then the number of particles in this narrow energy interval is only five. Finding $g(E)$ requires solving Schrödinger's equation (in three dimensions) for the allowed energy levels. The calculation is involved even for a crude model, but the result is simple:

Note:**Equation:**

$$g(E) = \frac{\pi V}{2} \left(\frac{8m_e}{h^2} \right)^{3/2} E^{1/2},$$

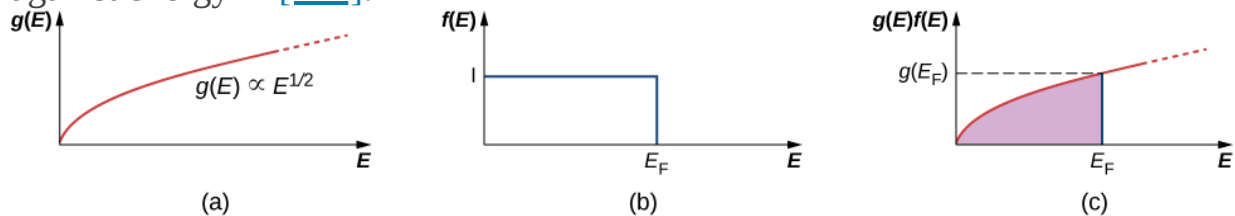
where V is the volume of the solid, m_e is the mass of the electron, and E is the energy of the state. Notice that the density of states increases with the square root of the energy. More states are available at high energy than at low energy. This expression does *not* provide information of the density of the electrons in physical space, but rather the density of energy levels in “energy space.” For example, in our study of the atomic structure, we learned that the energy levels of a hydrogen atom are much more widely spaced for small energy values (near than ground state) than for larger values.

This equation tells us how many electron states are available in a three-dimensional metallic solid. However, it does not tell us how likely these states will be filled. Thus, we need to determine the Fermi factor, F . Consider the simple case of $T = 0$ K. From classical physics, we expect that all the electrons ($\sim 10^{22} / \text{cm}^3$) would simply go into the ground state to achieve the lowest possible energy. However, this violates Pauli’s exclusion principle, which states that no two electrons can be in the same quantum state. Hence, when we begin filling the states with electrons, the states with lowest energy become occupied first, then states with progressively higher energies. The *last electron* we put in has the highest energy. This energy is the Fermi energy E_F of the free electron gas. A state with energy $E < E_F$ is occupied by a single electron, and a state with energy $E > E_F$ is unoccupied. To describe this in terms of a probability $F(E)$ that a state of energy E is occupied, we write for $T = 0$ K:

Equation:

$$\begin{aligned} F(E) &= 1 & (E < E_F) \\ F(E) &= 0 & (E > E_F). \end{aligned}$$

The density of states, Fermi factor, and electron number density are plotted against energy in [\[link\]](#).



(a) Density of states for a free electron gas; (b) probability that a state is occupied at $T = 0$ K; (c) density of occupied states at $T = 0$ K.

A few notes are in order. First, the electron number density (last row) distribution drops off sharply at the Fermi energy. According to the theory, this energy is given by

Note:
Equation:

$$E_F = \frac{h^2}{8m_e} \left(\frac{3N}{\pi V} \right)^{2/3}.$$

Fermi energies for selected materials are listed in the following table.

Element	Conduction Band Electron Density (10 ²⁸ m ⁻³)	Free-Electron Model Fermi Energy (eV)
Al	18.1	11.7
Ba	3.15	3.64
Cu	8.47	7.00
Au	5.90	5.53
Fe	17.0	11.1
Ag	5.86	5.49

Conduction Electron Densities and Fermi Energies for Some Metals

Note also that only the graph in part (c) of the figure, which answers the question, “How many particles are found in the energy range?” is checked by experiment. The **Fermi temperature** or effective “temperature” of an electron at the Fermi energy is

Note:

Equation:

$$T_F = \frac{E_F}{k_B}.$$

Example:

Fermi Energy of Silver

Metallic silver is an excellent conductor. It has 5.86×10^{28} conduction electrons per cubic meter. (a) Calculate its Fermi energy. (b) Compare this energy to the thermal energy $k_B T$ of the electrons at a room temperature of 300 K.

Solution

- a. From [\[link\]](#), the Fermi energy is

Equation:

$$\begin{aligned} E_F &= \frac{h^2}{2m_e} (3\pi^2 n_e)^{2/3} \\ &= \frac{(1.05 \times 10^{-34} \text{ J}\cdot\text{s})^2}{2(9.11 \times 10^{-31} \text{ kg})} \times [(3\pi^2 (5.86 \times 10^{28} \text{ m}^{-3}))^{2/3}] \\ &= 8.79 \times 10^{-19} \text{ J} = 5.49 \text{ eV}. \end{aligned}$$

This is a typical value of the Fermi energy for metals, as can be seen from [\[link\]](#).

- b. We can associate a Fermi temperature T_F with the Fermi energy by writing $k_B T_F = E_F$. We then find for the Fermi temperature

Equation:

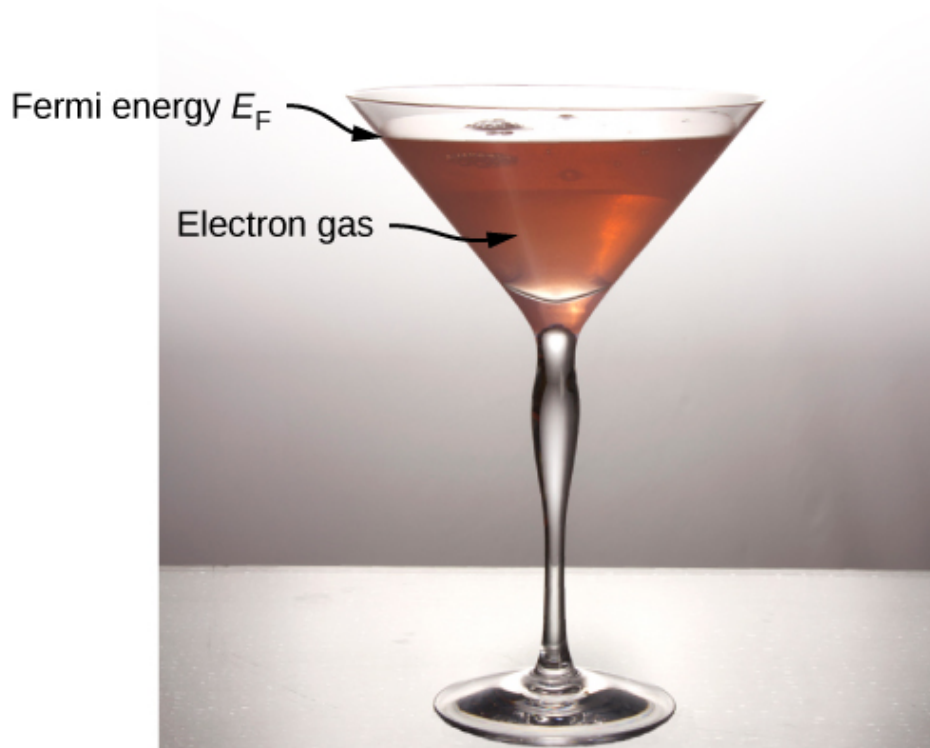
$$T_F = \frac{8.79 \times 10^{-19} \text{ J}}{1.38 \times 10^{-23} \text{ J/K}} = 6.37 \times 10^4 \text{ K},$$

which is much higher than room temperature and also the typical melting point ($\sim 10^3 \text{ K}$) of a metal. The ratio of the Fermi energy of silver to the room-temperature thermal energy is

Equation:

$$\frac{E_F}{k_B T} = \frac{T_F}{T} \approx 210.$$

To visualize how the quantum states are filled, we might imagine pouring water slowly into a glass, such as that of [\[link\]](#). The first drops of water (the electrons) occupy the bottom of the glass (the states with lowest energy). As the level rises, states of higher and higher energy are occupied. Furthermore, since the glass has a wide opening and a narrow stem, more water occupies the top of the glass than the bottom. This reflects the fact that the density of states $g(E)$ is proportional to $E^{1/2}$, so there is a relatively large number of higher energy electrons in a free electron gas. Finally, the level to which the glass is filled corresponds to the Fermi energy.



An analogy of how electrons fill energy states in a metal. As electrons fill energy states, lowest to highest, the number of available states increases. The highest energy state (corresponding to the water line) is the Fermi energy. (credit: modification of work by “Didriks”/Flickr)

Suppose that at $T = 0$ K, the number of conduction electrons per unit volume in our sample is n_e . Since each field state has one electron, the number of filled states per unit volume is the same as the number of electrons per unit volume.

Summary

- Metals conduct electricity, and electricity is composed of large numbers of randomly colliding and approximately free electrons.
- The allowed energy states of an electron are quantized. This quantization appears in the form of very large electron energies, even at $T = 0$ K.

- The allowed energies of free electrons in a metal depend on electron mass and on the electron number density of the metal.
- The density of states of an electron in a metal increases with energy, because there are more ways for an electron to fill a high-energy state than a low-energy state.
- Pauli's exclusion principle states that only two electrons (spin up and spin down) can occupy the same energy level. Therefore, in filling these energy levels (lowest to highest at $T = 0$ K), the last and largest energy level to be occupied is called the Fermi energy.

Conceptual Questions

Exercise:

Problem:

Why does the Fermi energy (E_F) increase with the number of electrons in a metal?

Exercise:

Problem:

If the electron number density (N/V) of a metal increases by a factor 8, what happens to the Fermi energy (E_F)?

Solution:

increases by a factor of $\sqrt[3]{8^2} = 4$

Exercise:

Problem:

Why does the horizontal line in the graph in [\[link\]](#) suddenly stop at the Fermi energy?

Exercise:

Problem: Why does the graph in [\[link\]](#) increase gradually from the origin?

Solution:

For larger energies, the number of accessible states increases.

Exercise:

Problem:

Why are the sharp transitions at the Fermi energy “smoothed out” by increasing the temperature?

Problems

Exercise:

Problem:

What is the difference in energy between the $n_x = n_y = n_z = 4$ state and the state with the next higher energy? What is the percentage change in the energy between the $n_x = n_y = n_z = 4$ state and the state with the next higher energy? (b) Compare these with the difference in energy and the percentage change in the energy between the $n_x = n_y = n_z = 400$ state and the state with the next higher energy.

Solution:

a. 4%; b. $4.2 \times 10^{-4}\%$; for very large values of the quantum numbers, the spacing between adjacent energy levels is very small (“in the continuum”). This is consistent with the expectation that for large quantum numbers, quantum and classical mechanics give approximately the same predictions.

Exercise:

Problem:

An electron is confined to a metal cube of $l = 0.8$ cm on each side. Determine the density of states at (a) $E = 0.80$ eV; (b) $E = 2.2$ eV; and (c) $E = 5.0$ eV.

Exercise:

Problem:

What value of energy corresponds to a density of states of $1.10 \times 10^{24} \text{ eV}^{-1}$?

Solution:

10.0 eV

Exercise:

Problem: Compare the density of states at 2.5 eV and 0.25 eV.

Exercise:**Problem:**

Consider a cube of copper with edges 1.50 mm long. Estimate the number of electron quantum states in this cube whose energies are in the range 3.75 to 3.77 eV.

Solution:

4.55×10^9

Exercise:**Problem:**

If there is one free electron per atom of copper, what is the electron number density of this metal?

Exercise:**Problem:**

Determine the Fermi energy and temperature for copper at $T = 0 \text{ K}$.

Solution:

Fermi energy, $E_F = 7.03 \text{ eV}$, Temperature, $T_F = 8.2 \times 10^4 \text{ K}$

Glossary

density of states

number of allowed quantum states per unit energy

electron number density

number of electrons per unit volume

Fermi energy

largest energy filled by electrons in a metal at $T = 0$ K

Fermi factor

number that expresses the probability that a state of given energy will be filled

Fermi temperature

effective temperature of electrons with energies equal to the Fermi energy

free electron model

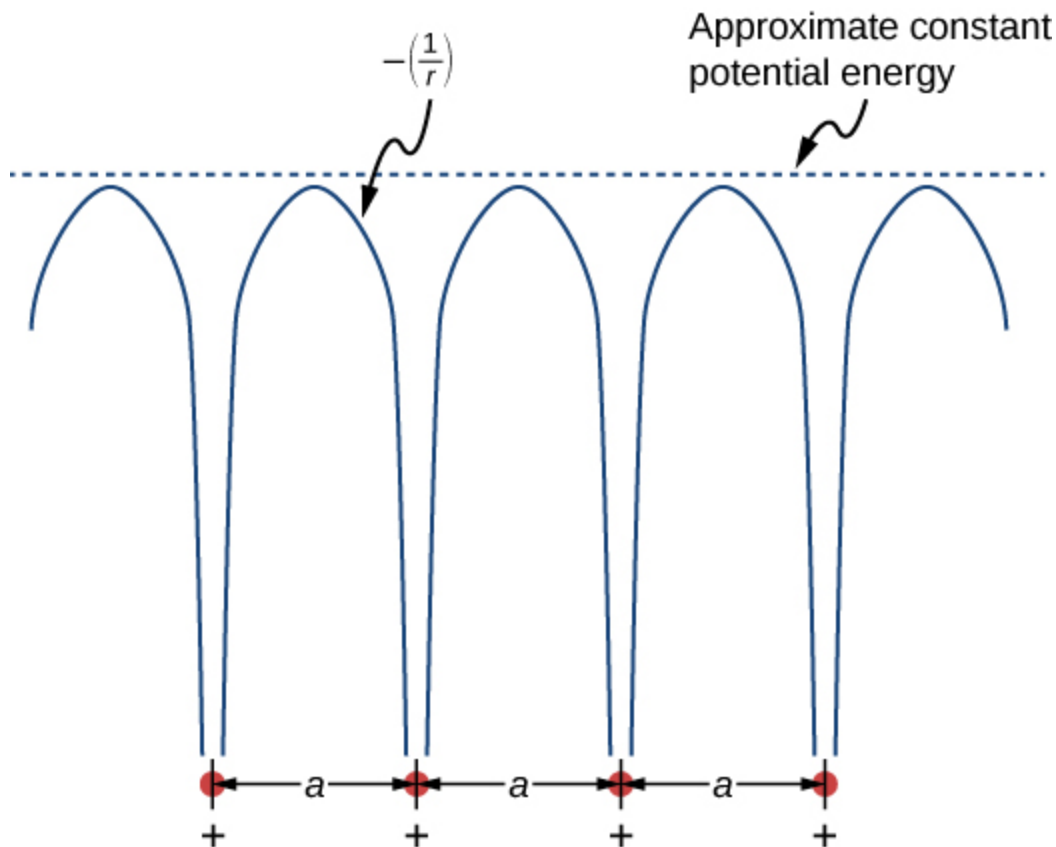
model of a metal that views electrons as a gas

Band Theory of Solids

By the end of this section, you will be able to:

- Describe two main approaches to determining the energy levels of an electron in a crystal
- Explain the presence of energy bands and gaps in the energy structure of a crystal
- Explain why some materials are good conductors and others are good insulators
- Differentiate between an insulator and a semiconductor

The free electron model explains many important properties of conductors but is weak in at least two areas. First, it assumes a constant potential energy within the solid. (Recall that a constant potential energy is associated with no forces.) [\[link\]](#) compares the assumption of a constant potential energy (dotted line) with the periodic Coulomb potential, which drops as $-1/r$ at each lattice point, where r is the distance from the ion core (solid line). Second, the free electron model assumes an impenetrable barrier at the surface. This assumption is not valid, because under certain conditions, electrons can escape the surface—such as in the photoelectric effect. In addition to these assumptions, the free electron model does not explain the dramatic differences in electronic properties of conductors, semiconductors, and insulators. Therefore, a more complete model is needed.



The periodic potential used to model electrons in a conductor. Each ion in the solid is the source of a Coulomb potential. Notice that the free electron model is productive because the average of this field is approximately constant.

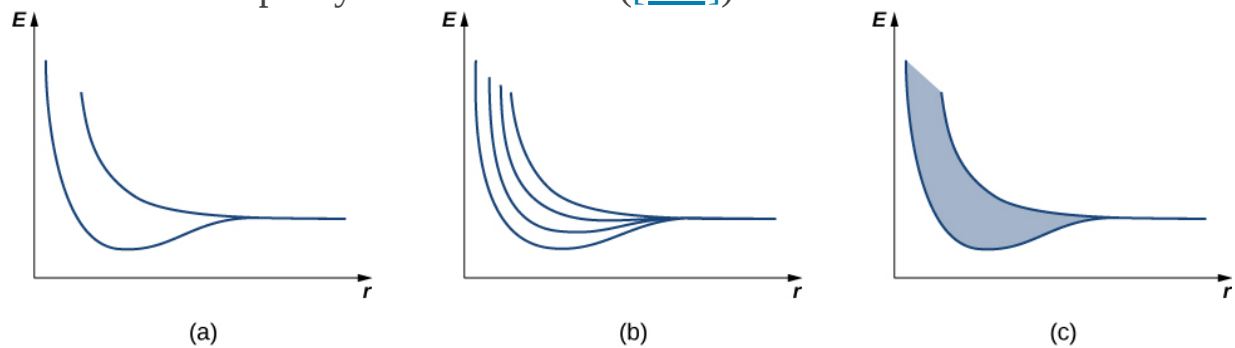
We can produce an improved model by solving Schrödinger's equation for the periodic potential shown in [\[link\]](#). However, the solution requires technical mathematics far beyond our scope. We again seek a qualitative argument based on quantum mechanics to find a way forward.

We first review the argument used to explain the energy structure of a covalent bond. Consider two identical hydrogen atoms so far apart that there is no interaction whatsoever between them. Further suppose that the electron in each atom is in the same ground state: a 1s electron with an energy of -13.6 eV (ignore spin). When the hydrogen atoms are brought

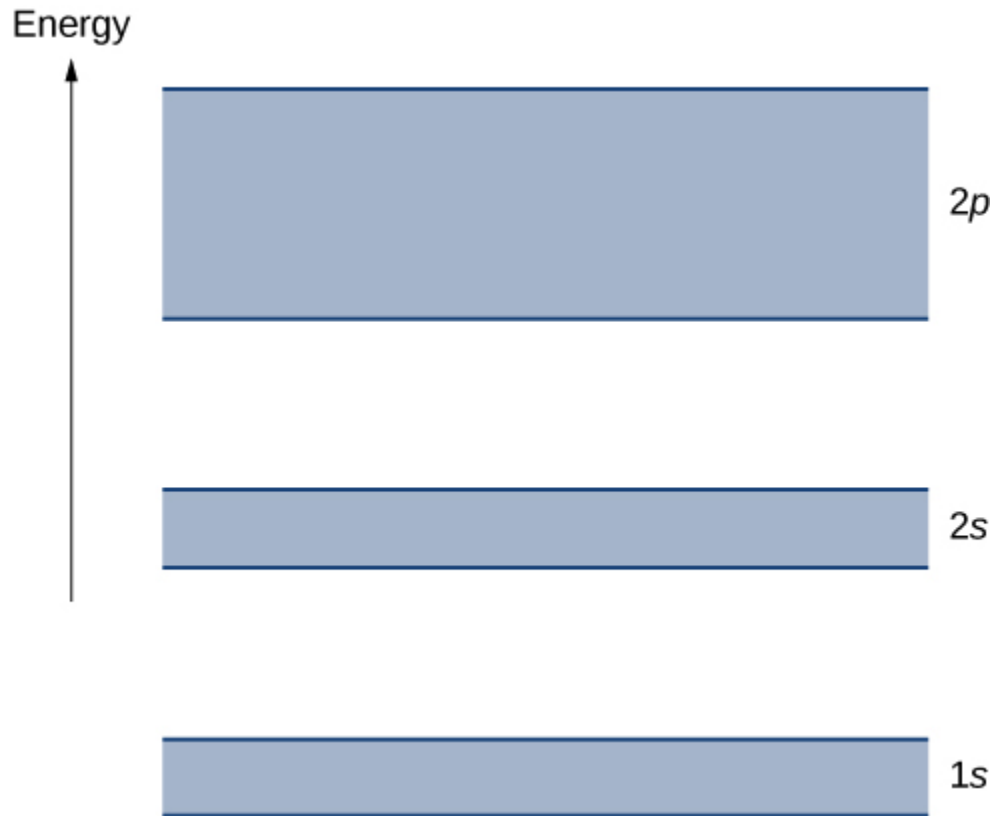
closer together, the individual wave functions of the electrons overlap and, by the exclusion principle, can no longer be in the same quantum state, which splits the original equivalent energy levels into two different energy levels. The energies of these levels depend on the interatomic distance, \propto ([link](#)).

If four hydrogen atoms are brought together, four levels are formed from the four possible symmetries—a single sine wave “hump” in each well, alternating up and down, and so on. In the limit of a very large number N of atoms, we expect a spread of nearly continuous bands of electronic energy levels in a solid (see [link](#)(c)). Each of these bands is known as an **energy band**. (The allowed states of energy and wave number are still technically quantized, but for large numbers of atoms, these states are so close together that they are considered to be continuous or “in the continuum.”)

Energy bands differ in the number of electrons they hold. In the 1s and 2s energy bands, each energy level holds up to two electrons (spin up and spin down), so this band has a maximum occupancy of $2N$ electrons. In the 2p energy band, each energy level holds up to six electrons, so this band has a maximum occupancy of $6N$ electrons ([link](#)).



The dependence of energy-level splitting on the average distance between (a) two atoms, (b) four atoms, and (c) a large number of atoms. For a large number of electrons, a continuous band of energies is produced.

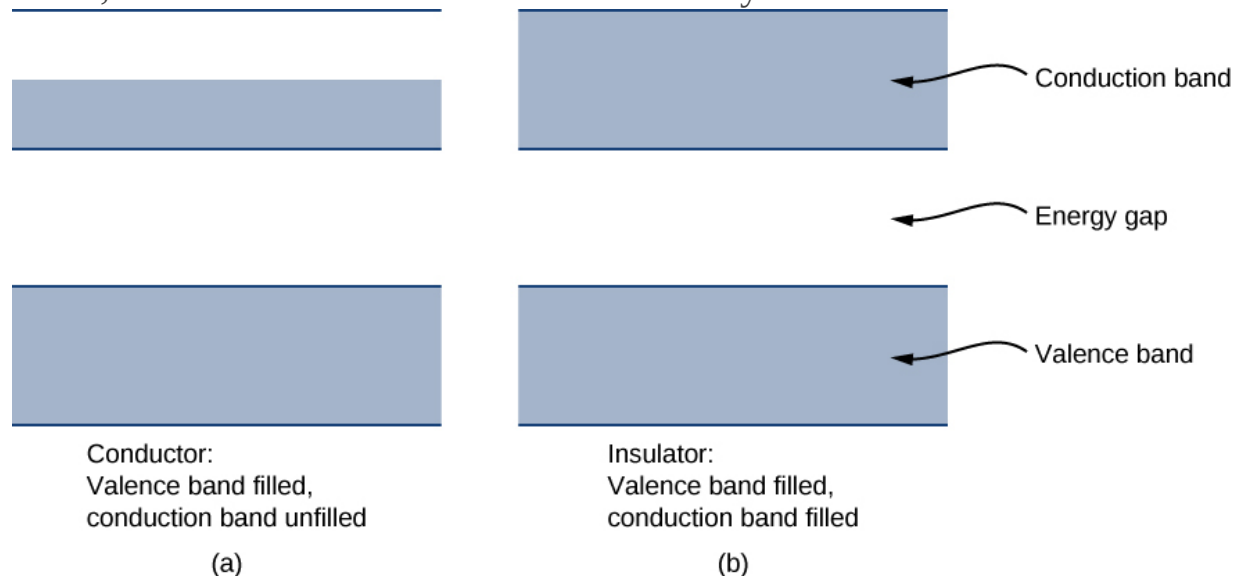


A simple representation of the energy structure of a solid. Electrons belong to energy bands separated by energy gaps.

Each energy band is separated from the other by an **energy gap**. The electrical properties of conductors and insulators can be understood in terms of energy bands and gaps. The highest energy band that is filled is known as a **valence band**. The next available band in the energy structure is known as a **conduction band**. In a conductor, the highest energy band that contains electrons is partially filled, whereas in an insulator, the highest energy band containing electrons is completely filled. The difference between a conductor and insulator is illustrated in [\[link\]](#).

A conductor differs from an insulator in how its electrons respond to an applied electric field. If a significant number of electrons are set into motion by the field, the material is a conductor. In terms of the band model,

electrons in the partially filled conduction band gain kinetic energy from the electric field by filling higher energy states in the conduction band. By contrast, in an insulator, electrons belong to completely filled bands. When the field is applied, the electrons cannot make such transitions (acquire kinetic energy from the electric field) due to the exclusion principle. As a result, the material does not conduct electricity.



Comparison of a conductor and insulator. The highest energy band is partially filled in a conductor but completely filled in an insulator.

Note:

Visit this [simulation](#) to learn about the origin of energy bands in crystals of atoms and how the structure of bands determines how a material conducts electricity. Explore how band structure creates a lattice of many wells.

A **semiconductor** has a similar energy structure to an insulator except it has a relatively small energy gap between the lowest completely filled band and the next available unfilled band. This type of material forms the basis of modern electronics. At $T = 0$ K, the semiconductor and insulator both have

completely filled bands. The only difference is in the size of the energy gap (or *band gap*) E_g between the highest energy band that is filled (the valence band) and the next-higher empty band (the conduction band). In a semiconductor, this gap is small enough that a substantial number of electrons from the valence band are thermally excited into the conduction band at room temperature. These electrons are then in a nearly empty band and can respond to an applied field. As a general rule of thumb, the band gap of a semiconductor is about 1 eV. (See [\[link\]](#) for silicon.) A band gap of greater than approximately 1 eV is considered an insulator. For comparison, the energy gap of diamond (an insulator) is several electron-volts.

Material	Energy Gap E_g (eV)
Si	1.14
Ge	0.67
GaAs	1.43
GaP	2.26
GaSb	0.69
InAs	0.35
InP	1.35
InSb	0.16
C (diamond)	5.48

Energy Gap for Various Materials at 300 K
 Note: Except for diamond, the materials listed are all semiconductors

materials listed are all semiconductors.

Summary

- The energy levels of an electron in a crystal can be determined by solving Schrödinger's equation for a periodic potential and by studying changes to the electron energy structure as atoms are pushed together from a distance.
- The energy structure of a crystal is characterized by continuous energy bands and energy gaps.
- The ability of a solid to conduct electricity relies on the energy structure of the solid.

Conceptual Questions

Exercise:

Problem:

What are the two main approaches used to determine the energy levels of electrons in a crystal?

Solution:

- (1) Solve Schrödinger's equation for the allowed states and energies.
- (2) Determine energy levels for the case of a very large lattice spacing and then determine the energy levels as this spacing is reduced.

Exercise:

Problem:

Describe two features of energy levels for an electron in a crystal.

Exercise:

Problem:

How does the number of energy levels in a band correspond to the number, N , of atoms.

Solution:

For N atoms spaced far apart, there are N different wave functions, all with the same energy (similar to the case of an electron in the double well of H_2). As the atoms are pushed together, the energies of these N different wave functions are split. By the exclusion principle, each electron must each have a unique set of quantum numbers, so the N atoms bringing N electrons together must have at least N states.

Exercise:**Problem:**

Why are some materials very good conductors and others very poor conductors?

Exercise:

Problem: Why are some materials semiconductors?

Solution:

For a semiconductor, there is a relatively large energy gap between the lowest completely filled band and the next available unfilled band. Typically, a number of electrons traverse the gap and therefore the electrical conductivity is small. The properties of a semiconductor are sensitivity to temperature: As the temperature is increased, thermal excitations promote charge carriers from the valence band across the gap and into the conduction band.

Exercise:**Problem:**

Why does the resistance of a semiconductor decrease as the temperature increases?

Problems

Exercise:**Problem:**

For a one-dimensional crystal, write the lattice spacing (a) in terms of the electron wavelength.

Exercise:**Problem:**

What is the main difference between an insulator and a semiconductor?

Solution:

For an insulator, the energy gap between the valence band and the conduction band is larger than for a semiconductor.

Exercise:**Problem:**

What is the longest wavelength for a photon that can excite a valence electron into the conduction band across an energy gap of 0.80 eV?

Exercise:**Problem:**

A valence electron in a crystal absorbs a photon of wavelength, $\lambda = 0.300$ nm. This is just enough energy to allow the electron to jump from the valence band to the conduction band. What is the size of the energy gap?

Solution:

4.13 keV

Glossary

conduction band

above the valence band, the next available band in the energy structure of a crystal

energy band

nearly continuous band of electronic energy levels in a solid

energy gap

gap between energy bands in a solid

semiconductor

solid with a relatively small energy gap between the lowest completely filled band and the next available unfilled band

valence band

highest energy band that is filled in the energy structure of a crystal

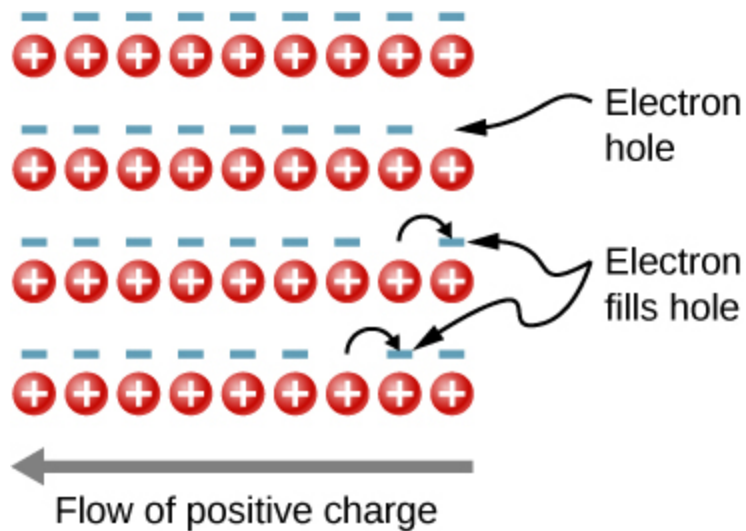
Semiconductors and Doping

By the end of this section, you will be able to:

- Describe changes to the energy structure of a semiconductor due to doping
- Distinguish between an n-type and p-type semiconductor
- Describe the Hall effect and explain its significance
- Calculate the charge, drift velocity, and charge carrier number density of a semiconductor using information from a Hall effect experiment

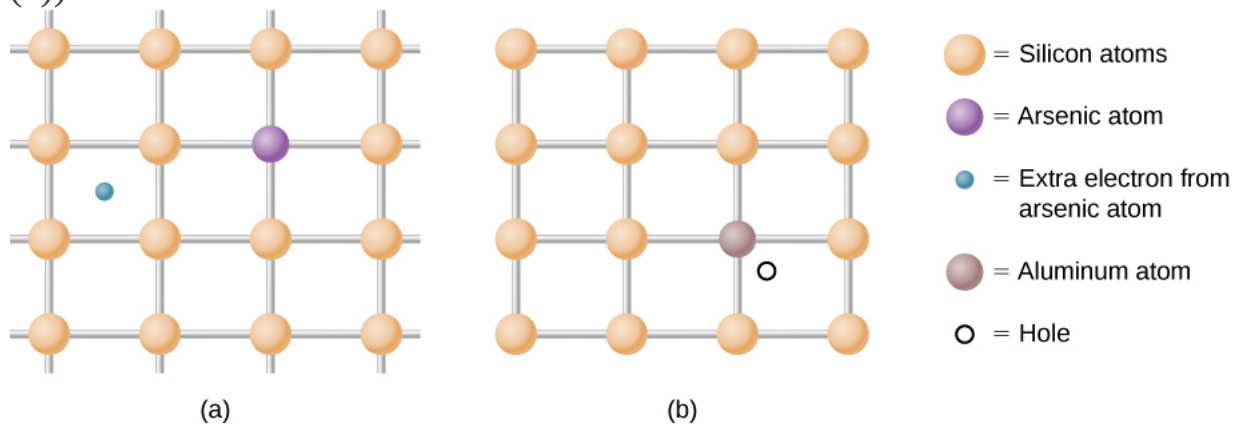
In the preceding section, we considered only the contribution to the electric current due to electrons occupying states in the conduction band. However, moving an electron from the valence band to the conduction band leaves an unoccupied state or **hole** in the energy structure of the valence band, which a nearby electron can move into. As these holes are filled by other electrons, new holes are created. The electric current associated with this filling can be viewed as the collective motion of many negatively charged electrons or the motion of the positively charged electron holes.

To illustrate, consider the one-dimensional lattice in [\[link\]](#). Assume that each lattice atom contributes one valence electron to the current. As the hole on the right is filled, this hole moves to the left. The current can be interpreted as the flow of positive charge to the left. The density of holes, or the number of holes per unit volume, is represented by p . Each electron that transitions into the conduction band leaves behind a hole. If the conduction band is originally empty, the conduction electron density p is equal to the hole density, that is, $n = p$.



The motion of holes in a crystal lattice.
As electrons shift to the right, an
electron hole moves to the left.

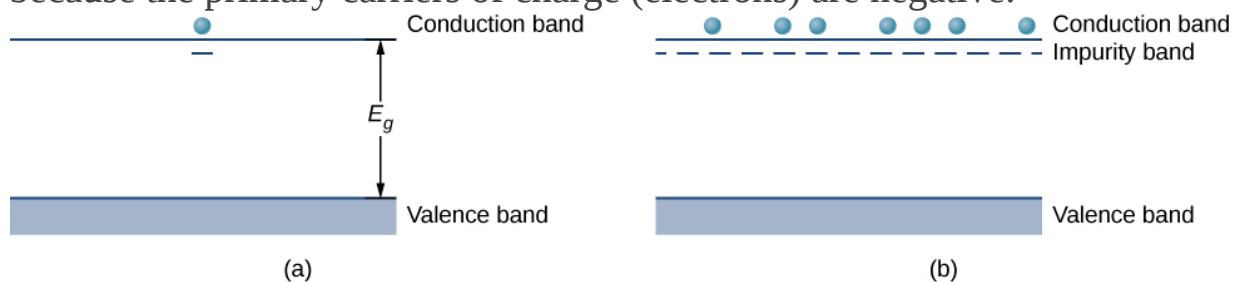
As mentioned, a semiconductor is a material with a filled valence band, an unfilled conduction band, and a relatively small energy gap between the bands. Excess electrons or holes can be introduced into the material by the substitution into the crystal lattice of an **impurity atom**, which is an atom of a slightly different valence number. This process is known as **doping**. For example, suppose we add an arsenic atom to a crystal of silicon ([link](#) (a)).



(a) A donor impurity and (b) an acceptor impurity. The introduction to

impurities and acceptors into a semiconductor significantly changes the electronic properties of this material.

Arsenic has five valence electrons, whereas silicon has only four. This extra electron must therefore go into the conduction band, since there is no room in the valence band. The arsenic ion left behind has a net positive charge that weakly binds the delocalized electron. The binding is weak because the surrounding atomic lattice shields the ion's electric field. As a result, the binding energy of the extra electron is only about 0.02 eV. In other words, the energy level of the impurity electron is in the band gap below the conduction band by 0.02 eV, a much smaller value than the energy of the gap, 1.14 eV. At room temperature, this impurity electron is easily excited into the conduction band and therefore contributes to the conductivity ([link](a)). An impurity with an extra electron is known as a **donor impurity**, and the doped semiconductor is called an ***n*-type semiconductor** because the primary carriers of charge (electrons) are negative.

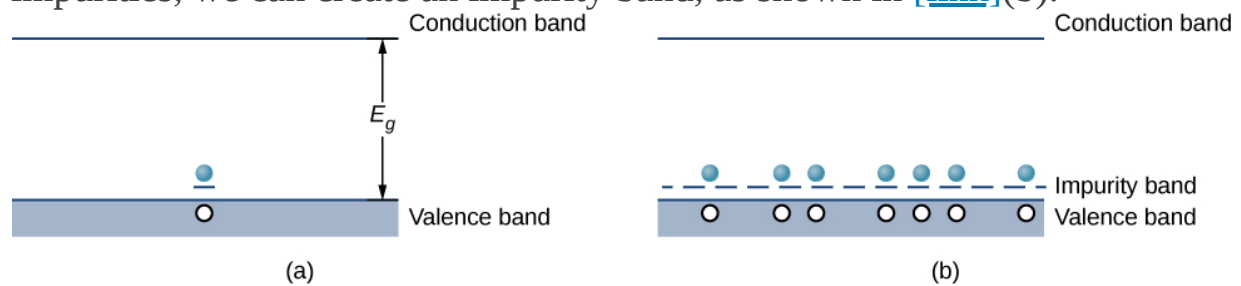


(a) The extra electron from a donor impurity is excited into the conduction band; (b) formation of an impurity band in an *n*-type semiconductor.

By adding more donor impurities, we can create an **impurity band**, a new energy band created by semiconductor doping, as shown in [link](b). The Fermi level is now between this band and the conduction band. At room temperature, many impurity electrons are thermally excited into the conduction band and contribute to the conductivity. Conduction can then also occur in the impurity band as vacancies are created there. Note that

changes in the energy of an electron correspond to a change in the motion (velocities or kinetic energy) of these charge carriers with the semiconductor, but not the bulk motion of the semiconductor itself.

Doping can also be accomplished using impurity atoms that typically have one *fewer* valence electron than the semiconductor atoms. For example, Al, which has three valence electrons, can be substituted for Si, as shown in [\[link\]](#)(b). Such an impurity is known as an **acceptor impurity**, and the doped semiconductor is called a ***p*-type semiconductor**, because the primary carriers of charge (holes) are positive. If a hole is treated as a positive particle weakly bound to the impurity site, then an empty electron state is created in the band gap just above the valence band. When this state is filled by an electron thermally excited from the valence band ([\[link\]](#)(a)), a mobile hole is created in the valence band. By adding more acceptor impurities, we can create an impurity band, as shown in [\[link\]](#)(b).



(a) An electron from the conduction band is excited into the empty state resulting from the acceptor impurity; (b) formation of an impurity band in a *p*-type semiconductor.

The electric current of a doped semiconductor can be due to the motion of a **majority carrier**, in which holes are contributed by an impurity atom, or due to a **minority carrier**, in which holes are contributed purely by thermal excitations of electrons across the energy gap. In an *n*-type semiconductor, majority carriers are free electrons contributed by impurity atoms, and minority carriers are free holes left by the filling of states due to thermal excitation of electrons across the gap. In a *p*-type semiconductor, the majority carriers are free electrons produced by thermal excitations from the valence to the conduction band. In general, the number of majority

carriers far exceeds the minority carriers. The concept of a majority and minority carriers will be used in the next section to explain the operation of diodes and transistors.

In studying p - and n -type doping, it is natural to ask: Do “electron holes” really act like particles? The existence of holes in a doped p -type semiconductor is demonstrated by the Hall effect. The Hall effect is the production of a potential difference due to the motion of a conductor through an external magnetic field (see [The Hall Effect](#)). A schematic of the Hall effect is shown in [\[link\]](#)(a). A semiconductor strip is bathed in a uniform magnetic field (which points into the paper). As the electron holes move from left to right through the semiconductor, a Lorentz force drives these charges toward the upper end of the strip. (Recall that the motion of the positively charged carriers is determined by the right-hand rule.) Positive charge continues to collect on the upper edge of the strip until the force associated with the downward electric field between the upper and lower edges of the strip ($F_E = Eq$) just balances the upward magnetic force ($F_B = qvB$). Setting these forces equal to each other, we have $E = vB$. The voltage that develops across the strip is therefore

Note:

Equation:

$$V_H = vBw,$$

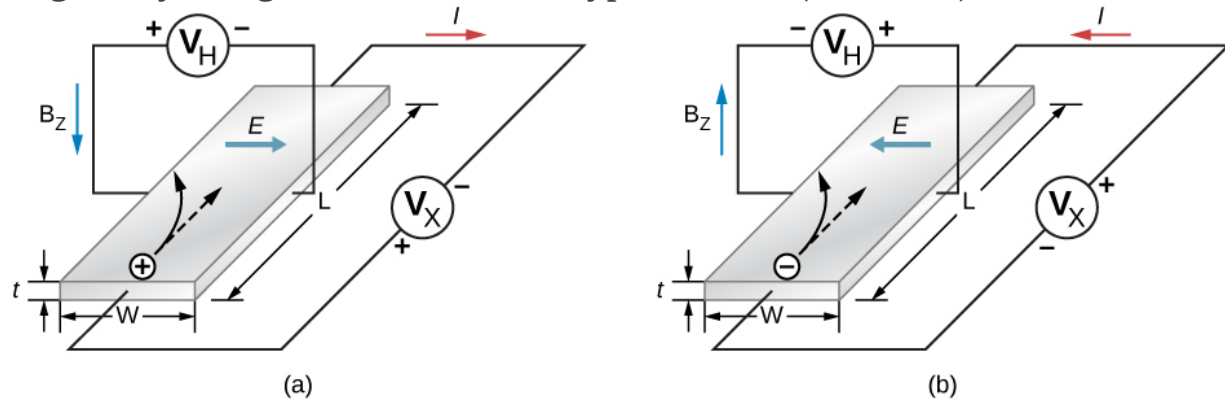
where V_H is the Hall voltage; v is the hole’s **drift velocity**, or average velocity of a particle that moves in a partially random fashion; B is the magnetic field strength; and w is the width of the strip. Note that the Hall voltage is transverse to the voltage that initially produces current through the material. A measurement of the sign of this voltage (or potential difference) confirms the collection of holes on the top side of the strip. The magnitude of the Hall voltage yields the drift velocity (v) of the majority carriers.

Additional information can also be extracted from the Hall voltage. Note that the electron current density (the amount of current per unit cross-sectional area of the semiconductor strip) is

Equation:

$$j = nqv,$$

where q is the magnitude of the charge, n is the number of charge carriers per unit volume, and v is the drift velocity. The current density is easily determined by dividing the total current by the cross-sectional area of the strip, q is charge of the hole (the magnitude of the charge of a single electron), and u is determined by the Hall effect [\[link\]](#). Hence, the above expression for the electron current density gives the number of charge carriers per unit volume, n . A similar analysis can be conducted for negatively charged carriers in an n -type material (see [\[link\]](#)).



The Hall effect. (a) Positively charged electron holes are drawn to the left by a uniform magnetic field that points downward. An electric field is generated to the right. (b) Negative charged electrons are drawn to the left by a magnetic field that points up. An electric field is generated to the left.

Summary

- The energy structure of a semiconductor can be altered by substituting one type of atom with another (doping).
- Semiconductor *n*-type doping creates and fills new energy levels just below the conduction band.
- Semiconductor *p*-type doping creates new energy levels just above the valence band.
- The Hall effect can be used to determine charge, drift velocity, and charge carrier number density of a semiconductor.

Conceptual Questions

Exercise:

Problem:

What kind of semiconductor is produced if germanium is doped with (a) arsenic, and (b) gallium?

Solution:

- a. Germanium has four valence electrons. If germanium doped with *arsenic* (five valence electrons), four are used in bonding and one electron will be left for conduction. This produces an *n*-type material.
- b. If germanium is doped with *gallium* (three valence electrons), all three electrons are used in bonding, leaving one hole for conduction. This results in a *p*-type material.

Exercise:

Problem:

What kind of semiconductor is produced if silicon is doped with (a) phosphorus, and (b) indium?

Exercise:

Problem: What is the Hall effect and what is it used for?

Solution:

The Hall effect is the production of a potential difference due to motion of a conductor through an external magnetic field. This effect can be used to determine the drift velocity of the charge carriers (electrons or hole). If the current density is measured, this effect can also determine the number of charge carriers per unit volume.

Exercise:

Problem:

For an n -type semiconductor, how do impurity atoms alter the energy structure of the solid?

Exercise:

Problem:

For a p -type semiconductor, how do impurity atoms alter the energy structure of the solid?

Solution:

It produces new unfilled energy levels just above the filled valence band. These levels accept electrons from the valence band.

Problems

Exercise:

Problem:

An experiment is performed to demonstrate the Hall effect. A thin rectangular strip of semiconductor with width 10 cm and length 30 cm is attached to a battery and immersed in a 1.50- T field perpendicular to its surface. This produced a Hall voltage of 12 V. What is the drift velocity of the charge carriers?

Exercise:

Problem:

Suppose that the cross-sectional area of the strip (the area of the face perpendicular to the electric current) presented to the in the preceding problem is 1 mm^2 and the current is independently measured to be 2 mA. What is the number density of the charge carriers?

Solution:

$$n = 1.56 \times 10^{19} \text{ holes/m}^3$$

Exercise:**Problem:**

A current-carrying copper wire with cross-section $\sigma = 2 \text{ mm}^2$ has a drift velocity of 0.02 cm/s. Find the total current running through the wire.

Exercise:**Problem:**

The Hall effect is demonstrated in the laboratory. A thin rectangular strip of semiconductor with width 5 cm and cross-sectional area 2 mm^2 is attached to a battery and immersed in a field perpendicular to its surface. The Hall voltage reads 12.5 V and the measured drift velocity is 50 m/s. What is the magnetic field?

Solution:

5 T

Glossary

acceptor impurity

atom substituted for another in a semiconductor that results in a free electron

donor impurity

atom substituted for another in a semiconductor that results in a free electron hole

doping

alteration of a semiconductor by the substitution of one type of atom with another

drift velocity

average velocity of a randomly moving particle

hole

unoccupied states in an energy band

impurity atom

acceptor or donor impurity atom

impurity band

new energy band create by semiconductor doping

majority carrier

free electrons (or holes) contributed by impurity atoms

minority carrier

free electrons (or holes) produced by thermal excitations across the energy gap

n-type semiconductor

doped semiconductor that conducts electrons

p-type semiconductor

doped semiconductor that conducts holes

Semiconductor Devices

By the end of this section, you will be able to:

- Describe what occurs when n- and p-type materials are joined together using the concept of diffusion and drift current (zero applied voltage)
- Explain the response of a p-n junction to a forward and reverse bias voltage
- Describe the function of a transistor in an electric circuit
- Use the concept of a p-n junction to explain its applications in audio amplifiers and computers

Semiconductors have many applications in modern electronics. We describe some basic semiconductor devices in this section. A great advantage of using semiconductors for circuit elements is the fact that many thousands or millions of semiconductor devices can be combined on the same tiny piece of silicon and connected by conducting paths. The resulting structure is called an integrated circuit (ic), and ic chips are the basis of many modern devices, from computers and smartphones to the internet and global communications networks.

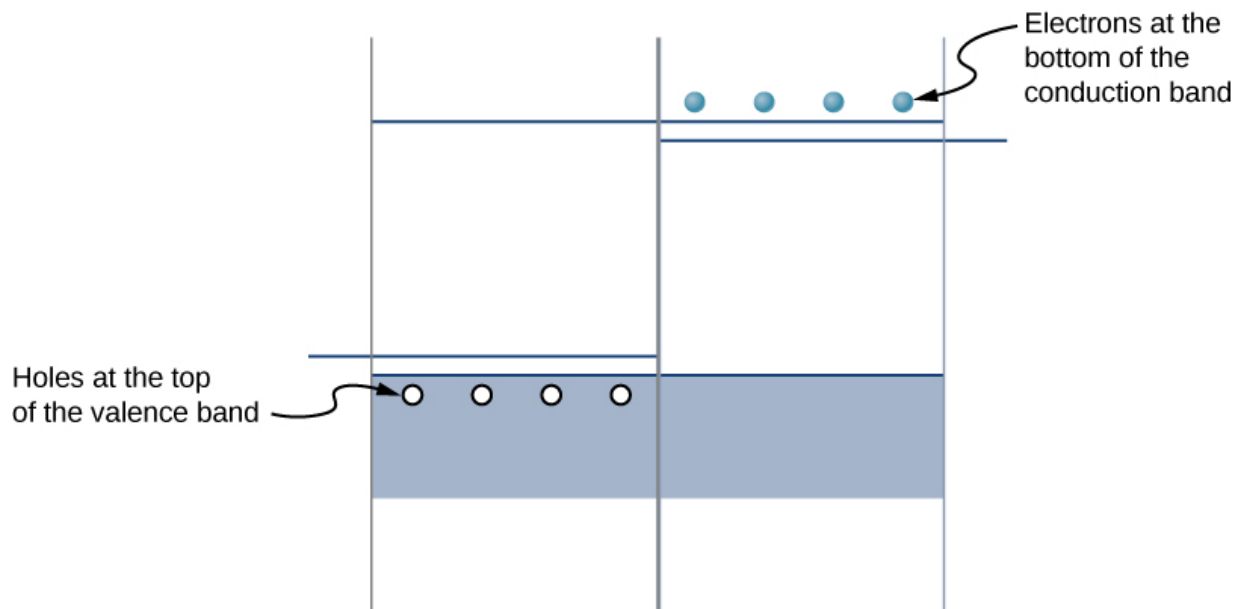
Diodes

Perhaps the simplest device that can be created with a semiconductor is a diode. A diode is a circuit element that allows electric current to flow in only one direction, like a one-way valve (see [Model of Conduction in Metals](#)). A diode is created by joining a *p*-type semiconductor to an *n*-type semiconductor ([\[link\]](#)). The junction between these materials is called a ***p-n* junction**. A comparison of the energy bands of a silicon-based diode is shown in [\[link\]\(b\)](#). The positions of the valence and conduction bands are the same, but the impurity levels are quite different. When a *p-n* junction is formed, electrons from the conduction band of the *n*-type material diffuse to the *p*-side, where they combine with holes in the valence band. This migration of charge leaves positive ionized donor ions on the *n*-side and negative ionized acceptor ions on the *p*-side, producing a narrow double layer of charge at the *p-n* junction called the **depletion layer**. The electric

field associated with the depletion layer prevents further diffusion. The potential energy for electrons across the p - n junction is given by [\[link\]](#).

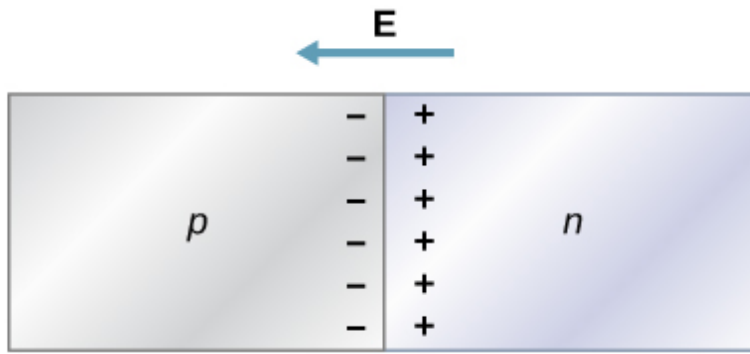


(a)

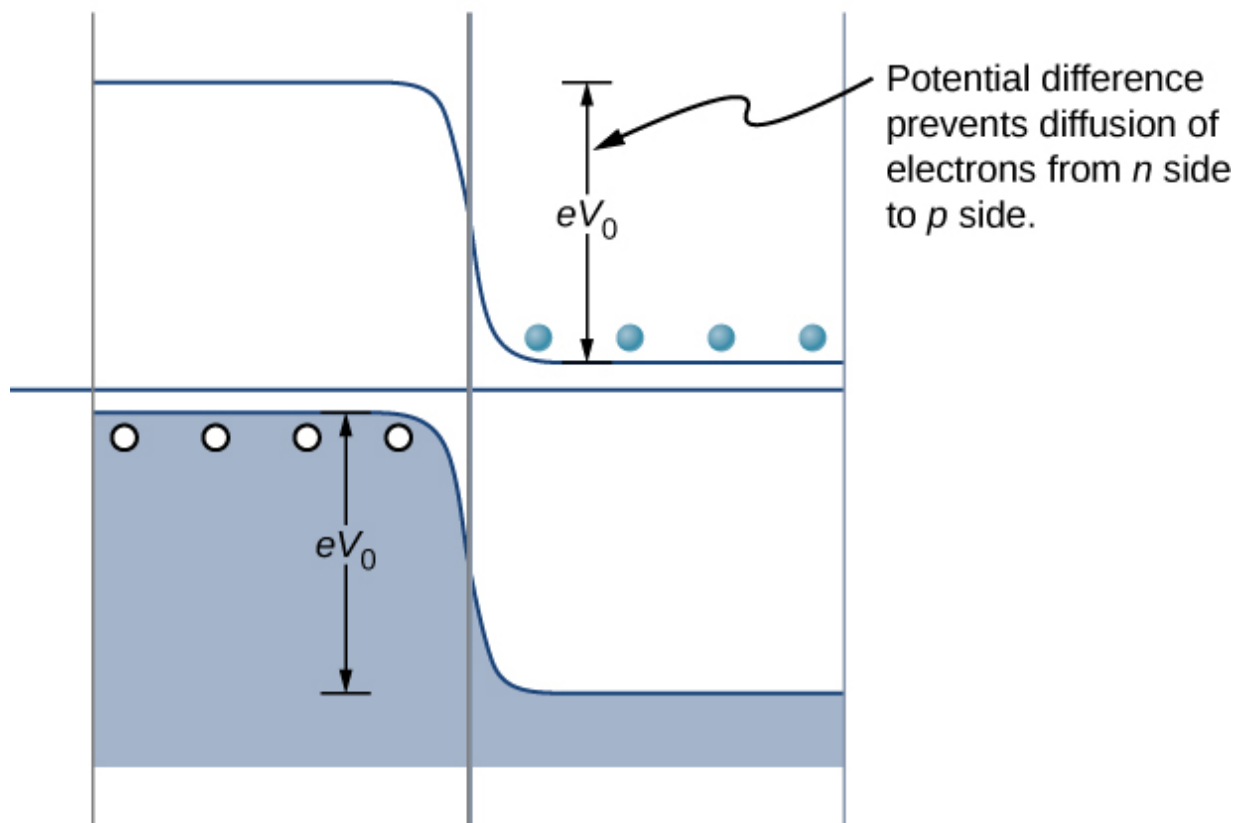


(b)

(a) Representation of a p - n junction. (b) A comparison of the energy bands of p -type and n -type silicon prior to equilibrium.



(a)



(b)

At equilibrium, (a) excess charge resides near the interface and the net current is zero, and (b) the potential energy difference for electrons (in light blue) prevents further diffusion of electrons into the p -side.

The behavior of a semiconductor diode can now be understood. If the positive side of the battery is connected to the n -type material, the depletion layer is widened, and the potential energy difference across the p - n junction is increased. Few or none of the electrons (holes) have enough energy to climb the potential barrier, and current is significantly reduced. This is called the **reverse bias configuration**. On the other hand, if the positive side of a battery is connected to the p -type material, the depletion layer is narrowed, the potential energy difference across the p - n junction is reduced, and electrons (holes) flow easily. This is called the **forward bias configuration** of the diode. In sum, the diode allows current to flow freely in one direction but prevents current flow in the opposite direction. In this sense, the semiconductor diode is a one-way valve.

We can estimate the mathematical relationship between the current and voltage for a diode using the electric potential concept. Consider N negatively charged majority carriers (electrons donated by impurity atoms) in the n -type material and a potential barrier V across the p - n junction. According to the Maxwell-Boltzmann distribution, the fraction of electrons that have enough energy to diffuse across the potential barrier is $Ne^{-eV/k_B T}$. However, if a battery of voltage V_b is applied in the forward-bias configuration, this fraction improves to $Ne^{-e(V-V_b)/k_B T}$. The electric current due to the majority carriers from the n -side to the p -side is therefore

Equation:

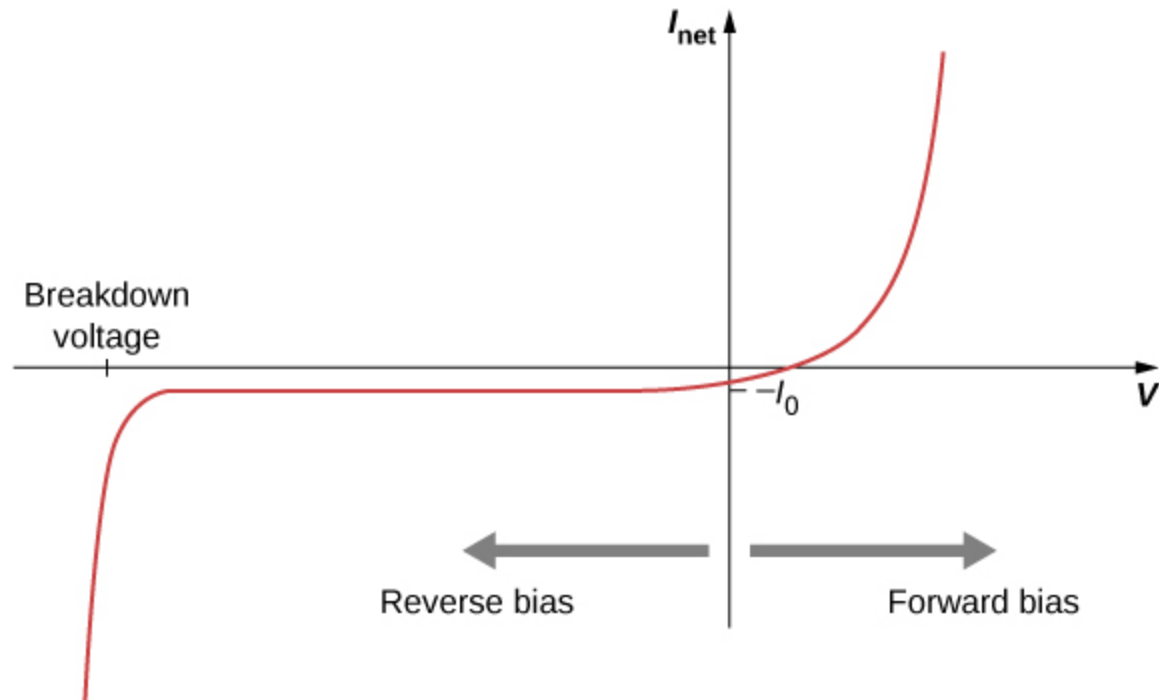
$$I = Ne^{-eV/k_B T} e^{eV_b/k_B T} = I_0 e^{eV_b/k_B T},$$

where I_0 is the current with no applied voltage and T is the temperature. Current due to the minority carriers (thermal excitation of electrons from the valence band to the conduction band on the p -side and subsequent attraction to the n -side) is $-I_0$, independent of the bias voltage. The net current is therefore

Note:
Equation:

$$I_{\text{net}} = I_0 \left(e^{eV_b/k_B T} - 1 \right).$$

A sample graph of the current versus bias voltage is given in [\[link\]](#). In the forward bias configuration, small changes in the bias voltage lead to large changes in the current. In the reverse bias configuration, the current is $I_{\text{net}} \approx -I_0$. For extreme values of reverse bias, the atoms in the material are ionized which triggers an avalanche of current. This case occurs at the **breakdown voltage**.



Current versus voltage across a p - n junction (diode). In the forward bias configuration, electric current flows easily. However, in the reverse bias configuration, electric current flow very little.

Example:**Diode Current**

Attaching the positive end of a battery to the p -side and the negative end to the n -side of a semiconductor diode produces a current of 4.5×10^{-1} A. The reverse saturation current is 2.2×10^{-8} A. (The reverse saturation current is the current of a diode in a reverse bias configuration such as this.) The battery voltage is 0.12 V. What is the diode temperature?

Strategy

The first arrangement is a forward bias configuration, and the second is the reverse bias configuration. In either case, [\[link\]](#) gives the current.

Solution

The current in the forward and reverse bias configurations is given by

Equation:

$$I_{\text{net}} = I_0 \left(e^{eV_b/k_B T} - 1 \right).$$

The current with no bias is related to the reverse saturation current by

Equation:

$$I_0 \approx -I_{\text{sat}} = 2.2 \times 10^{-8}.$$

Therefore

Equation:

$$\frac{I_{\text{net}}}{I_0} = \frac{4.5 \times 10^{-1} \text{ A}}{2.2 \times 10^{-8} \text{ A}} = 2.0 \times 10^8.$$

[\[link\]](#) can be written as

Equation:

$$\frac{I_{\text{net}}}{I_0} + 1 = e^{eV_b/k_B T}.$$

This ratio is much greater than one, so the second term on the left-hand side of the equation vanishes. Taking the natural log of both sides gives

Equation:

$$\frac{eV_b}{k_B T} = 19.$$

The temperature is therefore

Equation:

$$T = \frac{eV_b}{k_B} \left(\frac{1}{19} \right) = \frac{e(0.12 \text{ V})}{8.617 \times 10^{-5} \text{ eV/K}} \left(\frac{1}{19} \right) = 73 \text{ K}.$$

Significance

The current moving through a diode in the forward and reverse bias configuration is sensitive to the temperature of the diode. If the potential energy supplied by the battery is large compared to the thermal energy of the diode's surroundings, $k_B T$, then the forward bias current is very large compared to the reverse saturation current.

Note:

Exercise:

Problem:

Check Your Understanding How does the magnitude of the forward bias current compare with the reverse bias current?

Solution:

The forward bias current is much larger. To a good approximation, diodes permit current flow in only one direction.

Note:

Create a *p-n* junction and observe the behavior of a simple circuit for forward and reverse bias voltages. Visit this [site](#) to learn more about semiconductor diodes.

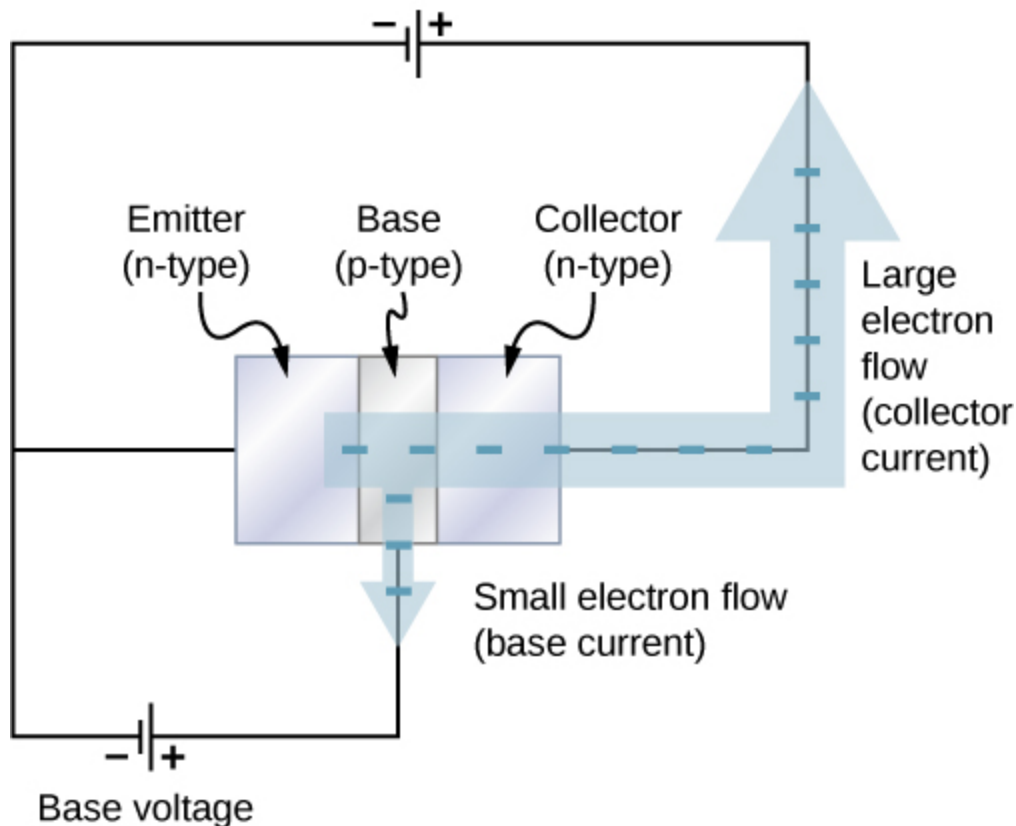
Junction Transistor

If diodes are one-way valves, transistors are one-way valves that can be carefully opened and closed to control current. A special kind of transistor is a junction transistor. A **junction transistor** has three parts, including an *n*-type semiconductor, also called the emitter; a thin *p*-type semiconductor, which is the base; and another *n*-type semiconductor, called the collector ([link](#)). When a positive terminal is connected to the *p*-type layer (the base), a small current of electrons, called the **base current** I_B , flows to the terminal. This causes a large **collector current** I_c to flow through the collector. The base current can be adjusted to control the large collector current. The current gain is therefore

Note:

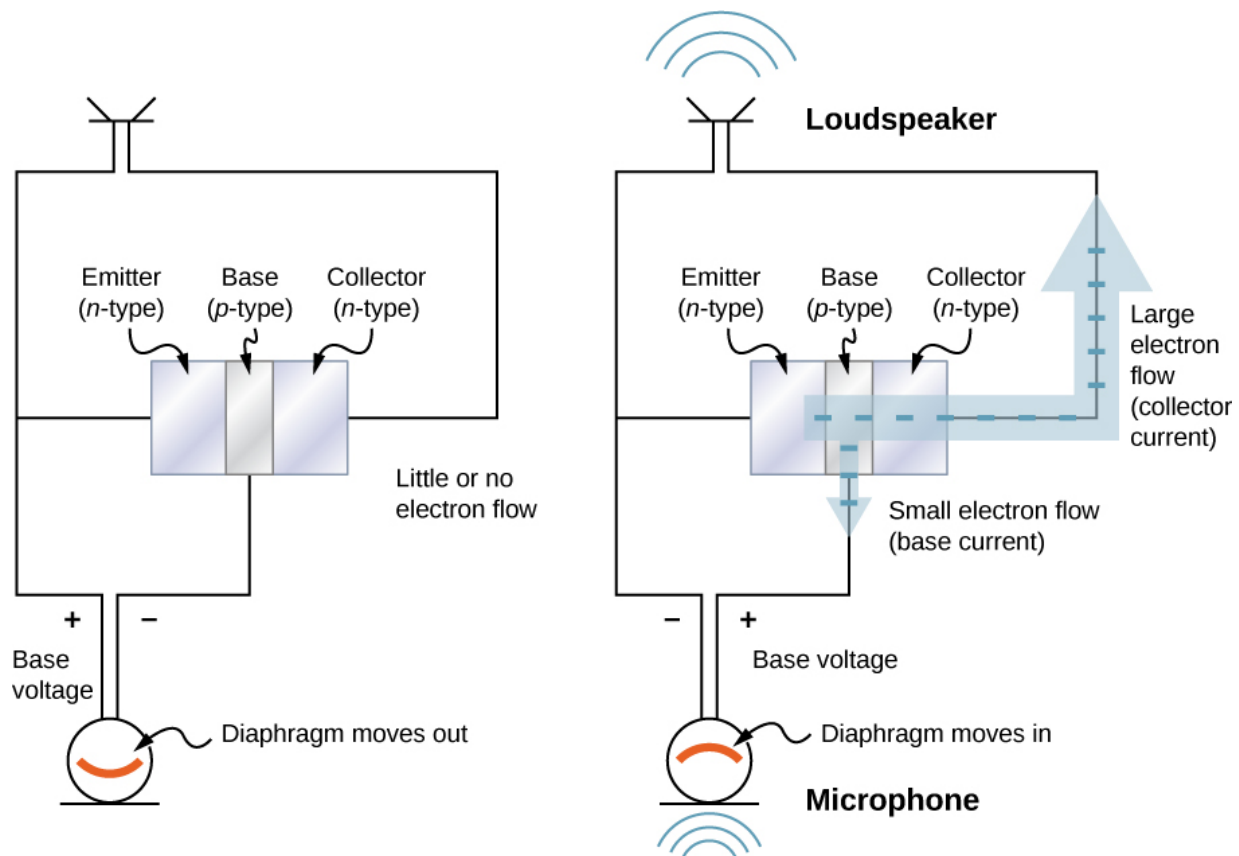
Equation:

$$I_c = \beta I_B.$$



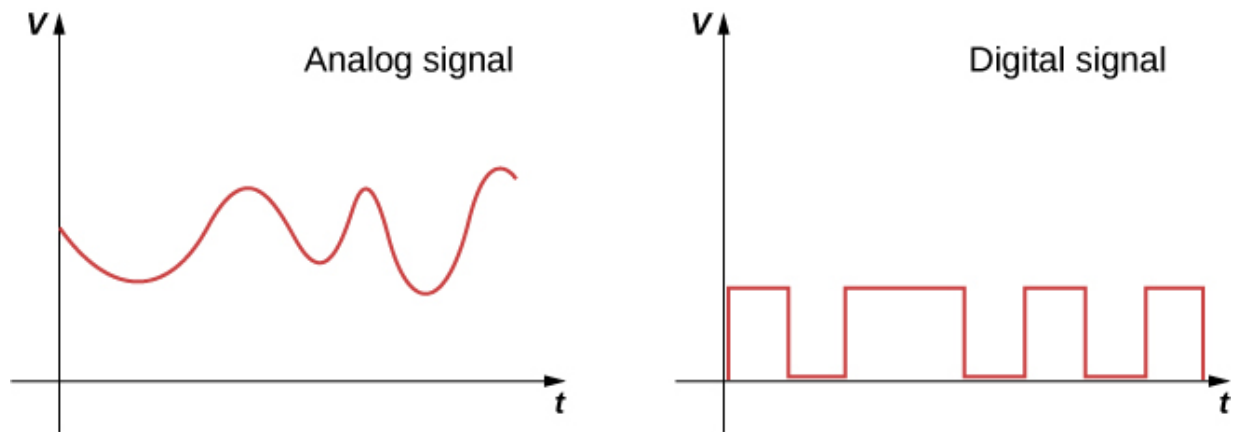
A junction transistor has three parts: emitter, base, and collector. Voltage applied to the base acts as a valve to control electric current from the emitter to the collector.

A junction transistor can be used to amplify the voltage from a microphone to drive a loudspeaker. In this application, sound waves cause a diaphragm inside the microphone to move in and out rapidly ([\[link\]](#)). When the diaphragm is in the “in” position, a tiny positive voltage is applied to the base of the transistor. This opens the transistor “valve” and allows a large electrical current flow to the loudspeaker. When the diaphragm is in the “out” position, a tiny negative voltage is applied to the base of the transistor, which shuts off the transistor valve so that no current flows to the loudspeaker. This shuts the transistor “valve” off so no current flows to the loudspeaker. In this way, current to the speaker is controlled by the sound waves, and the sound is amplified. Any electric device that amplifies a signal is called an **amplifier**.



An audio amplifier based on a junction transistor. Voltage applied to the base by a microphone acts as a valve to control a larger electric current that passes through a loudspeaker.

In modern electronic devices, digital signals are used with diodes and transistors to perform tasks such as data manipulation. Electric circuits carry two types of electrical signals: analog and digital ([\[link\]](#)). An analog signal varies continuously, whereas a digital signal switches between two fixed voltage values, such as plus 1 volt and zero volts. In digital circuits like those found in computers, a transistor behaves like an on-off switch. The transistor is either on, meaning the valve is completely open, or it is off, meaning the valve is completely closed. Integrated circuits contain vast collections of transistors on a single piece of silicon. They are designed to handle digital signals that represent ones and zeroes, which is also known as binary code. The invention of the ic helped to launch the modern computer revolution.



Real-world data are often analog, meaning data can vary continuously. Intensity values of sound or visual images are usually analog. These data are converted into digital signals for electronic processing in recording devices or computers. The digital signal is generated from the analog signal by requiring certain voltage cut-off value.

Summary

- A diode is produced by an n - p junction. A diode allows current to move in just one direction. In forward biased configuration of a diode, the current increases exponentially with the voltage.
- A transistor is produced by an n - p - n junction. A transistor is an electric valve that controls the current in a circuit.
- A transistor is a critical component in audio amplifiers, computers, and many other devices.

Conceptual Questions

Exercise:

Problem:

When p - and n -type materials are joined, why is a uniform electric field generated near the junction?

Exercise:**Problem:**

When p - and n -type materials are joined, why does the depletion layer not grow indefinitely?

Solution:

The electric field produced by the uncovered ions reduces further diffusion. In equilibrium, the diffusion and drift currents cancel so the net current is zero. Therefore, the resistance of the depletion region is large.

Exercise:**Problem:**

How do you know if a diode is in the *forward biased* configuration?

Exercise:**Problem:**

Why does the reverse bias configuration lead to a very small current?

Solution:

The positive terminal is applied to the n -side, which uncovers more ions near the junction (widens the depletion layer), increases the junction voltage difference, and therefore reduces the diffusion of holes across the junction.

Exercise:**Problem:**

What happens in the extreme case that where the n - and p -type materials are heavily doped?

Exercise:

Problem:

Explain how an audio amplifier works, using the transistor concept.

Solution:

Sound moves a diaphragm in and out, which varies the input or base current of the transistor circuit. The transistor amplifies this signal (*p-n-p* semiconductor). The output or collector current drives a speaker.

Problems**Exercise:**

Problem: Show that for V less than zero, $I_{\text{net}} \approx -I_0$.

Exercise:**Problem:**

A *p-n* diode has a reverse saturation current 1.44×10^{-8} A. It is forward biased so that it has a current of 6.78×10^{-1} A moving through it. What bias voltage is being applied if the temperature is 300 K?

Solution:

$$V_b = 0.458 \text{ V}$$

Exercise:**Problem:**

The collector current of a transistor is 3.4 A for a base current of 4.2 mA. What is the current gain?

Exercise:

Problem:

Applying the positive end of a battery to the p -side and the negative end to the n -side of a p - n junction, the measured current is 8.76×10^{-1} A. Reversing this polarity give a reverse saturation current of 4.41×10^{-8} A. What is the temperature if the bias voltage is 1.2 V?

Solution:

$$T = 829 \text{ K}$$

Exercise:**Problem:**

The base current of a transistor is 4.4 A, and its current gain 1126. What is the collector current?

Glossary

amplifier

electrical device that amplifies an electric signal

base current

current drawn from the base n -type material in a transistor

breakdown voltage

in a diode, the reverse bias voltage needed to cause an avalanche of current

collector current

current drawn from the collector p -type material

depletion layer

region near the p - n junction that produces an electric field

forward bias configuration

diode configuration that results in high current

junction transistor

electrical valve based on a $p-n-p$ junction

$p-n$ junction

junction formed by joining p - and n -type semiconductors

reverse bias configuration

diode configuration that results in low current

Superconductivity

By the end of this section, you will be able to:

- Describe the main features of a superconductor
- Describe the BCS theory of superconductivity
- Determine the critical magnetic field for $T = 0$ K from magnetic field data
- Calculate the maximum emf or current for a wire to remain superconducting

Electrical resistance can be considered as a measure of the frictional force in electrical current flow. Thus, electrical resistance is a primary source of energy dissipation in electrical systems such as electromagnets, electric motors, and transmission lines. Copper wire is commonly used in electrical wiring because it has one of the lowest room-temperature electrical resistivities among common conductors. (Actually, silver has a lower resistivity than copper, but the high cost and limited availability of silver outweigh its savings in energy over copper.)

Although our discussion of conductivity seems to imply that all materials must have electrical resistance, we know that this is not the case. When the temperature decreases below a critical value for many materials, their electrical resistivity drops to zero, and the materials become superconductors (see [Superconductors](#)).

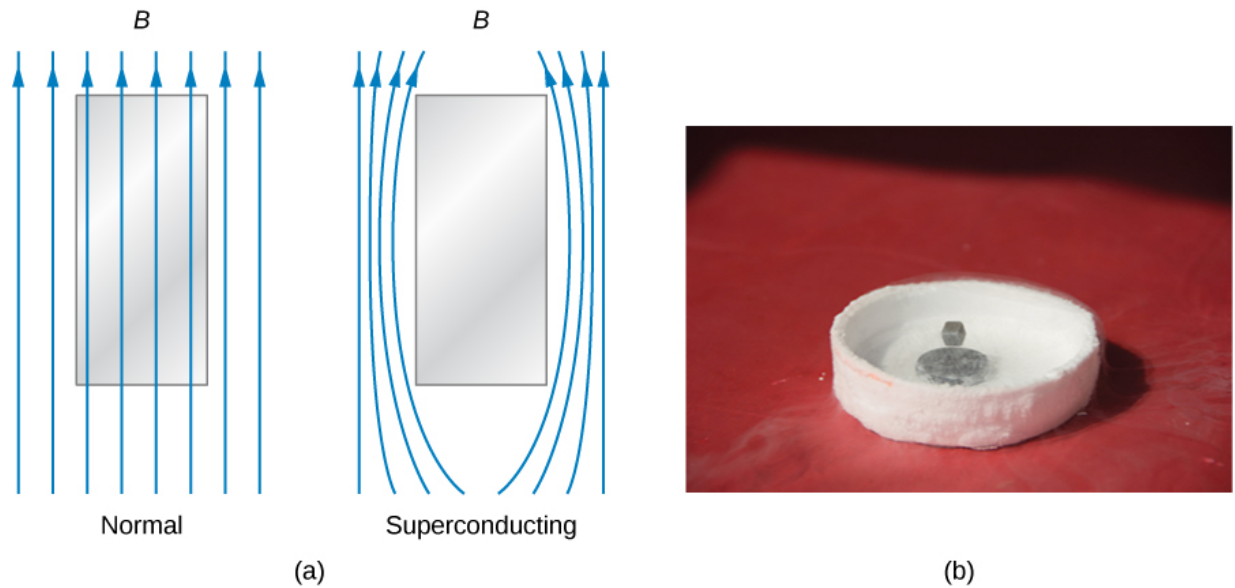
Note:

Watch this [NOVA video](#) excerpt, Making Stuff Colder, as an introduction to the topic of superconductivity and its many applications.

Properties of Superconductors

In addition to zero electrical resistance, superconductors also have perfect diamagnetism. In other words, in the presence of an applied magnetic field,

the net magnetic field within a superconductor is always zero ([\[link\]](#)). Therefore, any magnetic field lines that pass through a superconducting sample when it is in its normal state are expelled once the sample becomes superconducting. These are manifestations of the Meissner effect, which you learned about in the chapter on current and resistance.



(a) In the Meissner effect, a magnetic field is expelled from a material once it becomes superconducting. (b) A magnet can levitate above a superconducting material, supported by the force expelling the magnetic field. (credit b: modification of work by Kevin Jarrett)

Interestingly, the Meissner effect is not a consequence of the resistance being zero. To see why, suppose that a sample placed in a magnetic field undergoes a transition in which its resistance drops to zero. From Ohm's law, the current density, j , in the sample is related to the net internal electric field, E , and the resistivity ρ by $j = E/\rho$. If ρ is zero, E must also be zero so that j can remain finite. Now E and the magnetic flux Φ_m through the sample are related by Faraday's law as

Equation:

$$\oint E dI = -\frac{d\Phi_m}{dt}.$$

If E is zero, $d\Phi_m/dt$ is also zero, that is, the magnetic flux through the sample cannot change. The magnetic field lines within the sample should therefore not be expelled when the transition occurs. Hence, it does not follow that a material whose resistance goes to zero has to exhibit the Meissner effect. Rather, the Meissner effect is a special property of superconductors.

Another important property of a superconducting material is its **critical temperature**, T_c , the temperature below which the material is superconducting. The known range of critical temperatures is from a fraction of 1 K to slightly above 100 K. Superconductors with critical temperatures near this higher limit are commonly known as “high-temperature” superconductors. From a practical standpoint, superconductors for which $T_c \gg 77$ K are very important. At present, applications involving superconductors often still require that superconducting materials be immersed in liquid helium (4.2 K) in order to keep them below their critical temperature. The liquid helium baths must be continually replenished because of evaporation, and cooling costs can easily outweigh the savings in using a superconductor. However, 77 K is the temperature of liquid nitrogen, which is far more abundant and inexpensive than liquid helium. It would be much more cost-effective if we could easily fabricate and use high-temperature superconductor components that only need to be kept in liquid nitrogen baths to maintain their superconductivity.

High-temperature superconducting materials are presently in use in various applications. An example is the production of magnetic fields in some particle accelerators. The ultimate goal is to discover materials that are superconducting at room temperature. Without any cooling requirements, the bulk of electronic components and transmission lines could be superconducting, resulting in dramatic and unprecedented increases in efficiency and performance.

Another important property of a superconducting material is its **critical magnetic field** $B_c(T)$, which is the maximum applied magnetic field at a

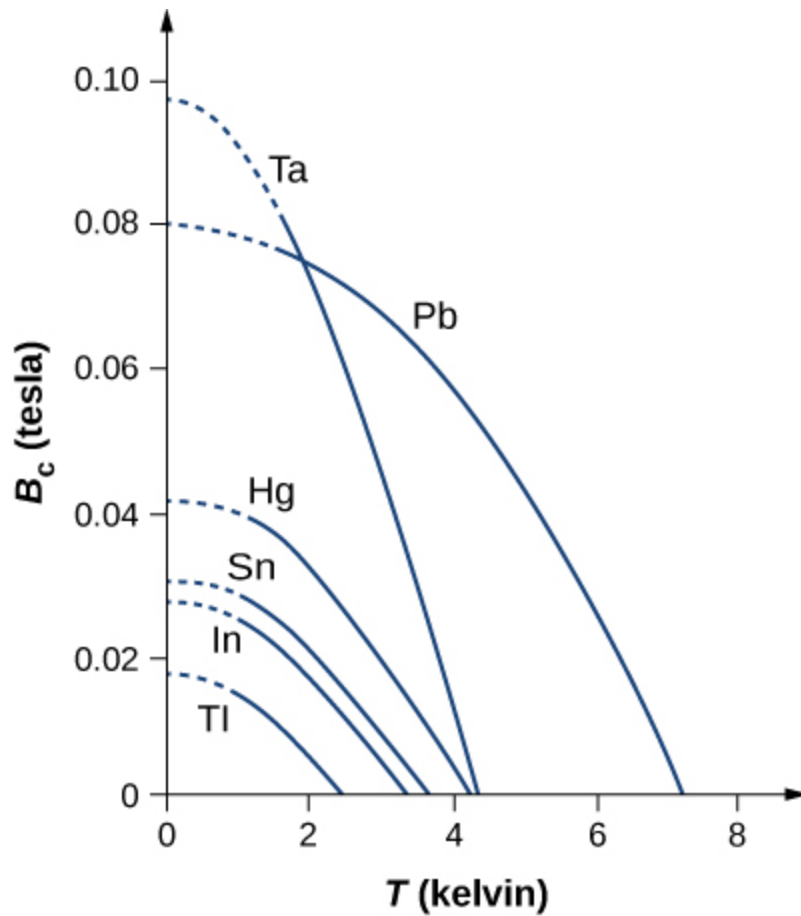
temperature T that will allow a material to remain superconducting. An applied field that is greater than the critical field will destroy the superconductivity. The critical field is zero at the critical temperature and increases as the temperature decreases. Plots of the critical field versus temperature for several superconducting materials are shown in [\[link\]](#). The temperature dependence of the critical field can be described approximately by

Note:

Equation:

$$B_c(T) = B_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right]$$

where $B_c(0)$ is the critical field at absolute zero temperature. [\[link\]](#) lists the critical temperatures and fields for two classes of superconductors: **type I superconductor** and **type II superconductor**. In general, type I superconductors are elements, such as aluminum and mercury. They are perfectly diamagnetic below a critical field $B_C(T)$, and enter the normal non-superconducting state once that field is exceeded. The critical fields of type I superconductors are generally quite low (well below one tesla). For this reason, they cannot be used in applications requiring the production of high magnetic fields, which would destroy their superconducting state.



The temperature dependence of the critical field for several superconductors.

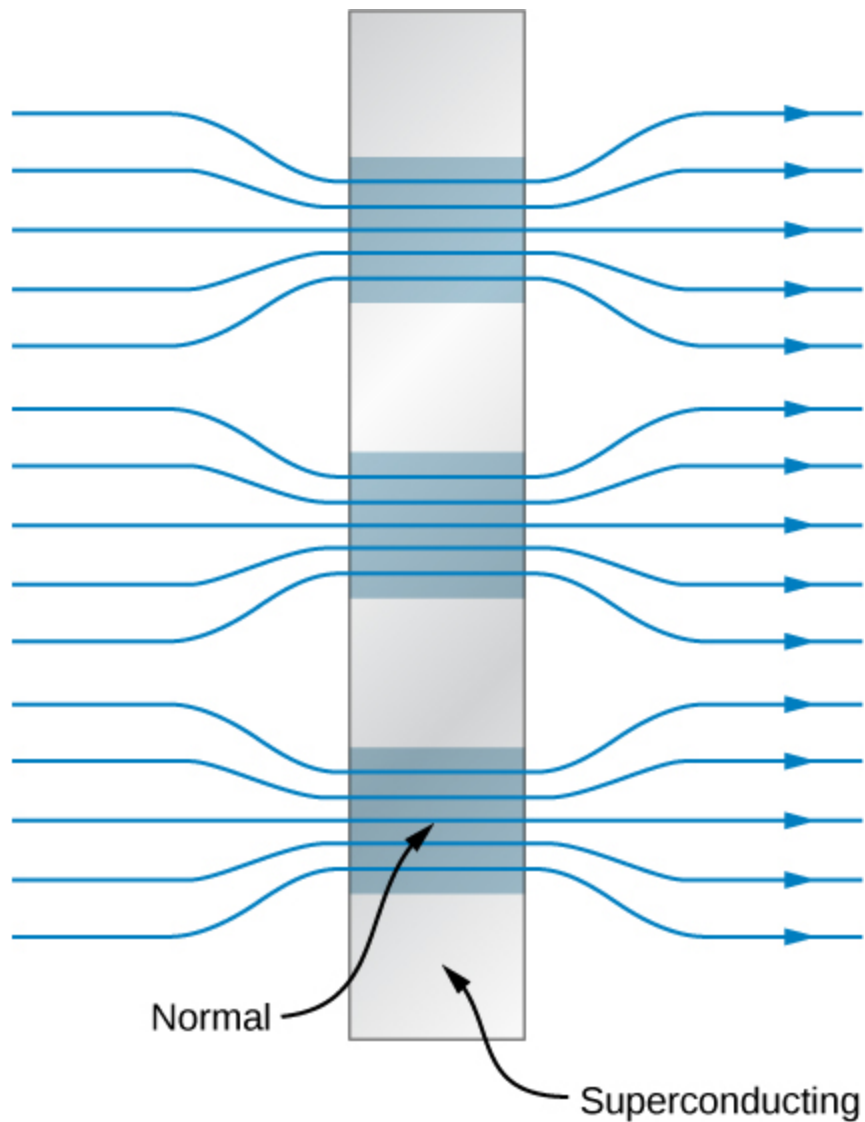
Superconductivity occurs for magnetic fields and temperatures below the curves shown.

Material	Critical Temperature (K)	Critical Magnetic Field (T)
----------	--------------------------	-----------------------------

Material	Critical Temperature (K)	Critical Magnetic Field (T)
Type I		
Al	1.2	0.011
Ga	1.1	0.0051
Hg (α)	4.2	0.041
In	3.4	0.029
Nb	9.3	0.20
Pb	7.2	0.080
Sn	3.7	0.031
Th	1.4	0.00016
Zn	0.87	0.0053
Type II		
Nb ₃ Al	18	32
Nb ₃ Ge	23	38
Nb ₃ Sn	18	25
NbTi	9.3	15
YBa ₂ Cu ₃ O ₇	92	>100

Critical Temperature and Critical Magnetic Field at $T = 0$ K for Various Superconductors

Type II superconductors are generally compounds or alloys involving transition metals or actinide series elements. Almost all superconductors with relatively high critical temperatures are type II. They have two critical fields, represented by $B_{c1}(T)$ and $B_{c2}(T)$. When the field is below $B_{c1}(T)$, type II superconductors are perfectly diamagnetic, and no magnetic flux penetration into the material can occur. For a field exceeding $B_{c2}(T)$, they are driven into their normal state. When the field is greater than $B_{c1}(T)$ but less than $B_{c2}(T)$, type II superconductors are said to be in a mixed state. Although there is some magnetic flux penetration in the mixed state, the resistance of the material is zero. Within the superconductor, filament-like regions exist that have normal electrical and magnetic properties interspersed between regions that are superconducting with perfect diamagnetism. A representation of this state is given in [\[link\]](#). The magnetic field is expelled from the superconducting regions but exists in the normal regions. In general, $B_{c2}(T)$ is very large compared with the critical fields of type I superconductors, so wire made of type II superconducting material is suitable for the windings of high-field magnets.



A schematic representation of the mixed state of a type II superconductor. Superconductors (the gray squares) expel magnetic fields in their vicinity.

Example:
Niobium Wire

In an experiment, a niobium (Nb) wire of radius 0.25 mm is immersed in liquid helium ($T = 4.2$ K) and required to carry a current of 300 A. Does the wire remain superconducting?

Strategy

The applied magnetic field can be determined from the radius of the wire and current. The critical magnetic field can be determined from [\[link\]](#), the properties of the superconductor, and the temperature. If the applied magnetic field is greater than the critical field, then superconductivity in the Nb wire is destroyed.

Solution

At $T = 4.2$ K, the critical field for Nb is, from [\[link\]](#) and [\[link\]](#),

Equation:

$$B_c(4.2 \text{ K}) = B_c(0) \left[1 - \left(\frac{4.2 \text{ K}}{9.3 \text{ K}} \right)^2 \right] = (0.20 \text{ T})(0.80) = 0.16 \text{ T}.$$

In an earlier chapter, we learned the magnetic field inside a current-carrying wire of radius a is given by

Equation:

$$B = \frac{\mu_0 I r}{2\pi a},$$

where r is the distance from the central axis of the wire. Thus, the field at the surface of the wire is $\frac{\mu_0 I r}{2\pi a}$. For the niobium wire, this field is

Equation:

$$B = \frac{(4\pi \times 10^{-7} \text{ T m/A})(300 \text{ A})}{2\pi(2.5 \times 10^{-4} \text{ m})} = 0.24 \text{ T}.$$

Since this exceeds the critical 0.16 T, the wire does not remain superconducting.

Significance

Superconductivity requires low temperatures and low magnetic fields. These simultaneous conditions are met less easily for Nb than for many

other metals. For example, aluminum superconducts at temperatures 7 times lower and magnetic fields 18 times lower.

Note:

Exercise:

Problem:

Check Your Understanding What conditions are necessary for superconductivity?

Solution:

a low temperature and low magnetic field

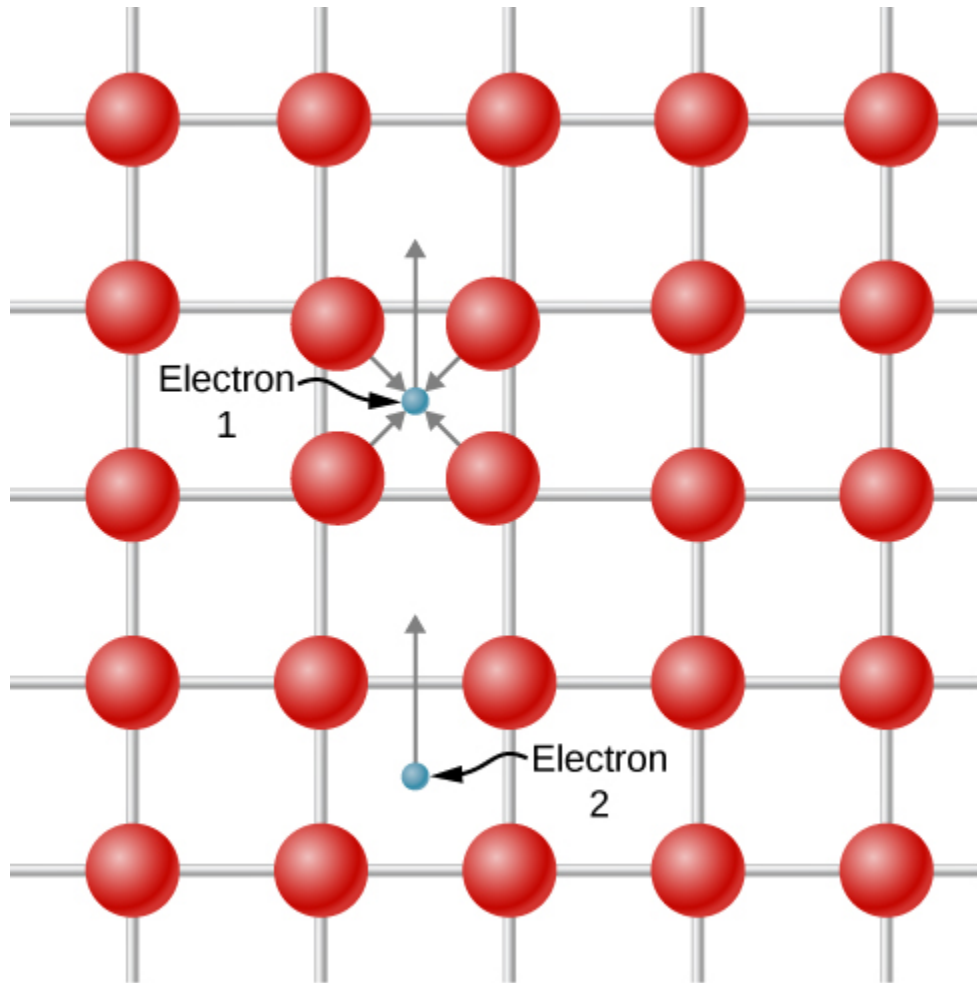
Theory of Superconductors

A successful theory of superconductivity was developed in the 1950s by John Bardeen, Leon Cooper, and J. Robert Schrieffer, for which they received the Nobel Prize in 1972. This theory is known as the **BCS theory**. BCS theory is complex, so we summarize it qualitatively below.

In a normal conductor, the electrical properties of the material are due to the most energetic electrons near the Fermi energy. In 1956, Cooper showed that if there is any attractive interaction between two electrons at the Fermi level, then the electrons can form a bound state in which their total energy is less than $2E_F$. Two such electrons are known as a **Cooper pair**.

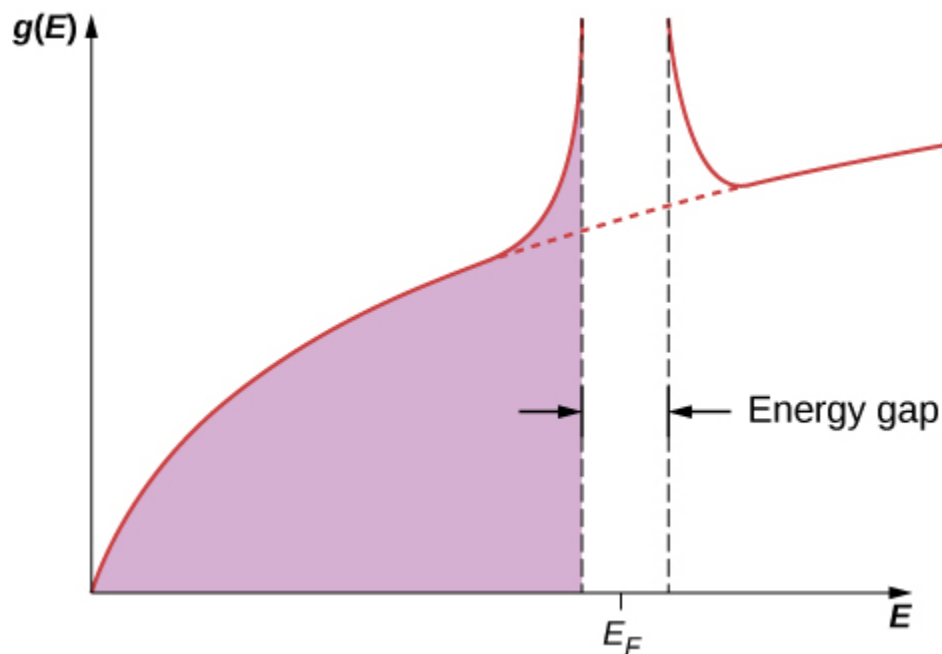
It is hard to imagine two electrons attracting each other, since they have like charge and should repel. However, the proposed interaction occurs only in the context of an atomic lattice. A depiction of the attraction is shown in [\[link\]](#). Electron 1 slightly displaces the positively charged atomic nuclei toward itself as it travels past because of the Coulomb attraction. Electron 2 “sees” a region with a higher density of positive charge relative to the

surroundings and is therefore attracted into this region and, therefore indirectly, to electron 1. Because of the exclusion principle, the two electrons of a Cooper pair must have opposite spin.



A Cooper pair can form as a result of the displacement of positive atomic nuclei. Electron 1 slightly displaces the positively charged atomic nuclei toward itself as it travels past because of the Coulomb attraction. Electron 2 “sees” a region with a higher density of positive charge relative to the surroundings and is therefore attracted into this region.

The BCS theory extends Cooper's ideas, which are for a single pair of electrons, to the entire free electron gas. When the transition to the superconducting state occurs, all the electrons pair up to form Cooper pairs. On an atomic scale, the distance between the two electrons making up a Cooper pair is quite large. Between these electrons are typically about 10^6 other electrons, each also pairs with a distant electron. Hence, there is considerable overlap between the wave functions of the individual Cooper pairs, resulting in a strong correlation among the motions of the pairs. They all move together "in step," like the members of a marching band. In the superconducting transition, the density of states becomes drastically changed near the Fermi level. As shown in [\[link\]](#), an energy gap appears around E_F because the collection of Cooper pairs has lower ground state energy than the Fermi gas of noninteracting electrons. The appearance of this gap characterizes the superconducting state. If this state is destroyed, then the gap disappears, and the density of states reverts to that of the free electron gas.



A relatively large energy gap is formed around the Fermi energy when a material becomes superconducting. If this state is destroyed, then the

gap disappears, and the density of states reverts to that of the free electron gas.

The BCS theory is able to predict many of the properties observed in superconductors. Examples include the Meissner effect, the critical temperature, the critical field, and, perhaps most importantly, the resistivity becoming zero at a critical temperature. We can think about this last phenomenon qualitatively as follows. In a normal conductor, resistivity results from the interaction of the conduction electrons with the lattice. In this interaction, the energy exchanged is on the order of $k_B T$, the thermal energy. In a superconductor, electric current is carried by the Cooper pairs. The only way for a lattice to scatter a Cooper pair is to break it up. The destruction of one pair then destroys the collective motion of all the pairs. This destruction requires energy on the order of 10^{-3} eV , which is the size of the energy gap. Below the critical temperature, there is not enough thermal energy available for this process, so the Cooper pairs travel unimpeded throughout the superconductor.

Finally, it is interesting to note that no evidence of superconductivity has been found in the best normal conductors, such as copper and silver. This is not unexpected, given the BCS theory. The basis for the formation of the superconducting state is an interaction between the electrons and the lattice. In the best conductors, the electron-lattice interaction is weakest, as evident from their minimal resistivity. We might expect then that in these materials, the interaction is so weak that Cooper pairs cannot be formed, and superconductivity is therefore precluded.

Summary

- A superconductor is characterized by two features: the conduction of electrons with zero electrical resistance and the repelling of magnetic field lines.
- A minimum temperature is required for superconductivity to occur.
- A strong magnetic field destroys superconductivity.
- Superconductivity can be explain in terms of Cooper pairs.

Key Equations

Electrostatic energy for equilibrium separation distance between atoms	$U_{\text{coul}} = -\frac{ke^2}{r_0}$
Energy change associated with ionic bonding	$U_{\text{form}} = E_{\text{transfer}} + U_{\text{coul}} + U_{\text{ex}}$
Critical magnetic field of a superconductor	$B_c(T) = B_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right]$
Rotational energy of a diatomic molecule	$E_r = l(l+1) \frac{\hbar^2}{2I}$
Characteristic rotational energy of a molecule	$E_{0r} = \frac{\hbar^2}{2I}$
Potential energy associated with the exclusion principle	$U_{\text{ex}} = \frac{A}{r^n}$
Dissociation energy of a solid	$U_{\text{diss}} = \alpha \frac{ke^2}{r_0} \left(1 - \frac{1}{n} \right)$
Moment of inertia of a diatomic molecule with reduced mass μ	$I = \mu r_0^2$
Electron energy in a metal	$E = \frac{\pi^2 \hbar^2}{2mL^2} (n_1^2 + n_2^2 + n_3^2)$
Electron density of states of a metal	$g(E) = \frac{\pi V}{2} \left(\frac{8m_e}{h^2} \right)^{3/2} E^{1/2}$

Fermi energy	$E_F = \frac{h^2}{8m_e} \left(\frac{3N}{\pi V} \right)^{2/3}$
Fermi temperature	$T_F = \frac{E_F}{k_B}$
Hall effect	$V_H = uBw$
Current versus bias voltage across p - n junction	$I_{\text{net}} = I_0 (e^{eV_b/k_B T} - 1)$
Current gain	$I_c = \beta I_B$
Selection rule for rotational energy transitions	$\Delta l = \pm 1$
Selection rule for vibrational energy transitions	$\Delta n = \pm 1$

Conceptual Questions

Exercise:

Problem: Describe two main features of a superconductor.

Exercise:

Problem: How does BCS theory explain superconductivity?

Solution:

BCS theory explains superconductivity in terms of the interactions between electron pairs (Cooper pairs). One electron in a pair interacts with the lattice, which interacts with the second electron. The combined electron-lattice-electron interaction binds the electron pair together in a way that overcomes their mutual repulsion.

Exercise:

Problem: What is the Meissner effect?

Exercise:

Problem:

What impact does an increasing magnetic field have on the critical temperature of a semiconductor?

Solution:

As the magnitude of the magnetic field is increased, the critical temperature decreases.

Problems

Exercise:

Problem:

At what temperature, in terms of T_C , is the critical field of a superconductor one-half its value at $T = 0$ K ?

Solution:

$$T = 0.707 T_c$$

Exercise:

Problem: What is the critical magnetic field for lead at $T = 2.8$ K ?

Exercise:

Problem:

A Pb wire wound in a tight solenoid of diameter of 4.0 mm is cooled to a temperature of 5.0 K. The wire is connected in series with a $50\text{-}\Omega$ resistor and a variable source of emf. As the emf is increased, what value does it have when the superconductivity of the wire is destroyed?

Solution:

61 kV

Exercise:**Problem:**

A tightly wound solenoid at 4.0 K is 50 cm long and is constructed from Nb wire of radius 1.5 mm. What maximum current can the solenoid carry if the wire is to remain superconducting?

Additional Problems**Exercise:****Problem:**

Potassium fluoride (KF) is a molecule formed by an ionic bond. At equilibrium separation the atoms are $r_0 = 0.255\text{ nm}$ apart. Determine the electrostatic potential energy of the atoms. The electron affinity of F is 3.40 eV and the ionization energy of K is 4.34 eV. Determine dissociation energy. (Neglect the energy of repulsion.)

Solution:

$$U_{\text{coul}} = -5.65\text{ eV}$$

$$E_{\text{form}} = -4.71\text{ eV}, E_{\text{diss}} = 4.71\text{ eV}$$

Exercise:

Problem:

For the preceding problem, sketch the potential energy versus separation graph for the bonding of K^+ and $F1^-$ ions. (a) Label the graph with the energy required to transfer an electron from K to Fl. (b) Label the graph with the dissociation energy.

Exercise:**Problem:**

The separation between hydrogen atoms in a H_2 molecule is about 0.075 nm. Determine the characteristic energy of rotation in eV.

Solution:

$$E_{0r} = 7.43 \times 10^{-3} \text{ eV}$$

Exercise:**Problem:**

The characteristic energy of the Cl_2 molecule is $2.95 \times 10^{-5} \text{ eV}$. Determine the separation distance between the nitrogen atoms.

Exercise:

Problem: Determine the lowest three rotational energy levels of H_2 .

Solution:

$$E_{0r} = 7.43 \times 10^{-3} \text{ eV}; l = 0; E_r = 0 \text{ eV (no rotation);}$$
$$l = 1; E_r = 1.49 \times 10^{-2} \text{ eV}; l = 2; E_r = 4.46 \times 10^{-2} \text{ eV}$$

Exercise:**Problem:**

A carbon atom can hybridize in the sp^2 configuration. (a) What is the angle between the hybrid orbitals?

Exercise:**Problem:**

List five main characteristics of ionic crystals that result from their high dissociation energy.

Solution:

- i. They are fairly hard and stable.
- ii. They vaporize at relatively high temperatures (1000 to 2000 K).
- iii. They are transparent to visible radiation, because photons in the visible portion of the spectrum are not energetic enough to excite an electron from its ground state to an excited state.
- iv. They are poor electrical conductors because they contain effectively no free electrons.
- v. They are usually soluble in water, because the water molecule has a large dipole moment whose electric field is strong enough to break the electrostatic bonds between the ions.

Exercise:**Problem:**

Why is bonding in H_2^+ favorable? Express your answer in terms of the symmetry of the electron wave function.

Exercise:**Problem:**

Astronomers claim to find evidence of He_2 from light spectra of a distant star. Do you believe them?

Solution:

No, He atoms do not contain valence electrons that can be shared in the formation of a chemical bond.

Exercise:

Problem:

Show that the moment of inertia of a diatomic molecule is $I = \mu r_0^2$, where μ is the reduced mass, and r_0 is the distance between the masses.

Exercise:**Problem:**

Show that the average energy of an electron in a one-dimensional metal is related to the Fermi energy by $\bar{E} = \frac{1}{2} E_F$.

Solution:

$$\sum_1^{N/2} n^2 = \frac{1}{3} \left(\frac{N}{2} \right)^3, \text{ so } \bar{E} = \frac{1}{3} E_F$$

Exercise:**Problem:**

Measurements of a superconductor's critical magnetic field (in T) at various temperatures (in K) are given below. Use a line of best fit to determine $B_c(0)$. Assume $T_c = 9.3$ K.

T (in K)	$B_c(T)$
3.0	0.18
4.0	0.16
5.0	0.14

T (in K)	$B_c(T)$
6.0	0.12
7.0	0.09
8.0	0.05
9.0	0.01

Exercise:

Problem:

Estimate the fraction of Si atoms that must be replaced by As atoms in order to form an impurity band.

Solution:

An impurity band will be formed when the density of the donor atoms is high enough that the orbits of the extra electrons overlap. We saw earlier that the orbital radius is about 50 Angstroms, so the maximum distance between the impurities for a band to form is 100 Angstroms. Thus if we use 1 Angstrom as the interatomic distance between the Si atoms, we find that 1 out of 100 atoms along a linear chain must be a donor atom. And in a three-dimensional crystal, roughly 1 out of 10^6 atoms must be replaced by a donor atom in order for an impurity band to form.

Exercise:

Problem:

Transition in the rotation spectrum are observed at ordinary room temperature ($T = 300$ K). According to your lab partner, a peak in the spectrum corresponds to a transition from the $l = 4$ to the $l = 1$ state. Is this possible? If so, determine the momentum of inertia of the molecule.

Exercise:**Problem:**

Determine the Fermi energies for (a) Mg, (b) Na, and (c) Zn.

Solution:

a. $E_F = 7.11$ eV; b. $E_F = 3.24$ eV; c. $E_F = 9.46$ eV

Exercise:

Problem: Find the average energy of an electron in a Zn wire.

Exercise:**Problem:**

What value of the repulsion constant, n , gives the measured dissociation energy of 158 kcal/mol for CsCl?

Solution:

$9.15 \approx 9$

Exercise:**Problem:**

A physical model of a diamond suggests a BCC packing structure. Why is this not possible?

Challenge Problems**Exercise:**

Problem:

For an electron in a three-dimensional metal, show that the average

energy is given by $\bar{E} = \frac{1}{N} \int_0^{E_F} E g(E) dE = \frac{3}{5} E_F$,

Where N is the total number electrons in the metal.

Solution:

In three dimensions, the energy of an electron is given by:

$E = R^2 E_1$, where $R^2 = n_1^2 + n_2^2 + n_3^2$. Each allowed energy state corresponds to node in N space (n_1, n_2, n_3) . The number of particles corresponds to the number of states (nodes) in the first octant, within a sphere of radius, R . This number is given by: $N = 2 \left(\frac{1}{8}\right) \left(\frac{4}{3}\right) \pi R^3$, where the factor 2 accounts for two states of spin. The density of states is found by differentiating this expression by energy:

$$g(E) = \frac{\pi V}{2} \left(\frac{8m_e}{h^2}\right)^{3/2} E^{1/2}. \text{ Integrating gives: } \bar{E} = \frac{3}{5} E_F.$$

Glossary**BCS theory**

theory of superconductivity based on electron-lattice-electron interactions

Cooper pair

coupled electron pair in a superconductor

critical magnetic field

maximum field required to produce superconductivity

critical temperature

maximum temperature to produce superconductivity

type I superconductor

superconducting element, such as aluminum or mercury

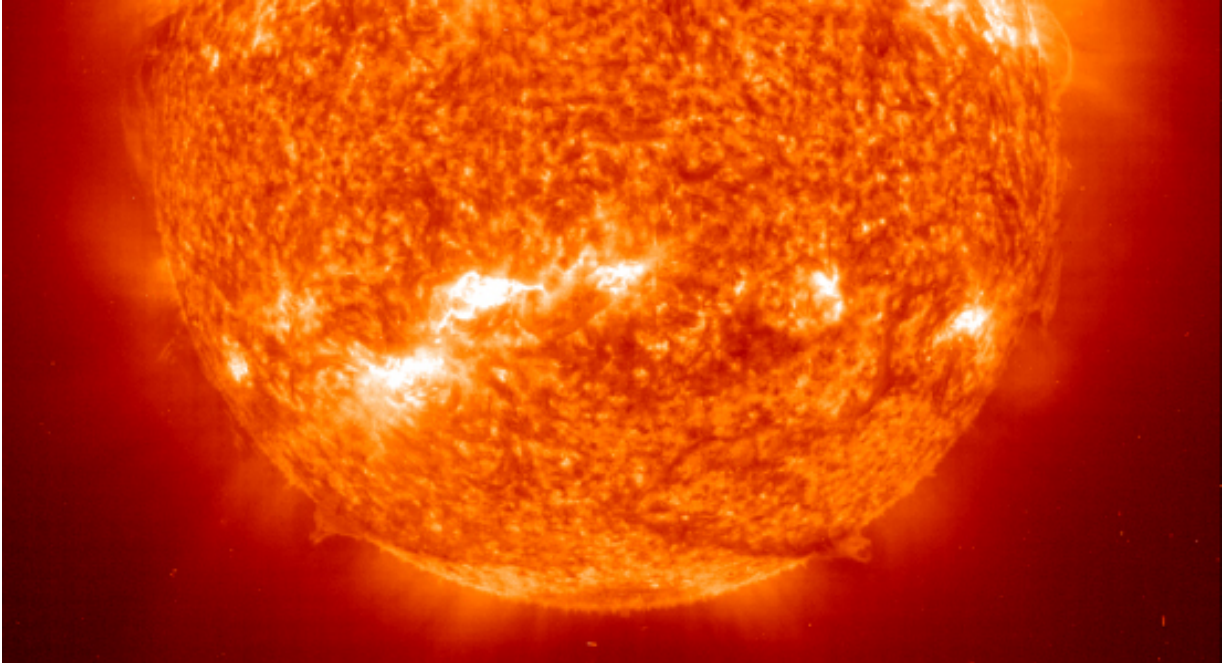
type II superconductor

superconducting compound or alloy, such as a transition metal or an actinide series element

Introduction

class="introduction"

The Sun is powered by nuclear fusion in its core. The core converts approximately 10^{38} protons/second into helium at a temperature of 14 million K. This process releases energy in the form of photons, neutrinos, and other particles.
(credit: modification of work by EIT SOHO Consortium, ESA, NASA)



In this chapter, we study the composition and properties of the atomic nucleus. The nucleus lies at the center of an atom, and consists of protons and neutrons. A deep understanding of the nucleus leads to numerous valuable technologies, including devices to date ancient rocks, map the galactic arms of the Milky Way, and generate electrical power.

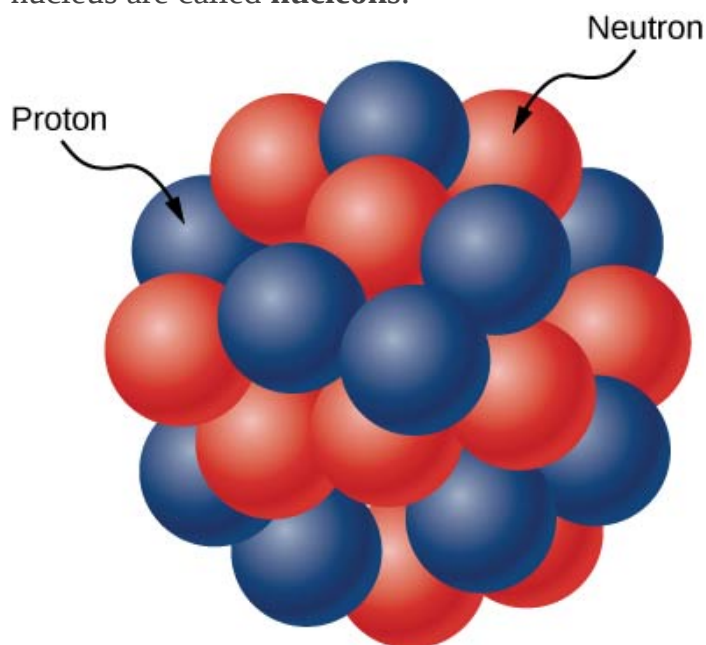
The Sun is the main source of energy in the solar system. The Sun is 109 Earth diameters across, and accounts for more than 99% of the total mass of the solar system. The Sun shines by fusing hydrogen nuclei—protons—deep inside its interior. Once this fuel is spent, the Sun will burn helium and, later, other nuclei. Nuclear fusion in the Sun is discussed toward the end of this chapter. In the meantime, we will investigate nuclear properties that govern all nuclear processes, including fusion.

Properties of Nuclei

By the end of this section, you will be able to:

- Describe the composition and size of an atomic nucleus
- Use a nuclear symbol to express the composition of an atomic nucleus
- Explain why the number of neutrons is greater than protons in heavy nuclei
- Calculate the atomic mass of an element given its isotopes

The **atomic nucleus** is composed of **protons** and **neutrons** ([link](#)). Protons and neutrons have approximately the same mass, but protons carry one unit of positive charge ($+e$), and neutrons carry no charge. These particles are packed together into an extremely small space at the center of an atom. According to scattering experiments, the nucleus is spherical or ellipsoidal in shape, and about $1/100,000$ th the size of a hydrogen atom. If an atom were the size of a major league baseball stadium, the nucleus would be roughly the size of the baseball. Protons and neutrons within the nucleus are called **nucleons**.



The atomic nucleus is composed of protons and neutrons. Protons are shown in blue, and neutrons are shown in red.

Counts of Nucleons

The number of protons in the nucleus is given by the **atomic number**, Z . The number of neutrons in the nucleus is the **neutron number**, N . The total number of nucleons is the **mass number**, A . These numbers are related by

Note:

Equation:

$$A = Z + N.$$

A nucleus is represented symbolically by

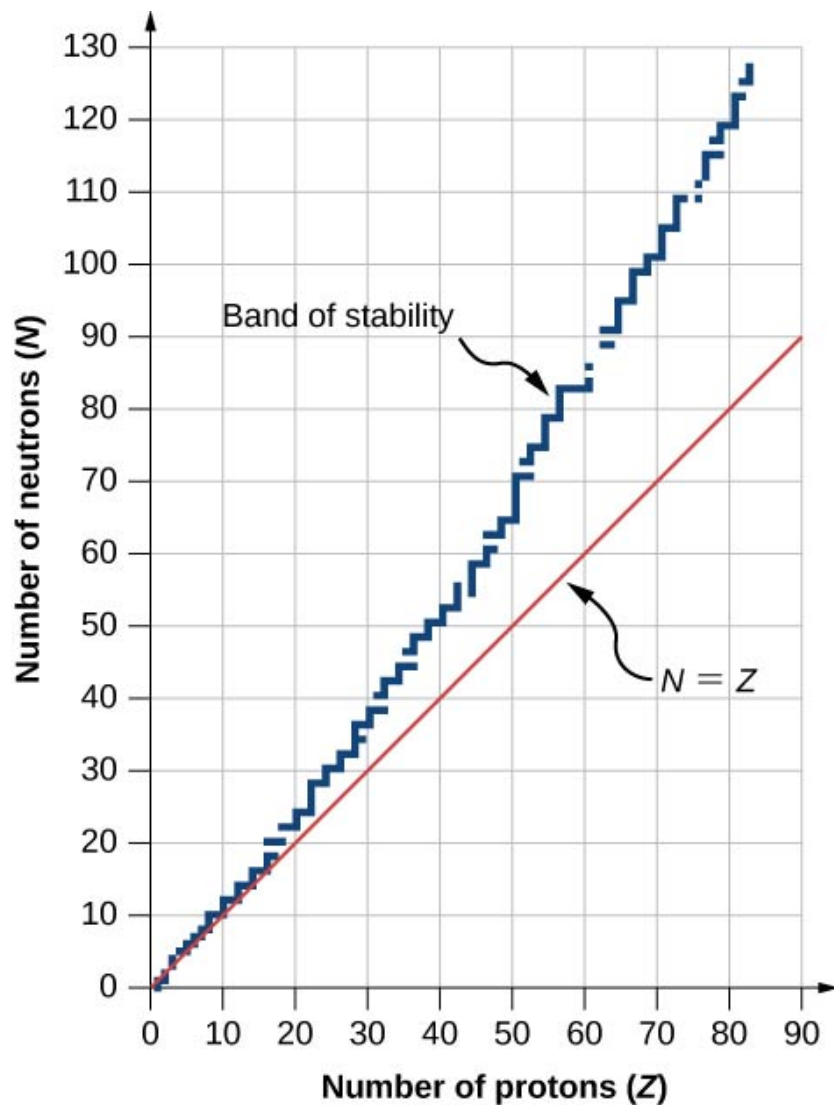
Note:

Equation:

$${}^A_Z\text{X},$$

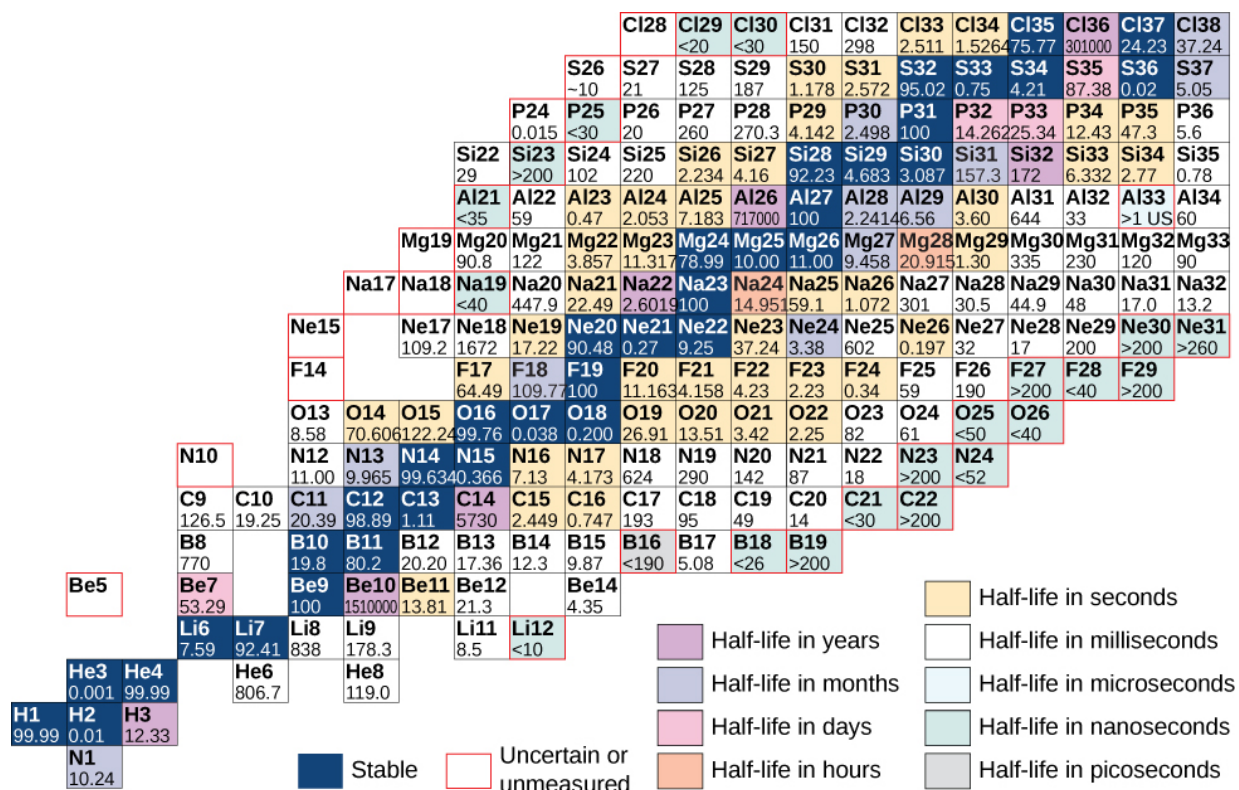
where X represents the chemical element, A is the mass number, and Z is the atomic number. For example, ${}^{12}_6\text{C}$ represents the carbon nucleus with six protons and six neutrons (or 12 nucleons).

A graph of the number N of neutrons versus the number Z of protons for a range of stable nuclei (**nuclides**) is shown in [\[link\]](#). For a given value of Z , multiple values of N (blue points) are possible. For small values of Z , the number of neutrons equals the number of protons ($N = P$), and the data fall on the red line. For large values of Z , the number of neutrons is greater than the number of protons ($N > P$), and the data points fall above the red line. The number of neutrons is generally greater than the number of protons for $Z > 15$.



This graph plots the number of neutrons N against the number of protons Z for stable atomic nuclei. Larger nuclei, have more neutrons than protons.

A chart based on this graph that provides more detailed information about each nucleus is given in [\[link\]](#). This chart is called a **chart of the nuclides**. Each cell or tile represents a separate nucleus. The nuclei are arranged in order of ascending Z (along the horizontal direction) and ascending N (along the vertical direction).



Partial chart of the nuclides. For stable nuclei (dark blue backgrounds), cell values represent the percentage of nuclei found on Earth with the same atomic number (percent abundance). For the unstable nuclei, the number represents the half-life.

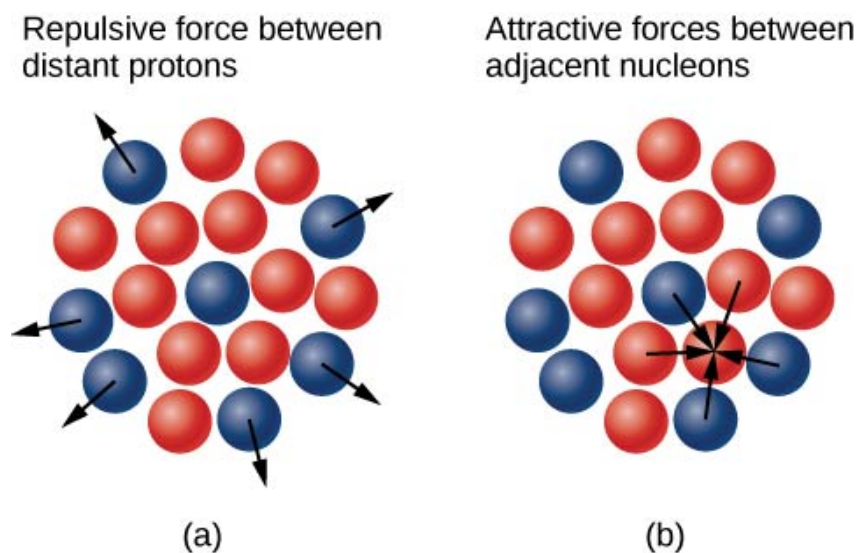
Atoms that contain nuclei with the same number of protons (Z) and different numbers of neutrons (N) are called **isotopes**. For example, hydrogen has three isotopes: normal hydrogen (1 proton, no neutrons), deuterium (one proton and one neutron), and tritium (one proton and two neutrons). Isotopes of a given atom share the same chemical properties, since these properties are determined by interactions between the outer electrons of the atom, and not the nucleons. For example, water that contains deuterium rather than hydrogen (“heavy water”) looks and tastes like normal water. The following table shows a list of common isotopes.

Element	Symbol	Mass Number	Mass (Atomic Mass Units)	Percent Abundance*	Half-life**
Hydrogen	H	1	1.0078	99.99	stable
	^2H or D	2	2.0141	0.01	stable
	^3H	3	3.0160	–	12.32 y
Carbon	^{12}C	12	12.0000	98.91	stable
	^{13}C	13	13.0034	1.1	stable
	^{14}C	14	14.0032	–	5730 y
Nitrogen	^{14}N	14	14.0031	99.6	stable
	^{15}N	15	15.0001	0.4	stable
	^{16}N	16	16.0061	–	7.13 s
Oxygen	^{16}O	16	15.9949	99.76	stable
	^{17}O	17	16.9991	0.04	stable
	^{18}O	18	17.9992	0.20	stable
	^{19}O	19	19.0035	–	26.46 s

Common Isotopes*No entry if less than 0.001 (trace amount).

Why do neutrons outnumber protons in heavier nuclei ([link](#))? The answer to this question requires an understanding of forces inside the nucleus. Two types of forces exist: (1) the long-range electrostatic (Coulomb) force that makes the positively charged protons repel one another; and (2) the short-range **strong nuclear force** that makes all nucleons in the nucleus attract one another. You may also have heard of a

“weak” nuclear force. This force is responsible for some nuclear decays, but as the name implies, it does not play a role in stabilizing the nucleus against the strong Coulomb repulsion it experiences. We discuss strong nuclear force in more detail in the next chapter when we cover particle physics. Nuclear stability occurs when the attractive forces between nucleons compensate for the repulsive, long-range electrostatic forces between all protons in the nucleus. For heavy nuclei ($Z > 15$), excess neutrons are necessary to keep the electrostatic interactions from breaking the nucleus apart, as shown in [\[link\]](#).



(a) The electrostatic force is repulsive and has long range. The arrows represent outward forces on protons (in blue) at the nuclear surface by a proton (also in blue) at the center. (b) The strong nuclear force acts between neighboring nucleons. The arrows represent attractive forces exerted by a neutron (in red) on its nearest neighbors.

Because of the existence of stable isotopes, we must take special care when quoting the mass of an element. For example, Copper (Cu) has two stable isotopes:

Equation:

$${}^{63}_{29}\text{Cu} \text{ (62.929595 g/mol) with an abundance of 69.09\%}$$

Equation:

$^{65}_{29}\text{Cu}$ (64.927786 g/mol) with an abundance of 30.91%

Given these two “versions” of Cu, what is the mass of this element? The **atomic mass** of an element is defined as the weighted average of the masses of its isotopes. Thus, the atomic mass of Cu is

$m_{\text{Cu}} = (62.929595) (0.6909) + (64.927786) (0.3091) = 63.55 \text{ g/mol}$. The mass of an individual nucleus is often expressed in **atomic mass units** (u), where $u = 1.66054 \times 10^{-27} \text{ kg}$. (An atomic mass unit is defined as 1/12th the mass of a ^{12}C nucleus.) In atomic mass units, the mass of a helium nucleus ($A = 4$) is approximately 4 u. A helium nucleus is also called an alpha (α) particle.

Nuclear Size

The simplest model of the nucleus is a densely packed sphere of nucleons. The volume V of the nucleus is therefore proportional to the number of nucleons A , expressed by

Equation:

$$V = \frac{4}{3} \pi r^3 = kA,$$

where r is the **radius of a nucleus** and k is a constant with units of volume. Solving for r , we have

Note:

Equation:

$$r = r_0 A^{1/3}$$

where r_0 is a constant. For hydrogen ($A = 1$), r_0 corresponds to the radius of a single proton. Scattering experiments support this general relationship for a wide range of nuclei, and they imply that neutrons have approximately the same radius as protons. The experimentally measured value for r_0 is approximately 1.2 femtometer (recall that $1 \text{ fm} = 10^{-15} \text{ m}$).

Example:**The Iron Nucleus**

Find the radius (r) and approximate density (ρ) of a Fe-56 nucleus. Assume the mass of the Fe-56 nucleus is approximately 56 u.

Strategy

(a) Finding the radius of ^{56}Fe is a straightforward application of $r = r_0 A^{1/3}$, given $A = 56$. (b) To find the approximate density of this nucleus, assume the nucleus is spherical. Calculate its volume using the radius found in part (a), and then find its density from $\rho = m/V$.

Solution

- a. The radius of a nucleus is given by

Equation:

$$r = r_0 A^{1/3}.$$

Substituting the values for r_0 and A yields

Equation:

$$\begin{aligned} r &= (1.2 \text{ fm})(56)^{1/3} = (1.2 \text{ fm})(3.83) \\ &= 4.6 \text{ fm}. \end{aligned}$$

- b. Density is defined to be $\rho = m/V$, which for a sphere of radius r is

Equation:

$$\rho = \frac{m}{V} = \frac{m}{(4/3)\pi r^3}.$$

Substituting known values gives

Equation:

$$\rho = \frac{56 \text{ u}}{(1.33)(3.14)(4.6 \text{ fm})^3} = 0.138 \text{ u/fm}^3.$$

Converting to units of kg/m^3 , we find

Equation:

$$\rho = (0.138 \text{ u/fm}^3)(1.66 \times 10^{-27} \text{ kg/u}) \left(\frac{1 \text{ fm}}{10^{-15} \text{ m}} \right) = 2.3 \times 10^{17} \text{ kg/m}^3.$$

Significance

- a. The radius of the Fe-56 nucleus is found to be approximately 5 fm, so its diameter is about 10 fm, or 10^{-14} m. In previous discussions of Rutherford's scattering experiments, a light nucleus was estimated to be 10^{-15} m in diameter. Therefore, the result shown for a mid-sized nucleus is reasonable.
- b. The density found here may seem incredible. However, it is consistent with earlier comments about the nucleus containing nearly all of the mass of the atom in a tiny region of space. One cubic meter of nuclear matter has the same mass as a cube of water 61 km on each side.

Note:**Exercise:****Problem:**

Check Your Understanding Nucleus X is two times larger than nucleus Y. What is the ratio of their atomic masses?

Solution:

eight

Summary

- The atomic nucleus is composed of protons and neutrons.
- The number of protons in the nucleus is given by the atomic number, Z . The number of neutrons in the nucleus is the neutron number, N . The number of nucleons is mass number, A .
- Atomic nuclei with the same atomic number, Z , but different neutron numbers, N , are isotopes of the same element.
- The atomic mass of an element is the weighted average of the masses of its isotopes.

Conceptual Questions

Exercise:

Problem:

Define and make clear distinctions between the terms neutron, nucleon, nucleus, and nuclide.

Solution:

The nucleus of an atom is made of one or more nucleons. A nucleon refers to either a proton or neutron. A nuclide is a stable nucleus.

Exercise:**Problem:**

What are isotopes? Why do isotopes of the same atom share the same chemical properties?

Problems**Exercise:****Problem:**

Find the atomic numbers, mass numbers, and neutron numbers for (a) ${}^{58}_{29}\text{Cu}$, (b) ${}^{24}_{11}\text{Na}$, (c) ${}^{210}_{84}\text{Po}$, (d) ${}^{45}_{20}\text{Ca}$, and (e) ${}^{206}_{82}\text{Pb}$.

Solution:

Use the rule $A = Z + N$.

	Atomic Number (Z)	Neutron Number (N)	Mass Number (A)
(a)	29	29	58
(b)	11	13	24
(c)	84	126	210

	Atomic Number (Z)	Neutron Number (N)	Mass Number (A)
(d)	20	25	45
(e)	82	124	206

Exercise:

Problem:

Silver has two stable isotopes. The nucleus, $^{107}_{47}\text{Ag}$, has atomic mass 106.905095 g/mol with an abundance of 51.83%; whereas $^{109}_{47}\text{Ag}$ has atomic mass 108.904754 g/mol with an abundance of 48.17%. Find the atomic mass of the element silver.

Exercise:

Problem:

The mass (M) and the radius (r) of a nucleus can be expressed in terms of the mass number, A . (a) Show that the density of a nucleus is independent of A . (b) Calculate the density of a gold (Au) nucleus. Compare your answer to that for iron (Fe).

Solution:

a. $r = r_0 A^{1/3}, \rho = \frac{3u}{4\pi r_0^3};$

b. $\rho = 2.3 \times 10^{17} \text{ kg/m}^3$

Exercise:

Problem:

A particle has a mass equal to 10 u. If this mass is converted completely into energy, how much energy is released? Express your answer in mega-electron volts (MeV). (Recall that $1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$.)

Exercise:

Problem:

Find the length of a side of a cube having a mass of 1.0 kg and the density of nuclear matter.

Solution:

side length = $1.6\ \mu\text{m}$

Exercise:**Problem:**

The detail that you can observe using a probe is limited by its wavelength. Calculate the energy of a particle that has a wavelength of $1 \times 10^{-16}\text{m}$, small enough to detect details about one-tenth the size of a nucleon.

Glossary

atomic mass

total mass of the protons, neutrons, and electrons in a single atom

atomic mass unit

unit used to express the mass of an individual nucleus, where
 $1\text{u} = 1.66054 \times 10^{-27}\text{ kg}$

atomic nucleus

tightly packed group of nucleons at the center of an atom

atomic number

number of protons in a nucleus

chart of the nuclides

graph comprising stable and unstable nuclei

isotopes

nuclei having the same number of protons but different numbers of neutrons

mass number

number of nucleons in a nucleus

neutron number

number of neutrons in a nucleus

nucleons

protons and neutrons found inside the nucleus of an atom

nuclide

nucleus

radius of a nucleus

radius of a nucleus is defined as $r = r_0 A^{1/3}$

strong nuclear force

force that binds nucleons together in the nucleus

Nuclear Binding Energy

By the end of this section, you will be able to:

- Calculate the mass defect and binding energy for a wide range of nuclei
- Use a graph of binding energy per nucleon (BEN) versus mass number (A) graph to assess the relative stability of a nucleus
- Compare the binding energy of a nucleon in a nucleus to the ionization energy of an electron in an atom

The forces that bind nucleons together in an atomic nucleus are much greater than those that bind an electron to an atom through electrostatic attraction. This is evident by the relative sizes of the atomic nucleus and the atom (10^{-15} and 10^{-10} m, respectively). The energy required to pry a nucleon from the nucleus is therefore much larger than that required to remove (or ionize) an electron in an atom. In general, all nuclear changes involve large amounts of energy per particle undergoing the reaction. This has numerous practical applications.

Mass Defect

According to nuclear particle experiments, the total mass of a nucleus (m_{nuc}) is *less* than the sum of the masses of its constituent nucleons (protons and neutrons). The mass difference, or **mass defect**, is given by

Note:

Equation:

$$\Delta m = Zm_p + (A - Z)m_n - m_{\text{nuc}}$$

where Zm_p is the total mass of the protons, $(A - Z)m_n$ is the total mass of the neutrons, and m_{nuc} is the mass of the nucleus. According to

Einstein's special theory of relativity, mass is a measure of the total energy of a system ($E = mc^2$). Thus, the total energy of a nucleus is less than the sum of the energies of its constituent nucleons. The formation of a nucleus from a system of isolated protons and neutrons is therefore an exothermic reaction—meaning that it releases energy. The energy emitted, or radiated, in this process is $(\Delta m)c^2$.

Now imagine this process occurs in reverse. Instead of forming a nucleus, energy is put into the system to break apart the nucleus ([link](#)). The amount of energy required is called the total **binding energy (BE)**, E_b .

Note:

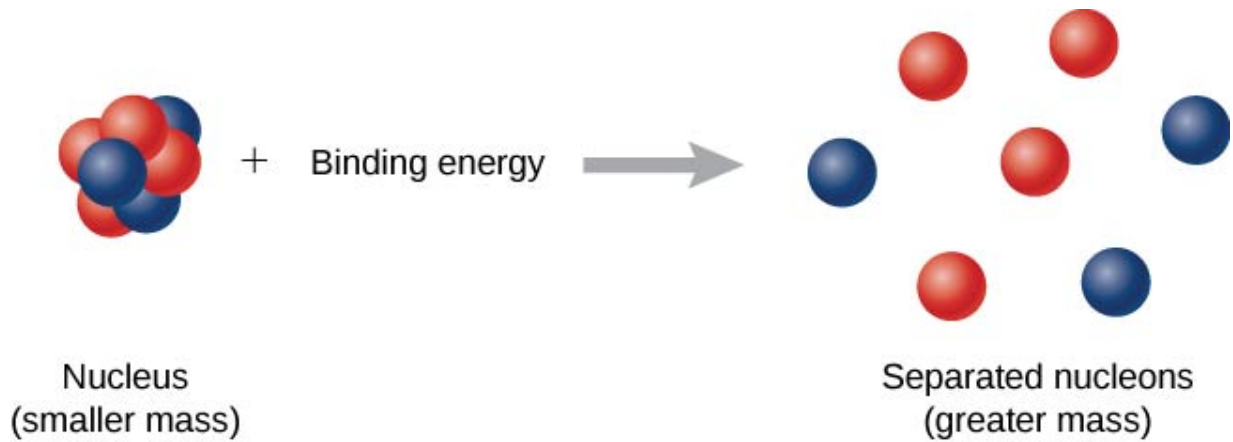
Binding Energy

The binding energy is equal to the amount of energy released in forming the nucleus, and is therefore given by

Equation:

$$E_b = (\Delta m)c^2.$$

Experimental results indicate that the binding energy for a nucleus with mass number $A > 8$ is roughly proportional to the total number of nucleons in the nucleus, A . The binding energy of a magnesium nucleus (^{24}Mg), for example, is approximately two times greater than for the carbon nucleus (^{12}C).



The binding energy is the energy required to break a nucleus into its constituent protons and neutrons. A system of separated nucleons has a greater mass than a system of bound nucleons.

Example:

Mass Defect and Binding Energy of the Deuteron

Calculate the mass defect and the binding energy of the deuteron. The mass of the deuteron is $m_D = 3.34359 \times 10^{-27} \text{kg}$ or $1875.61 \text{ MeV}/c^2$.

Solution

From [\[link\]](#), the mass defect for the deuteron is

Equation:

$$\begin{aligned}\Delta m &= m_p + m_n - m_D \\ &= 938.28 \text{ MeV}/c^2 + 939.57 \text{ MeV}/c^2 - 1875.61 \text{ MeV}/c^2 \\ &= 2.24 \text{ MeV}/c^2.\end{aligned}$$

The binding energy of the deuteron is then

Equation:

$$E_b = (\Delta m)c^2 = (2.24 \text{ MeV}/c^2) (c^2) = 2.24 \text{ MeV}.$$

Over two million electron volts are needed to break apart a deuteron into a proton and a neutron. This very large value indicates the great strength of the nuclear force. By comparison, the greatest amount of energy required to liberate an electron bound to a hydrogen atom by an attractive Coulomb force (an electromagnetic force) is about 10 eV.

Graph of Binding Energy per Nucleon

In nuclear physics, one of the most important experimental quantities is the **binding energy per nucleon (BEN)**, which is defined by

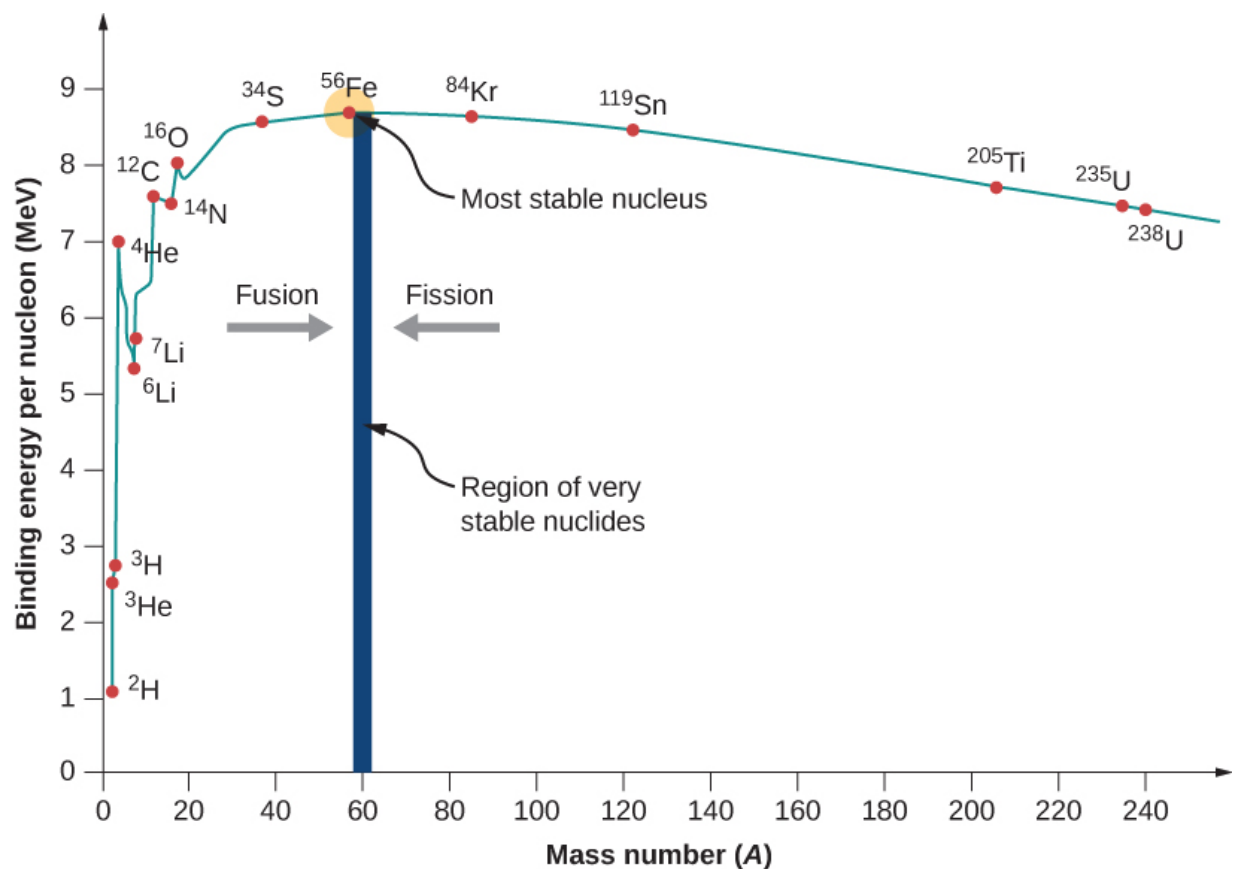
Note:
Equation:

$$BEN = \frac{E_b}{A}$$

This quantity is the average energy required to remove an individual nucleon from a nucleus—analogous to the ionization energy of an electron in an atom. If the BEN is relatively large, the nucleus is relatively stable. BEN values are estimated from nuclear scattering experiments.

A graph of binding energy per nucleon versus mass number A is given in [\[link\]](#). This graph is considered by many physicists to be one of the most important graphs in physics. Two notes are in order. First, typical BEN values range from 6–10 MeV, with an average value of about 8 MeV. In other words, it takes several million electron volts to pry a nucleon from a typical nucleus, as compared to just 13.6 eV to ionize an electron in the ground state of hydrogen. This is why nuclear force is referred to as the “strong” nuclear force.

Second, the graph rises at low A , peaks very near iron (Fe , $A = 56$), and then tapers off at high A . The peak value suggests that the iron nucleus is the most stable nucleus in nature (it is also why nuclear fusion in the cores of stars ends with Fe). The reason the graph rises and tapers off has to do with competing forces in the nucleus. At low values of A , attractive nuclear forces between nucleons dominate over repulsive electrostatic forces between protons. But at high values of A , repulsive electrostatic forces between forces begin to dominate, and these forces tend to break apart the nucleus rather than hold it together.



In this graph of binding energy per nucleon for stable nuclei, the BEN is greatest for nuclei with a mass near ^{56}Fe . Therefore, fusion of nuclei with mass numbers much less than that of Fe , and fission of nuclei with mass numbers greater than that of Fe , are exothermic processes.

As we will see, the BEN-versus- A graph implies that nuclei divided or combined release an enormous amount of energy. This is the basis for a wide range of phenomena, from the production of electricity at a nuclear power plant to sunlight.

Example:

Tightly Bound Alpha Nuclides

Calculate the binding energy per nucleon of an ${}^4\text{He}$ (α particle).

Strategy

Determine the total binding energy (BE) using the equation $\text{BE} = (\Delta m)c^2$, where Δm is the mass defect. The binding energy per nucleon (BEN) is BE divided by A .

Solution

For ${}^4\text{He}$, we have $Z = N = 2$. The total binding energy is

Equation:

$$\text{BE} = \{[2m_p + 2m_n] - m({}^4\text{He})\}c^2.$$

These masses are $m({}^4\text{He}) = 4.002602 \text{ u}$, $m_p = 1.007825 \text{ u}$, and $m_n = 1.008665 \text{ u}$. Thus we have,

Equation:

$$\text{BE} = (0.030378 \text{ u})c^2.$$

Noting that $1 \text{ u} = 931.5 \text{ MeV}/c^2$, we find

Equation:

$$\begin{aligned}\text{BE} &= (0.030378) (931.5 \text{ MeV}/c^2)c^2 \\ &= 28.3 \text{ MeV}.\end{aligned}$$

Since $A = 4$, the total binding energy per nucleon is

Equation:

$$\text{BEN} = 7.07 \text{ MeV/nucleon}.$$

Significance

Notice that the binding energy per nucleon for ${}^4\text{He}$ is much greater than for the hydrogen isotopes (only $\approx 3 \text{ MeV/nucleon}$). Therefore, helium nuclei cannot break down hydrogen isotopes without energy being put into the system.

Note:

Exercise:

Problem:

Check Your Understanding If the binding energy per nucleon is large, does this make it harder or easier to strip off a nucleon from a nucleus?

Solution:

harder

Summary

- The mass defect of a nucleus is the difference between the total mass of a nucleus and the sum of the masses of all its constituent nucleons.
- The binding energy (BE) of a nucleus is equal to the amount of energy released in forming the nucleus, or the mass defect multiplied by the speed of light squared.
- A graph of binding energy per nucleon (BEN) versus atomic number A implies that nuclei divided or combined release an enormous amount of energy.
- The binding energy of a nucleon in a nucleus is analogous to the ionization energy of an electron in an atom.

Conceptual Questions

Exercise:**Problem:**

Explain why a bound system should have less mass than its components. Why is this not observed traditionally, say, for a building made of bricks?

Solution:

A bound system should have less mass than its components because of energy-mass equivalence ($E = mc^2$). If the energy of a system is reduced, the total mass of the system is reduced. If two bricks are placed next to one another, the attraction between them is purely gravitational, assuming the bricks are electrically neutral. The gravitational force between the bricks is relatively small (compared to the strong nuclear force), so the mass defect is much too small to be observed. If the bricks are glued together with cement, the mass defect is likewise small because the electrical interactions between the electrons involved in the bonding are still relatively small.

Exercise:**Problem:**

Why is the number of neutrons greater than the number of protons in stable nuclei that have an A greater than about 40? Why is this effect more pronounced for the heaviest nuclei?

Exercise:**Problem:**

To obtain the most precise value of the binding energy per nucleon, it is important to take into account forces between nucleons at the surface of the nucleus. Will surface effects increase or decrease estimates of BEN?

Solution:

Nucleons at the surface of a nucleus interact with fewer nucleons. This reduces the binding energy per nucleon, which is based on an average over all the nucleons in the nucleus.

Problems

Exercise:

Problem:

How much energy would be released if six hydrogen atoms and six neutrons were combined to form $^{12}_6\text{C}$?

Solution:

92.4 MeV

Exercise:

Problem:

Find the mass defect and the binding energy for the helium-4 nucleus.

Exercise:

Problem:

^{56}Fe is among the most tightly bound of all nuclides. It makes up more than 90% of natural iron. Note that ^{56}Fe has even numbers of protons and neutrons. Calculate the binding energy per nucleon for ^{56}Fe and compare it with the approximate value obtained from the graph in [\[link\]](#).

Solution:

8.790 MeV \approx graph's value

Exercise:

Problem:

^{209}Bi is the heaviest stable nuclide, and its BEN is low compared with medium-mass nuclides. Calculate BEN for this nucleus and compare it with the approximate value obtained from the graph in [\[link\]](#).

Exercise:**Problem:**

(a) Calculate BEN for ^{235}U , the rarer of the two most common uranium isotopes; (b) Calculate BEN for ^{238}U . (Most of uranium is ^{238}U .)

Solution:

a. 7.570 MeV; b. 7.591 MeV \approx graph's value

Exercise:**Problem:**

The fact that BEN peaks at roughly $A = 60$ implies that the *range* of the strong nuclear force is about the diameter of this nucleus.

(a) Calculate the diameter of $A = 60$ nucleus.

(b) Compare BEN for ^{58}Ni and ^{90}Sr . The first is one of the most tightly bound nuclides, whereas the second is larger and less tightly bound.

Glossary

binding energy (BE)

energy needed to break a nucleus into its constituent protons and neutrons

binding energy per nucleon (BEN)

energy need to remove a nucleon from a nucleus

mass defect

difference between the mass of a nucleus and the total mass of its constituent nucleons

Radioactive Decay

By the end of this section, you will be able to:

- Describe the decay of a radioactive substance in terms of its decay constant and half-life
- Use the radioactive decay law to estimate the age of a substance
- Explain the natural processes that allow the dating of living tissue using ^{14}C

In 1896, Antoine Becquerel discovered that a uranium-rich rock emits invisible rays that can darken a photographic plate in an enclosed container. Scientists offer three arguments for the nuclear origin of these rays. First, the effects of the radiation do not vary with chemical state; that is, whether the emitting material is in the form of an element or compound. Second, the radiation does not vary with changes in temperature or pressure—both factors that in sufficient degree can affect electrons in an atom. Third, the very large energy of the invisible rays (up to hundreds of eV) is not consistent with atomic electron transitions (only a few eV). Today, this radiation is explained by the conversion of mass into energy deep within the nucleus of an atom. The spontaneous emission of radiation from nuclei is called nuclear **radioactivity** ([\[link\]](#)).



The international ionizing radiation symbol is universally

recognized as the warning symbol for nuclear radiation.

Radioactive Decay Law

When an individual nucleus transforms into another with the emission of radiation, the nucleus is said to **decay**. Radioactive decay occurs for all nuclei with $Z > 82$, and also for some unstable isotopes with $Z < 83$. The decay rate is proportional to the number of original (undecayed) nuclei N in a substance. The number of nuclei lost to decay, $-dN$ in time interval dt , is written

Note:

Equation:

$$-\frac{dN}{dt} = \lambda N$$

where λ is called the **decay constant**. (The minus sign indicates the number of original nuclei decreases over time.) In other words, the more nuclei available to decay, the more that do decay (in time dt). This equation can be rewritten as

Equation:

$$\frac{dN}{N} = -\lambda dt.$$

Integrating both sides of the equation, and defining N_0 to be the number of nuclei at $t = 0$, we obtain

Equation:

$$\int_{N_0}^N \frac{dN'}{N} = - \int_0^t \lambda dt'.$$

This gives us

Equation:

$$\ln \frac{N}{N_0} = -\lambda t.$$

Taking the left and right sides of the equation as a power of e , we have the **radioactive decay law**.

Note:

Radioactive Decay Law

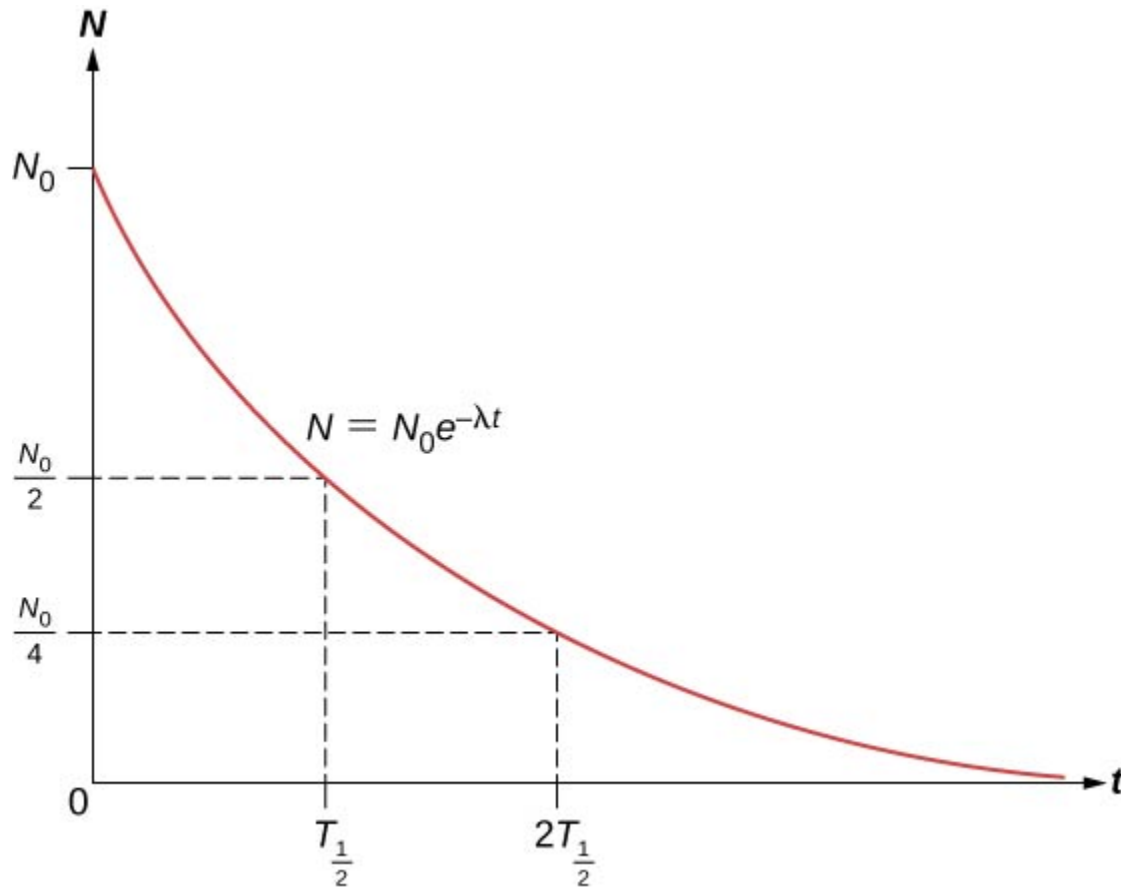
The total number N of radioactive nuclei remaining after time t is

Equation:

$$N = N_0 e^{-\lambda t}$$

where λ is the decay constant for the particular nucleus.

The total number of nuclei drops very rapidly at first, and then more slowly ([link](#)).



A plot of the radioactive decay law demonstrates that the number of nuclei remaining in a decay sample drops dramatically during the first moments of decay.

The **half-life** ($T_{1/2}$) of a radioactive substance is defined as the time for half of the original nuclei to decay (or the time at which half of the original nuclei remain). The half-lives of unstable isotopes are shown in the chart of nuclides in [\[link\]](#). The number of radioactive nuclei remaining after an integer (n) number of half-lives is therefore

Equation:

$$N = \frac{N_0}{2^n}$$

If the decay constant (λ) is large, the half-life is small, and vice versa. To determine the relationship between these quantities, note that when $t = T_{1/2}$, then $N = N_0/2$. Thus, [\[link\]](#) can be rewritten as

Equation:

$$\frac{N_0}{2} = N_0 e^{-\lambda T_{1/2}}.$$

Dividing both sides by N_0 and taking the natural logarithm yields

Equation:

$$\ln \frac{1}{2} = \ln e^{-\lambda T_{1/2}}$$

which reduces to

Note:

Equation:

$$\lambda = \frac{0.693}{T_{1/2}}.$$

Thus, if we know the half-life $T_{1/2}$ of a radioactive substance, we can find its decay constant. The **lifetime** \bar{T} of a radioactive substance is defined as the average amount of time that a nucleus exists before decaying. The lifetime of a substance is just the reciprocal of the decay constant, written as

Note:

Equation:

$$T = \frac{1}{\lambda}.$$

The **activity** A is defined as the magnitude of the decay rate, or
Equation:

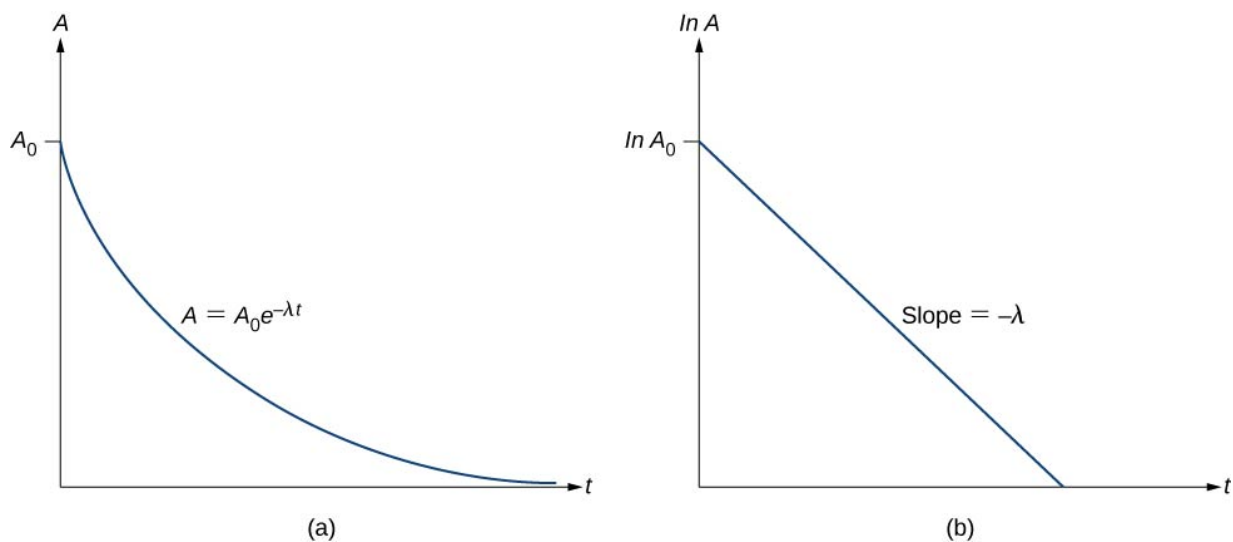
$$A = -\frac{dN}{dt} = \lambda N = \lambda N_0 e^{-\lambda t}.$$

The infinitesimal change dN in the time interval dt is negative because the number of parent (undecayed) particles is decreasing, so the activity (A) is positive. Defining the initial activity as $A_0 = \lambda N_0$, we have

Note:
Equation:

$$A = A_0 e^{-\lambda t}.$$

Thus, the activity A of a radioactive substance decreases exponentially with time ([link](#)).



(a) A plot of the activity as a function of time (b) If we measure the activity at different times, we can plot $\ln A$ versus t , and obtain a straight line.

Example:

Decay Constant and Activity of Strontium-90

The half-life of strontium-90, $^{90}_{38}\text{Sr}$, is 28.8 y. Find (a) its decay constant and (b) the initial activity of 1.00 g of the material.

Strategy

We can find the decay constant directly from [\[link\]](#). To determine the activity, we first need to find the number of nuclei present.

Solution

a. The decay constant is found to be

Equation:

$$\lambda = \frac{0.693}{T_{1/2}} = \left(\frac{0.693}{T_{1/2}} \right) \left(\frac{1 \text{ yr}}{3.16 \times 10^7 \text{ s}} \right) = 7.61 \times 10^{-10} \text{ s}^{-1}.$$

b. The atomic mass of $^{90}_{38}\text{Sr}$ is 89.91 g. Using Avogadro's number $N_A = 6.022 \times 10^{23}$ atoms/mol, we find the initial number of nuclei in 1.00 g of the material:

Equation:

$$N_0 = \frac{1.00 \text{ g}}{89.91 \text{ g}} (N_A) = 6.70 \times 10^{21} \text{ nuclei.}$$

From this, we find that the activity A_0 at $t = 0$ for 1.00 g of strontium-90 is

Equation:

$$\begin{aligned} A_0 &= \lambda N_0 \\ &= (7.61 \times 10^{-10} \text{ s}^{-1})(6.70 \times 10^{21} \text{ nuclei}) \\ &= 5.10 \times 10^{12} \text{ decays/s.} \end{aligned}$$

Expressing λ in terms of the half-life of the substance, we get

Equation:

$$A = A_0 e^{-(0.693/T_{1/2})T_{1/2}} = A_0 e^{-0.693} = A_0/2.$$

Therefore, the activity is halved after one half-life. We can determine the decay constant λ by measuring the activity as a function of time. Taking the natural logarithm of the left and right sides of [\[link\]](#), we get

Note:

Equation:

$$\ln A = -\lambda t + \ln A_0.$$

This equation follows the linear form $y = mx + b$. If we plot $\ln A$ versus t , we expect a straight line with slope $-\lambda$ and y-intercept $\ln A_0$ ([link](#))(b)).

Activity A is expressed in units of **becquerels** (Bq), where one

1 Bq = 1 decay per second. This quantity can also be expressed in decays per minute or decays per year. One of the most common units for activity is the **curie (Ci)**, defined to be the activity of 1 g of ^{226}Ra . The relationship between the Bq and Ci is

Equation:

$$1 \text{ Ci} = 3.70 \times 10^{10} \text{ Bq}.$$

Example:

What is ^{14}C Activity in Living Tissue?

Approximately 20% of the human body by mass is carbon. Calculate the activity due to ^{14}C in 1.00 kg of carbon found in a living organism.

Express the activity in units of Bq and Ci.

Strategy

The activity of ^{14}C is determined using the equation $A_0 = \lambda N_0$, where λ is the decay constant and N_0 is the number of radioactive nuclei. The number of ^{14}C nuclei in a 1.00-kg sample is determined in two steps. First, we determine the number of ^{12}C nuclei using the concept of a mole.

Second, we multiply this value by 1.3×10^{-12} (the known abundance of ^{14}C in a carbon sample from a living organism) to determine the number of ^{14}C nuclei in a living organism. The decay constant is determined from the known half-life of ^{14}C (available from [link](#)).

Solution

One mole of carbon has a mass of 12.0 g, since it is nearly pure ^{12}C . Thus, the number of carbon nuclei in a kilogram is

Equation:

$$N(^{12}\text{C}) = \frac{6.02 \times 10^{23} \text{ mol}^{-1}}{12.0 \text{ g/mol}} \times (1000 \text{ g}) = 5.02 \times 10^{25}.$$

The number of ^{14}C nuclei in 1 kg of carbon is therefore

Equation:

$$N(^{14}\text{C}) = (5.02 \times 10^{25}) (1.3 \times 10^{-12}) = 6.52 \times 10^{13}.$$

Now we can find the activity A by using the equation $A = \frac{0.693 N}{t_{1/2}}$.

Entering known values gives us

Equation:

$$A = \frac{0.693 (6.52 \times 10^{13})}{5730 \text{ y}} = 7.89 \times 10^9 \text{ y}^{-1}$$

or 7.89×10^9 decays per year. To convert this to the unit Bq, we simply convert years to seconds. Thus,

Equation:

$$A = (7.89 \times 10^9 \text{ y}^{-1}) \frac{1.00 \text{ y}}{3.16 \times 10^7 \text{ s}} = 250 \text{ Bq},$$

or 250 decays per second. To express A in curies, we use the definition of a curie,

Equation:

$$A = \frac{250 \text{ Bq}}{3.7 \times 10^{10} \text{ Bq/Ci}} = 6.76 \times 10^{-9} \text{ Ci}.$$

Thus,

Equation:

$$A = 6.76 \text{ nCi}.$$

Significance

Approximately 20% of the human body by weight is carbon. Hundreds of ^{14}C decays take place in the human body every second. Carbon-14 and other naturally occurring radioactive substances in the body compose a person's background exposure to nuclear radiation. As we will see later in this chapter, this activity level is well below the maximum recommended dosages.

Radioactive Dating

Radioactive dating is a technique that uses naturally occurring radioactivity to determine the age of a material, such as a rock or an ancient artifact. The basic approach is to estimate the original number of nuclei in a material and the present number of nuclei in the material (after decay), and then use the known value of the decay constant λ and [\[link\]](#) to calculate the total time of the decay, t .

An important method of radioactive dating is **carbon-14 dating**. Carbon-14 nuclei are produced when high-energy solar radiation strikes ^{14}N nuclei in the upper atmosphere and subsequently decay with a half-life of 5730 years. Radioactive carbon has the same chemistry as stable carbon, so it combines with the ecosphere and eventually becomes part of every living organism. Carbon-14 has an abundance of 1.3 parts per trillion of normal carbon. Therefore, if you know the number of carbon nuclei in an object, you multiply that number by 1.3×10^{-12} to find the number of ^{14}C nuclei in that object. When an organism dies, carbon exchange with the environment ceases, and ^{14}C is not replenished as it decays.

By comparing the abundance of ^{14}C in an artifact, such as mummy wrappings, with the normal abundance in living tissue, it is possible to determine the mummy's age (or the time since the person's death). Carbon-14 dating can be used for biological tissues as old as 50,000 years, but is generally most accurate for younger samples, since the abundance of ^{14}C nuclei in them is greater. Very old biological materials contain no ^{14}C at all. The validity of carbon dating can be checked by other means, such as by historical knowledge or by tree-ring counting.

Example:

An Ancient Burial Cave

In an ancient burial cave, your team of archaeologists discovers ancient wood furniture. Only 80% of the original ^{14}C remains in the wood. How old is the furniture?

Strategy

The problem statement implies that $N/N_0 = 0.80$. Therefore, the equation $N = N_0 e^{-\lambda t}$ can be used to find the product, λt . We know the half-life of ^{14}C is 5730 y, so we also know the decay constant, and therefore the total decay time t .

Solution

Solving the equation $N = N_0 e^{-\lambda t}$ for N/N_0 gives us

Equation:

$$\frac{N}{N_0} = e^{-\lambda t}.$$

Thus,

Equation:

$$0.80 = e^{-\lambda t}.$$

Taking the natural logarithm of both sides of the equation yields

Equation:

$$\ln 0.80 = -\lambda t,$$

so that

Equation:

$$-0.223 = -\lambda t.$$

Rearranging the equation to isolate t gives us

Equation:

$$t = \frac{0.223}{\lambda},$$

where

Equation:

$$\lambda = \frac{0.693}{t_{1/2}} = \frac{0.693}{5730 \text{ y}}.$$

Combining this information yields

Equation:

$$t = \frac{0.223}{\left(\frac{0.693}{5730 \text{ y}}\right)} = 1844 \text{ y.}$$

Significance

The furniture is almost 2000 years old—an impressive discovery. The typical uncertainty on carbon-14 dating is about 5%, so the furniture is anywhere between 1750 and 1950 years old. This date range must be confirmed by other evidence, such as historical records.

Note:**Exercise:****Problem:**

Check Your Understanding A radioactive nuclide has a high decay rate. What does this mean for its half-life and activity?

Solution:

Half-life is inversely related to decay rate, so the half-life is short. Activity depends on both the number of decaying particles and the decay rate, so the activity can be great or small.

Note:

Visit the [Radioactive Dating Game](#) to learn about the types of radiometric dating and try your hand at dating some ancient objects.

Summary

- In the decay of a radioactive substance, if the decay constant (λ) is large, the half-life is small, and vice versa.
- The radioactive decay law, $N = N_0 e^{-\lambda t}$, uses the properties of radioactive substances to estimate the age of a substance.
- Radioactive carbon has the same chemistry as stable carbon, so it mixes into the ecosphere and eventually becomes part of every living organism. By comparing the abundance of ^{14}C in an artifact with the normal abundance in living tissue, it is possible to determine the artifact's age.

Conceptual Questions

Exercise:

Problem:

How is the initial activity rate of a radioactive substance related to its half-life?

Exercise:

Problem:

For the carbon dating described in this chapter, what important assumption is made about the time variation in the intensity of cosmic rays?

Solution:

That it is constant.

Problems

Exercise:

Problem:

A sample of radioactive material is obtained from a very old rock. A plot $\ln A$ versus t yields a slope value of -10^{-9} s^{-1} (see [\[link\]](#)(b)). What is the half-life of this material?

Solution:

The decay constant is equal to the negative value of the slope or 10^{-9} s^{-1} . The half-life of the nuclei, and thus the material, is $T_{1/2} = 693$ million years.

Exercise:

Problem: Show that: $T = \frac{1}{\lambda}$.

Exercise:**Problem:**

The half-life of strontium-91, ${}^{91}_{38}\text{Sr}$ is 9.70 h. Find (a) its decay constant and (b) for an initial 1.00-g sample, the activity after 15 hours.

Solution:

a. The decay constant is $\lambda = 1.99 \times 10^{-5} \text{ s}^{-1}$. b. Since strontium-91 has an atomic mass of 90.90 g, the number of nuclei in a 1.00-g sample is initially

$$N_0 = 6.63 \times 10^{21} \text{ nuclei.}$$

The initial activity for strontium-91 is

$$A_0 = \lambda N_0$$

$$= 1.32 \times 10^{17} \text{ decays/s}$$

The activity at $t = 15.0 \text{ h} = 5.40 \times 10^4 \text{ s}$ is

$$A = 4.51 \times 10^{16} \text{ decays/s.}$$

Exercise:

Problem:

A sample of pure carbon-14 ($T_{1/2} = 5730 \text{ y}$) has an activity of $1.0 \mu \text{ Ci}$. What is the mass of the sample?

Exercise:**Problem:**

A radioactive sample initially contains $2.40 \times 10^{-2} \text{ mol}$ of a radioactive material whose half-life is 6.00 h. How many moles of the radioactive material remain after 6.00 h? After 12.0 h? After 36.0 h?

Solution:

$1.20 \times 10^{-2} \text{ mol}$; $6.00 \times 10^{-3} \text{ mol}$; $3.75 \times 10^{-4} \text{ mol}$

Exercise:**Problem:**

An old campfire is uncovered during an archaeological dig. Its charcoal is found to contain less than 1/1000 the normal amount of ^{14}C . Estimate the minimum age of the charcoal, noting that $2^{10} = 1024$.

Exercise:**Problem:**

Calculate the activity R , in curies of 1.00 g of ^{226}Ra . (b) Explain why your answer is not exactly 1.00 Ci, given that the curie was originally supposed to be exactly the activity of a gram of radium.

Solution:

a. 0.988 Ci; b. The half-life of ^{226}Ra is more precisely known than it was when the Ci unit was established.

Exercise:

Problem:

Natural uranium consists of ^{235}U (percent abundance = 0.7200%, $\lambda = 3.12 \times 10^{-17}/\text{s}$) and ^{238}U (percent abundance = 99.27%, $\lambda = 4.92 \times 10^{-18}/\text{s}$). What were the values for percent abundance of ^{235}U and ^{238}U when Earth formed 4.5×10^9 years ago?

Exercise:**Problem:**

World War II aircraft had instruments with glowing radium-painted dials. The activity of one such instrument was 1.0×10^5 Bq when new. (a) What mass of ^{226}Ra was present? (b) After some years, the phosphors on the dials deteriorated chemically, but the radium did not escape. What is the activity of this instrument 57.0 years after it was made?

Solution:

a. $2.73\mu\text{g}$; b. 9.76×10^4 Bq

Exercise:**Problem:**

The ^{210}Po source used in a physics laboratory is labeled as having an activity of $1.0\mu\text{Ci}$ on the date it was prepared. A student measures the radioactivity of this source with a Geiger counter and observes 1500 counts per minute. She notices that the source was prepared 120 days before her lab. What fraction of the decays is she observing with her apparatus?

Exercise:

Problem:

Armor-piercing shells with depleted uranium cores are fired by aircraft at tanks. (The high density of the uranium makes them effective.) The uranium is called depleted because it has had its ^{235}U removed for reactor use and is nearly pure ^{238}U . Depleted uranium has been erroneously called nonradioactive. To demonstrate that this is wrong: (a) Calculate the activity of 60.0 g of pure ^{238}U . (b) Calculate the activity of 60.0 g of natural uranium, neglecting the ^{234}U and all daughter nuclides.

Solution:

a. $7.46 \times 10^5 \text{ Bq}$; b. $7.75 \times 10^5 \text{ Bq}$

Glossary

activity

magnitude of the decay rate for radioactive nuclides

becquerel (Bq)

SI unit for the decay rate of a radioactive material, equal to 1 decay/second

carbon-14 dating

method to determine the age of formerly living tissue using the ratio $^{14}\text{C}/^{12}\text{C}$

curie (Ci)

unit of decay rate, or the activity of 1 g of ^{226}Ra , equal to $3.70 \times 10^{10} \text{ Bq}$

decay

process by which an individual atomic nucleus of an unstable atom loses mass and energy by emitting ionizing particles

decay constant

quantity that is inversely proportional to the half-life and that is used in equation for number of nuclei as a function of time

half-life

time for half of the original nuclei to decay (or half of the original nuclei remain)

lifetime

average time that a nucleus exists before decaying

radioactive dating

application of radioactive decay in which the age of a material is determined by the amount of radioactivity of a particular type that occurs

radioactive decay law

describes the exponential decrease of parent nuclei in a radioactive sample

radioactivity

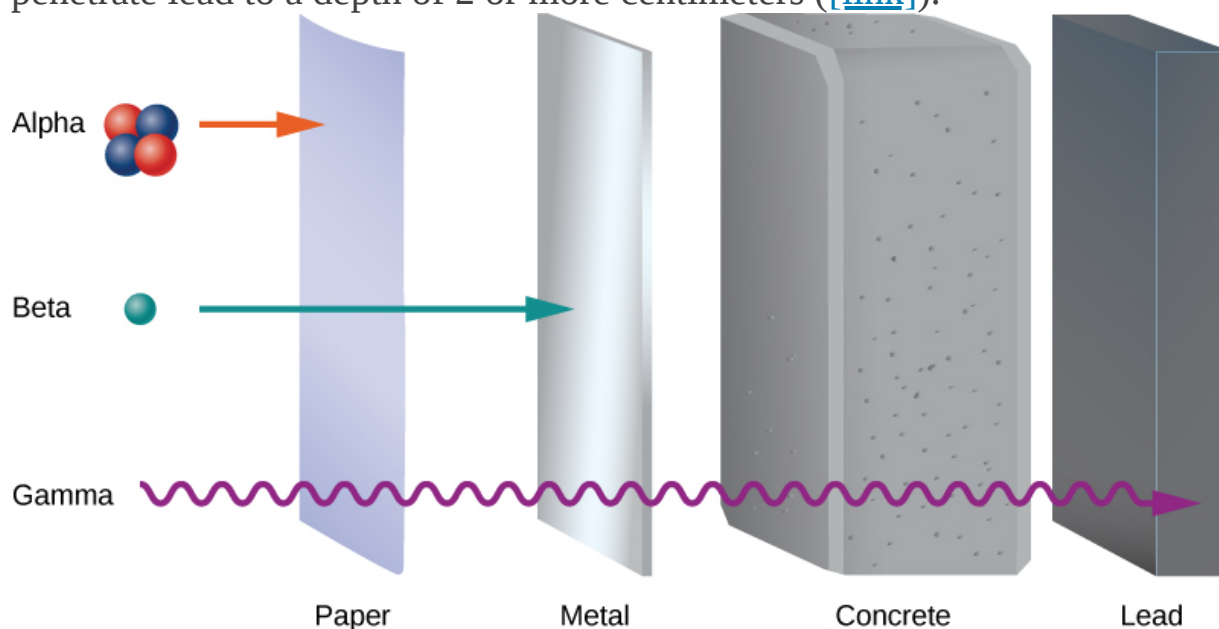
spontaneous emission of radiation from nuclei

Nuclear Reactions

By the end of this section, you will be able to:

- Describe and compare three types of nuclear radiation
- Use nuclear symbols to describe changes that occur during nuclear reactions
- Describe processes involved in the decay series of heavy elements

Early experiments revealed three types of nuclear “rays” or radiation: **alpha** (α) **rays**, **beta** (β) **rays**, and **gamma** (γ) **rays**. These three types of radiation are differentiated by their ability to penetrate matter. Alpha radiation is barely able to pass through a thin sheet of paper. Beta radiation can penetrate aluminum to a depth of about 3 mm, and gamma radiation can penetrate lead to a depth of 2 or more centimeters ([\[link\]](#)).

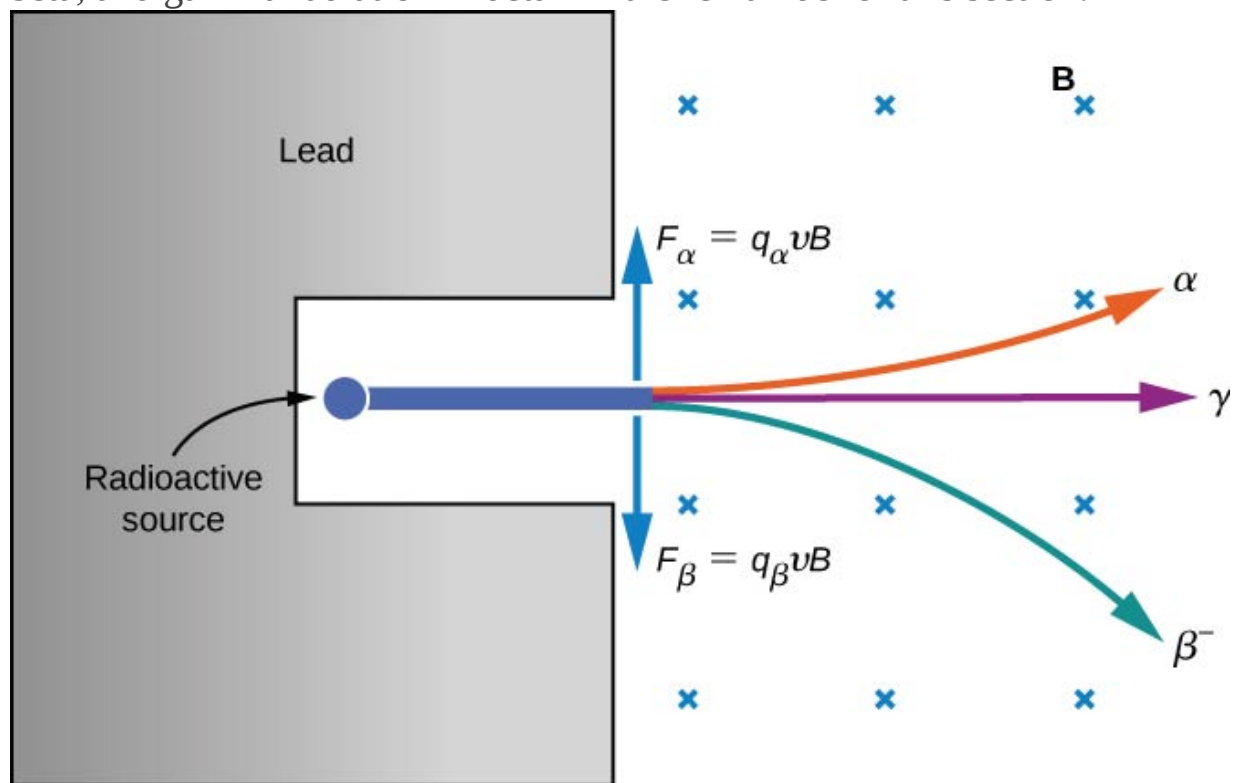


A comparison of the penetration depths of alpha (α), beta (β), and gamma (γ) radiation through various materials.

The electrical properties of these three types of radiation are investigated by passing them through a uniform magnetic field, as shown in [\[link\]](#).

According to the magnetic force equation $\vec{F} = q\vec{v} \times \vec{B}$, positively charged

particles are deflected upward, negatively charged particles are deflected downward, and particles with no charge pass through the magnetic field undeflected. Eventually, α rays were identified with helium nuclei (${}^4\text{He}$), β rays with electrons and **positrons** (positively charged electrons or **antielectrons**), and γ rays with high-energy photons. We discuss alpha, beta, and gamma radiation in detail in the remainder of this section.



The effect of a magnetic field on alpha (α), beta (β), and gamma (γ) radiation. This figure is a schematic only. The relative paths of the particles depend on their masses and initial kinetic energies.

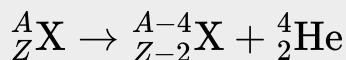
Alpha Decay

Heavy unstable nuclei emit α radiation. In α -particle decay (or **alpha decay**), the nucleus loses two protons and two neutrons, so the atomic number decreases by two, whereas its mass number decreases by four. Before the decay, the nucleus is called the **parent nucleus**. The nucleus or

nuclei produced in the decay are referred to as the **daughter nucleus** or daughter nuclei. We represent an α decay symbolically by

Note:

Equation:



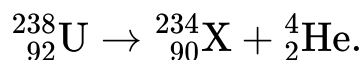
where ${}^A_Z\text{X}$ is the parent nucleus, ${}^{A-4}_{Z-2}\text{X}$ is the daughter nucleus, and ${}^4_2\text{He}$ is the α particle. In α decay, a nucleus of atomic number Z decays into a nucleus of atomic number $Z - 2$ and atomic mass $A - 4$. Interestingly, the dream of the ancient alchemists to turn other metals into gold is scientifically feasible through the alpha-decay process. The efforts of the alchemists failed because they relied on chemical interactions rather than nuclear interactions.

Note:

Watch alpha particles escape from a polonium nucleus, causing radioactive alpha decay. See how random decay times relate to the half-life. To try a simulation of alpha decay, visit [alpha particles](#)

An example of alpha decay is uranium-238:

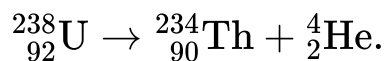
Equation:



The atomic number has dropped from 92 to 90. The chemical element with $Z = 90$ is thorium. Hence, Uranium-238 has decayed to Thorium-234 by

the emission of an α particle, written

Equation:



Subsequently, ${}_{90}^{234}\text{Th}$ decays by β emission with a half-life of 24 days. The energy released in this alpha decay takes the form of kinetic energies of the thorium and helium nuclei, although the kinetic energy of thorium is smaller than helium due to its heavier mass and smaller velocity.

Example:

Plutonium Alpha Decay

Find the energy emitted in the α decay of ${}^{239}\text{Pu}$.

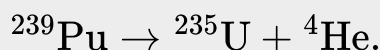
Strategy

The energy emitted in the α decay of ${}^{239}\text{Pu}$ can be found using the equation $E = (\Delta m)c^2$. We must first find Δm , the difference in mass between the parent nucleus and the products of the decay.

Solution

The decay equation is

Equation:



Thus, the pertinent masses are those of ${}^{239}\text{Pu}$, ${}^{235}\text{U}$, and the α particle or ${}^4\text{He}$, all of which are known. The initial mass was $m({}^{239}\text{Pu}) = 239.052157 \text{ u}$. The final mass is the sum

Equation:

$$\begin{aligned} m({}^{235}\text{U}) + m({}^4\text{He}) &= 235.043924 \text{ u} + 4.002602 \text{ u} \\ &= 239.046526 \text{ u}. \end{aligned}$$

Thus,

Equation:

$$\begin{aligned}
 \Delta m &= m(^{239}\text{Pu}) - [m(^{235}\text{U}) + m(^4\text{He})] \\
 &= 239.052157 \text{ u} - 239.046526 \text{ u} \\
 &= 0.0005631 \text{ u}.
 \end{aligned}$$

Now we can find E by entering Δm into the equation:

Equation:

$$E = (\Delta m)c^2 = (0.005631 \text{ u})c^2.$$

We know $1 \text{ u} = 931.5 \text{ MeV}/c^2$, so we have

Equation:

$$\begin{aligned}
 E &= (0.005631) (931.5 \text{ MeV}/c^2) (c^2) \\
 &= 5.25 \text{ MeV}.
 \end{aligned}$$

Significance

The energy released in this α decay is in the MeV range, many times greater than chemical reaction energies. Most of this energy becomes kinetic energy of the α particle (or ^4He nucleus), which moves away at high speed. The energy carried away by the recoil of the ^{235}U nucleus is much smaller due to its relatively large mass. The ^{235}U nucleus can be left in an excited state to later emit photons (γ rays).

Beta Decay

In most β particle decays (or **beta decay**), either an electron (β^-) or positron (β^+) is emitted by a nucleus. A positron has the same mass as the electron, but its charge is $+e$. For this reason, a positron is sometimes called an antielectron. How does β decay occur? A possible explanation is the electron (positron) is confined to the nucleus prior to the decay and somehow escapes. To obtain a rough estimate of the escape energy, consider a simplified model of an electron trapped in a box (or in the terminology of quantum mechanics, a one-dimensional square well) that has the width of a typical nucleus (10^{-14}m). According to the Heisenberg uncertainty

principle in [Quantum Mechanics](#), the uncertainty of the momentum of the electron is:

Equation:

$$\Delta p > \frac{h}{\Delta x} = \frac{6.6 \times 10^{-34} \text{ m}^2 \cdot \text{kg/s}}{10^{-14} \text{ m}} = 6.6 \times 10^{-20} \text{ kg} \cdot \text{m/s}.$$

Taking this momentum value (an underestimate) to be the “true value,” the kinetic energy of the electron on escape is approximately

Equation:

$$\frac{(\Delta p)^2}{2m_e} = \frac{(6.6 \times 10^{-20} \text{ kg} \cdot \text{m/s})^2}{2 (9.1 \times 10^{-31} \text{ kg})} = 2.0 \times 10^{-9} \text{ J} = 12,400 \text{ MeV}.$$

Experimentally, the electrons emitted in β^- decay are found to have kinetic energies of the order of only a few MeV. We therefore conclude that the electron is somehow produced in the decay rather than escaping the nucleus. Particle production (annihilation) is described by theories that combine quantum mechanics and relativity, a subject of a more advanced course in physics.

Nuclear beta decay involves the conversion of one nucleon into another. For example, a neutron can decay to a proton by the emission of an electron (β^-) and a nearly massless particle called an **antineutrino** ($\bar{\nu}$):

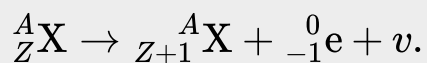
Equation:

$${}_0^1\text{n} \rightarrow {}_1^1\text{p} + {}_{-1}^0\text{e} + \bar{\nu}.$$

The notation ${}_{-1}^0\text{e}$ is used to designate the electron. Its mass number is 0 because it is not a nucleon, and its atomic number is -1 to signify that it has a charge of $-e$. The proton is represented by ${}_1^1\text{p}$ because its mass number and atomic number are 1. When this occurs within an atomic nucleus, we have the following equation for beta decay:

Note:

Equation:



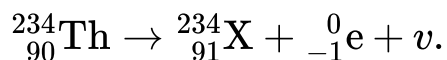
As discussed in another chapter, this process occurs due to the weak nuclear force.

Note:

Watch [beta decay](#) occur for a collection of nuclei or for an individual nucleus.

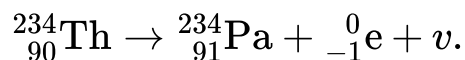
As an example, the isotope ${}_{90}^{234}\text{Th}$ is unstable and decays by β^- emission with a half-life of 24 days. Its decay can be represented as

Equation:



Since the chemical element with atomic number 91 is protactinium (Pa), we can write the β^- decay of thorium as

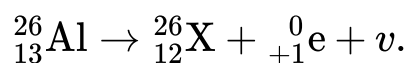
Equation:



The reverse process is also possible: A proton can decay to a neutron by the emission of a positron (β^+) and a nearly massless particle called a **neutrino** (ν). This reaction is written as ${}_1^1\text{p} \rightarrow {}_0^1\text{n} + {}_{+1}^0e + \nu$.

The positron ${}_{+1}^0\text{e}$ is emitted with the neutrino ν , and the neutron remains in the nucleus. (Like β^- decay, the positron does not precede the decay but is produced in the decay.) For an isolated proton, this process is impossible because the neutron is heavier than the proton. However, this process is possible within the nucleus because the proton can receive energy from other nucleons for the transition. As an example, the isotope of aluminum ${}_{13}^{26}\text{Al}$ decays by β^+ emission with a half-life of $7.40 \times 10^5\text{y}$. The decay is written as

Equation:



The atomic number 12 corresponds to magnesium. Hence,

Equation:



As a nuclear reaction, positron emission can be written as

Note:

Equation:



The neutrino was not detected in the early experiments on β decay. However, the laws of energy and momentum seemed to require such a particle. Later, neutrinos were detected through their interactions with nuclei.

Example:**Bismuth Alpha and Beta Decay**

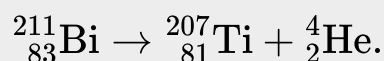
The $^{211}_{83}\text{Bi}$ nucleus undergoes both α and β^- decay. For each case, what is the daughter nucleus?

Strategy

We can use the processes described by [\[link\]](#) and [\[link\]](#), as well as the Periodic Table, to identify the resulting elements.

Solution

The atomic number and the mass number for the α particle are 2 and 4, respectively. Thus, when a bismuth-211 nucleus emits an α particle, the daughter nucleus has an atomic number of 81 and a mass number of 207. The element with an atomic number of 81 is thallium, so the decay is given by

Equation:

In β^- decay, the atomic number increases by 1, while the mass number stays the same. The element with an atomic number of 84 is polonium, so the decay is given by

Equation:**Note:****Exercise:****Problem:**

Check Your Understanding In radioactive beta decay, does the atomic mass number, A , increase or decrease?

Solution:

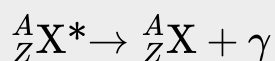
Neither; it stays the same.

Gamma Decay

A nucleus in an excited state can decay to a lower-level state by the emission of a “gamma-ray” photon, and this is known as **gamma decay**. This is analogous to de-excitation of an atomic electron. Gamma decay is represented symbolically by

Note:

Equation:



where the asterisk (*) on the nucleus indicates an excited state. In γ decay, neither the atomic number nor the mass number changes, so the type of nucleus does not change.

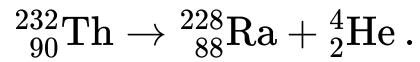
Radioactive Decay Series

Nuclei with $Z > 82$ are unstable and decay naturally. Many of these nuclei have very short lifetimes, so they are not found in nature. Notable exceptions include ${}^{232}_{90}\text{Th}$ (or Th-232) with a half-life of 1.39×10^{10} years, and ${}^{238}_{92}\text{U}$ (or U-238) with a half-life of 7.04×10^8 years. When a heavy nucleus decays to a lighter one, the lighter daughter nucleus can become the parent nucleus for the next decay, and so on. This process can produce a long series of nuclear decays called a **decay series**. The series ends with a stable nucleus.

To illustrate the concept of a decay series, consider the decay of Th-232 series ([\[link\]](#)). The neutron number, N , is plotted on the vertical y-axis, and the atomic number, Z , is plotted on the horizontal x-axis, so Th-232 is found

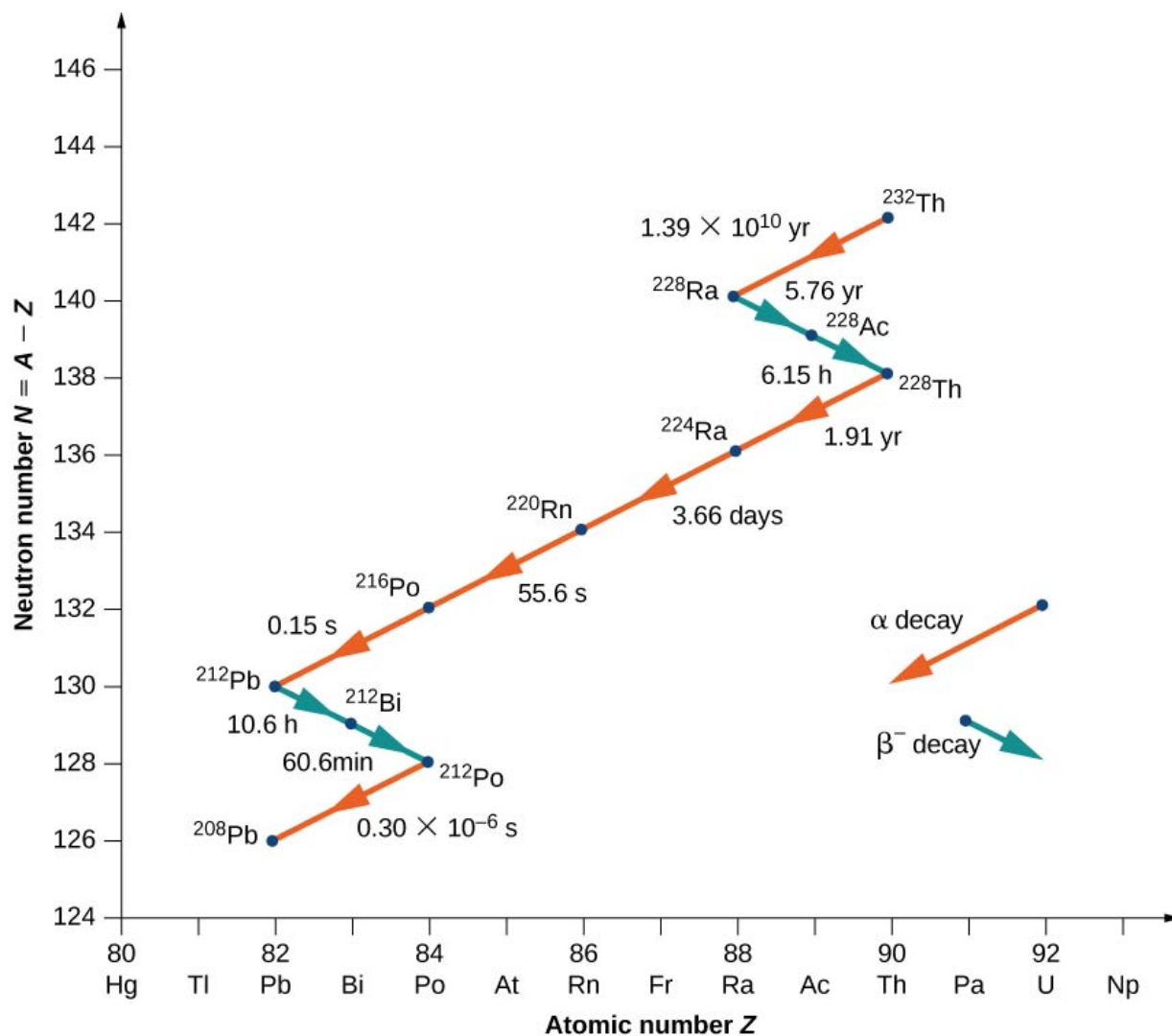
at the coordinates $(N, Z) = (142, 90)$. Th-232 decays by α emission with a half-life of 1.39×10^{10} years. Alpha decay decreases the atomic number by 2 and the mass number by 4, so we have

Equation:



The neutron number for Radium-228 is 140, so it is found in the diagram at the coordinates $(N, Z) = (140, 88)$. Radium-228 is also unstable and decays by β^- emission with a half-life of 5.76 years to Actinium-228. The atomic number increases by 1, the mass number remains the same, and the neutron number decreases by 1. Notice that in the graph, α emission appears as a line sloping downward to the left, with both N and Z decreasing by 2. Beta emission, on the other hand, appears as a line sloping downward to the right with N decreasing by 1, and Z increasing by 1. After several additional alpha and beta decays, the series ends with the stable nucleus Pb-208.

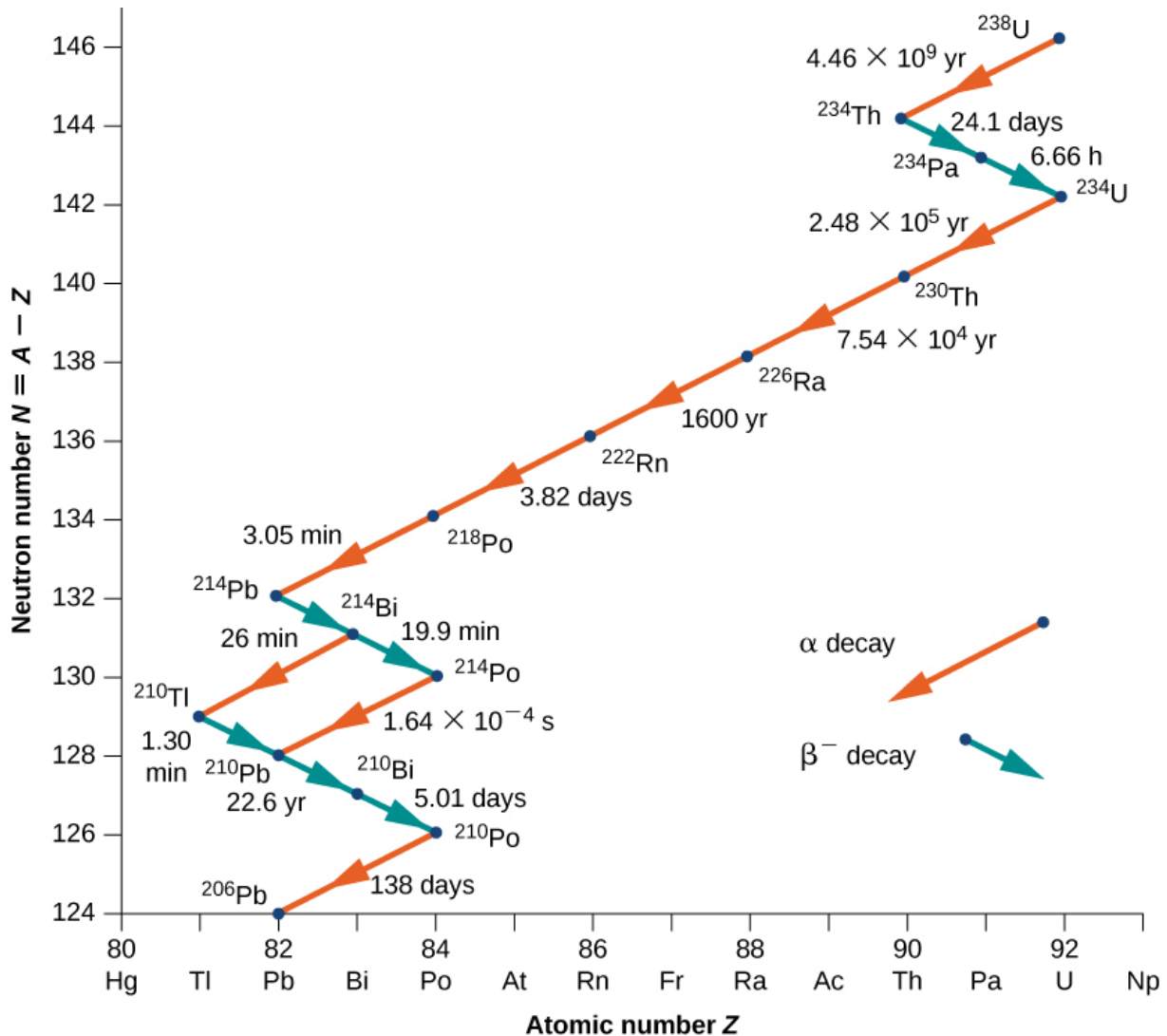
The relative frequency of different types of radioactive decays (alpha, beta, and gamma) depends on many factors, including the strength of the forces involved and the number of ways a given reaction can occur without violating the conservation of energy and momentum. How often a radioactive decay occurs often depends on a sensitive balance of the strong and electromagnetic forces. These forces are discussed in [Particle Physics and Cosmology](#).



In the thorium $^{232}_{90}\text{Th}$ decay series, alpha (α) decays reduce the atomic number, as indicated by the red arrows. Beta (β^-) decays increase the atomic number, as indicated by the blue arrows. The series ends at the stable nucleus Pb-208.

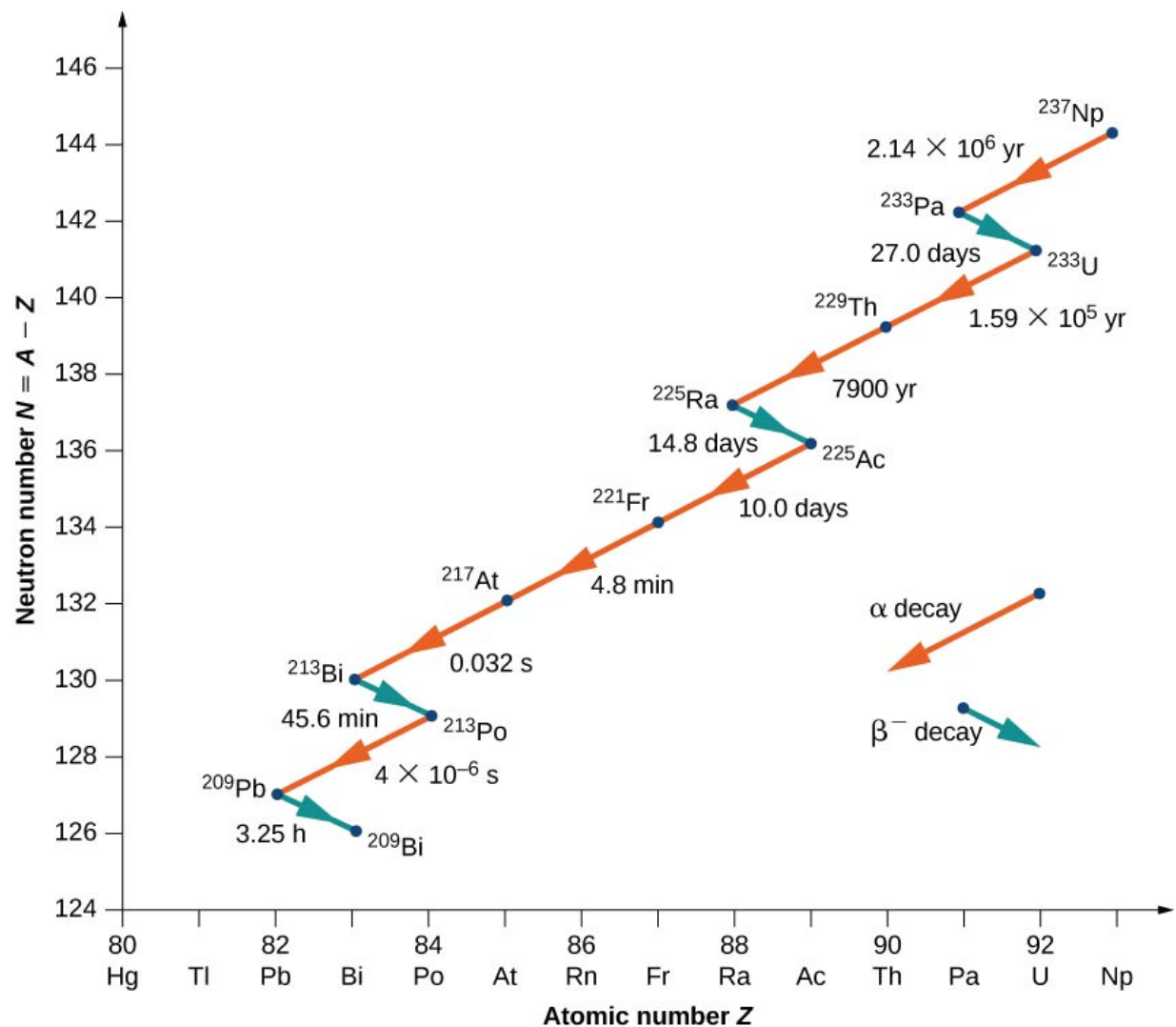
As another example, consider the U-238 decay series shown in [\[link\]](#). After numerous alpha and beta decays, the series ends with the stable nucleus Pb-206. An example of a decay whose parent nucleus no longer exists naturally is shown in [\[link\]](#). It starts with Neptunium-237 and ends in the stable nucleus Bismuth-209. Neptunium is called a **transuranic element** because it lies beyond uranium in the periodic table. Uranium has the highest atomic

number ($Z = 92$) of any element found in nature. Elements with $Z > 92$ can be produced only in the laboratory. They most probably also existed in nature at the time of the formation of Earth, but because of their relatively short lifetimes, they have completely decayed. There is nothing fundamentally different between naturally occurring and artificial elements.



In the Uranium-238 decay series, alpha (α) decays reduce the atomic number, as indicated by the red arrows. Beta (β^-) decays increase the atomic number, as indicated by the blue arrows. The series ends at the stable nucleus Pb-206.

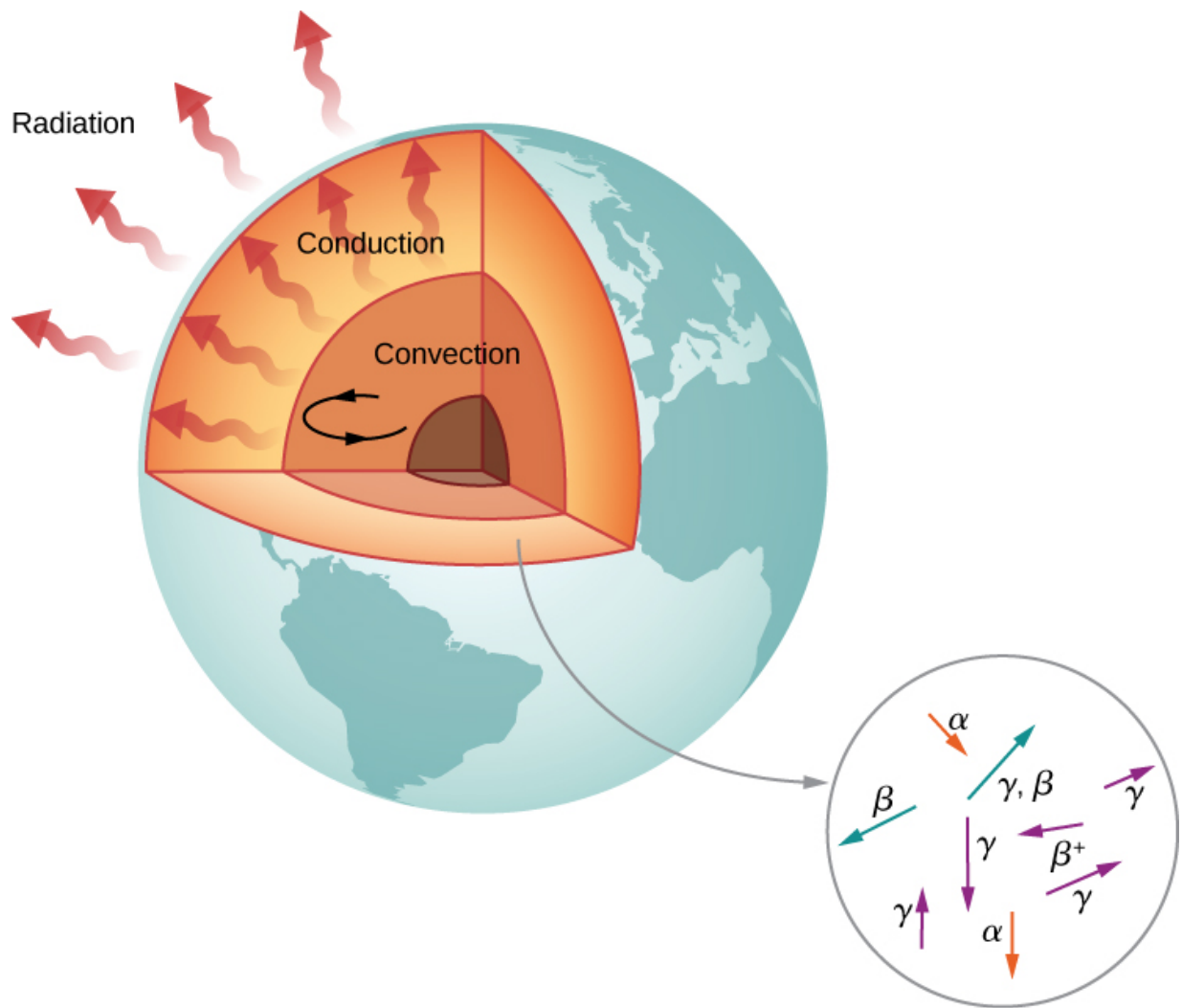
Notice that for Bi (21), the decay may proceed through either alpha or beta decay.



In the Neptunium-237 decay series, alpha (α) decays reduce the atomic number, as indicated by the red arrows. Beta (β^-) decays increase the atomic number, as indicated by the blue arrows. The series ends at the stable nucleus Bi-209.

Radioactivity in the Earth

According to geologists, if there were no heat source, Earth should have cooled to its present temperature in no more than 1 billion years. Yet, Earth is more than 4 billion years old. Why is Earth cooling so slowly? The answer is nuclear radioactivity, that is, high-energy particles produced in radioactive decays heat Earth from the inside ([link](#)).



Earth is heated by nuclear reactions (alpha, beta, and gamma decays).
Without these reactions, Earth's core and mantle would be much cooler than it is now.

Candidate nuclei for this heating model are ^{238}U and ^{40}K , which possess half-lives similar to or longer than the age of Earth. The energy produced by these decays (per second per cubic meter) is small, but the energy cannot escape easily, so Earth's core is very hot. Thermal energy in Earth's core is transferred to Earth's surface and away from it through the processes of convection, conduction, and radiation.

Summary

- The three types of nuclear radiation are alpha (α) rays, beta (β) rays, and gamma (γ) rays.
- We represent α decay symbolically by ${}^A_Z\text{X} \rightarrow {}^{A-4}_{Z-2}\text{X} + {}^4_2\text{He}$. There are two types of β decay: either an electron (β^-) or a positron (β^+) is emitted by a nucleus. γ decay is represented symbolically by ${}^A_Z\text{X}^* \rightarrow {}^A_Z\text{X} + \gamma$.
- When a heavy nucleus decays to a lighter one, the lighter daughter nucleus can become the parent nucleus for the next decay, and so on, producing a decay series.

Conceptual Questions

Exercise:

Problem:

What is the key difference and the key similarity between beta (β^-) decay and alpha decay?

Exercise:

Problem:

What is the difference between γ rays and characteristic X-rays and visible light?

Solution:

Gamma (γ) rays are produced by nuclear interactions and X-rays and light are produced by atomic interactions. Gamma rays are typically shorter wavelength than X-rays, and X-rays are shorter wavelength than light.

Exercise:

Problem:

What characteristics of radioactivity show it to be nuclear in origin and not atomic?

Exercise:

Problem:

Consider [\[link\]](#). If the magnetic field is replaced by an electric field pointed in toward the page, in which directions will the α -, β^+ -, and γ rays bend?

Solution:

Assume a rectangular coordinate system with an xy -plane that corresponds to the plane of the paper. α bends into the page (trajectory parabolic in the xz -plane); β^+ bends into the page (trajectory parabolic in the xz -plane); and γ is unbent.

Exercise:

Problem: Why is Earth's core molten?

Problems

Exercise:

Problem:

^{249}Cf undergoes alpha decay. (a) Write the reaction equation. (b) Find the energy released in the decay.

Exercise:**Problem:**

(a) Calculate the energy released in the α decay of ^{238}U . (b) What fraction of the mass of a single ^{238}U is destroyed in the decay? The mass of ^{234}Th is 234.043593 u. (c) Although the fractional mass loss is large for a single nucleus, it is difficult to observe for an entire macroscopic sample of uranium. Why is this?

Solution:

a. 4.273 MeV; b. 1.927×10^{-5} ; c. Since ^{238}U is a slowly decaying substance, only a very small number of nuclei decay on human timescales; therefore, although those nuclei that decay lose a noticeable fraction of their mass, the change in the total mass of the sample is not detectable for a macroscopic sample.

Exercise:**Problem:**

The β^- particles emitted in the decay of ^3H (tritium) interact with matter to create light in a glow-in-the-dark exit sign. At the time of manufacture, such a sign contains 15.0 Ci of ^3H . (a) What is the mass of the tritium? (b) What is its activity 5.00 y after manufacture?

Exercise:**Problem:**

(a) Write the complete β^- decay equation for ^{90}Sr , a major waste product of nuclear reactors. (b) Find the energy released in the decay.

Solution:

a. $^{90}_{38}\text{Sr}_{52} \rightarrow ^{90}_{39}\text{Y}_{51} + \beta^- + \bar{\nu}_e$; b. 0.546 MeV

Exercise:

Problem:

Write a nuclear β^- decay reaction that produces the ^{90}Y nucleus.
(*Hint:* The parent nuclide is a major waste product of reactors and has chemistry similar to calcium, so that it is concentrated in bones if ingested.)

Exercise:**Problem:**

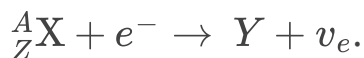
Write the complete decay equation in the complete ${}^A_Z\text{X}_N$ notation for the beta (β^-) decay of ${}^3\text{H}$ (tritium), a manufactured isotope of hydrogen used in some digital watch displays, and manufactured primarily for use in hydrogen bombs.

Solution:**Exercise:****Problem:**

If a 1.50-cm-thick piece of lead can absorb 90.0% of the rays from a radioactive source, how many centimeters of lead are needed to absorb all but 0.100% of the rays?

Exercise:**Problem:**

An electron can interact with a nucleus through the beta-decay process:



(a) Write the complete reaction equation for electron capture by ${}^7\text{Be}$.

(b) Calculate the energy released.

Solution:

a. ${}^7_4\text{Be}_3 + e^- \rightarrow {}^7_3\text{Li}_4 + \nu_e$; b. 0.862 MeV

Exercise:**Problem:**

(a) Write the complete reaction equation for electron capture by ${}^{15}\text{O}$.

(b) Calculate the energy released.

Exercise:**Problem:**

A rare decay mode has been observed in which ${}^{222}\text{Ra}$ emits a ${}^{14}\text{C}$ nucleus. (a) The decay equation is ${}^{222}\text{Ra} \rightarrow {}^A\text{X} + {}^{14}\text{C}$. Identify the nuclide ${}^A\text{X}$. (b) Find the energy emitted in the decay. The mass of ${}^{222}\text{Ra}$ is 222.015353 u.

Solution:

a. $\text{X} = {}^{208}_{82}\text{Pb}_{126}$; b. 33.05 MeV

Glossary

alpha decay

radioactive nuclear decay associated with the emission of an alpha particle

alpha (α) rays

one of the types of rays emitted from the nucleus of an atom as alpha particles

antielectrons

another term for positrons

antineutrino

antiparticle of an electron's neutrino in β^- decay

beta decay

radioactive nuclear decay associated with the emission of a beta particle

beta (β) rays

one of the types of rays emitted from the nucleus of an atom as beta particles

daughter nucleus

nucleus produced by the decay of a parent nucleus

decay series

series of nuclear decays ending in a stable nucleus

gamma decay

radioactive nuclear decay associated with the emission of gamma radiation

gamma (γ) rays

one of the types of rays emitted from the nucleus of an atom as gamma particles

neutrino

subatomic elementary particle which has no net electric charge

parent nucleus

original nucleus before decay

positron

electron with positive charge

transuranic element

element that lies beyond uranium in the periodic table

Fission

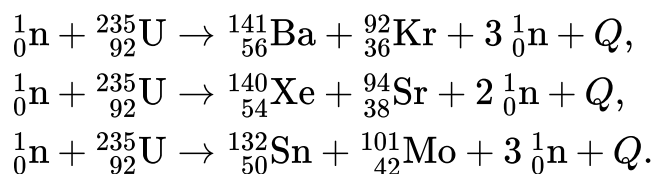
By the end of this section, you will be able to:

- Describe the process of nuclear fission in terms of its product and reactants
- Calculate the energies of particles produced by a fission reaction
- Explain the fission concept in the context of fission bombs and nuclear reactions

In 1934, Enrico Fermi bombarded chemical elements with neutrons in order to create isotopes of other elements. He assumed that bombarding uranium with neutrons would make it unstable and produce a new element. Unfortunately, Fermi could not determine the products of the reaction. Several years later, Otto Hahn and Fritz Strassman reproduced these experiments and discovered that the products of these reactions were smaller nuclei. From this, they concluded that the uranium nucleus had split into two smaller nuclei.

The splitting of a nucleus is called **fission**. Interestingly, U-235 fission does not always produce the same fragments. Example fission reactions include:

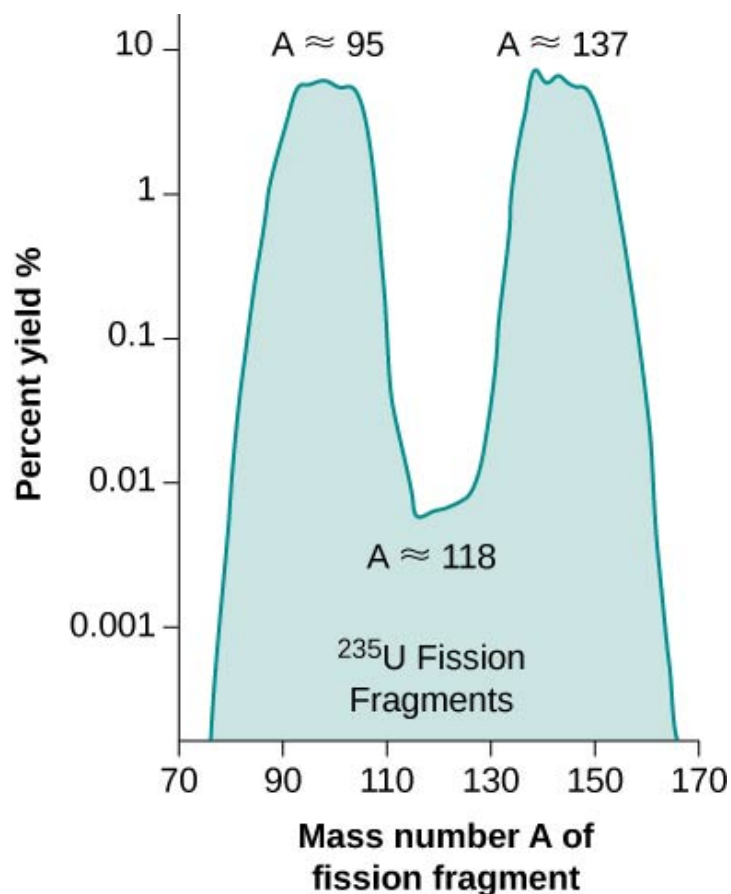
Equation:



In each case, the sum of the masses of the product nuclei are less than the masses of the reactants, so the fission of uranium is an exothermic process ($Q > 0$). This is the idea behind the use of fission reactors as sources of energy ([\[link\]](#)). The energy carried away by the reaction takes the form of particles with kinetic energy. The percent yield of fragments from a U-235 fission is given in [\[link\]](#).



The Phillipsburg Nuclear Power Plant in Germany uses a fission reactor to generate electricity.

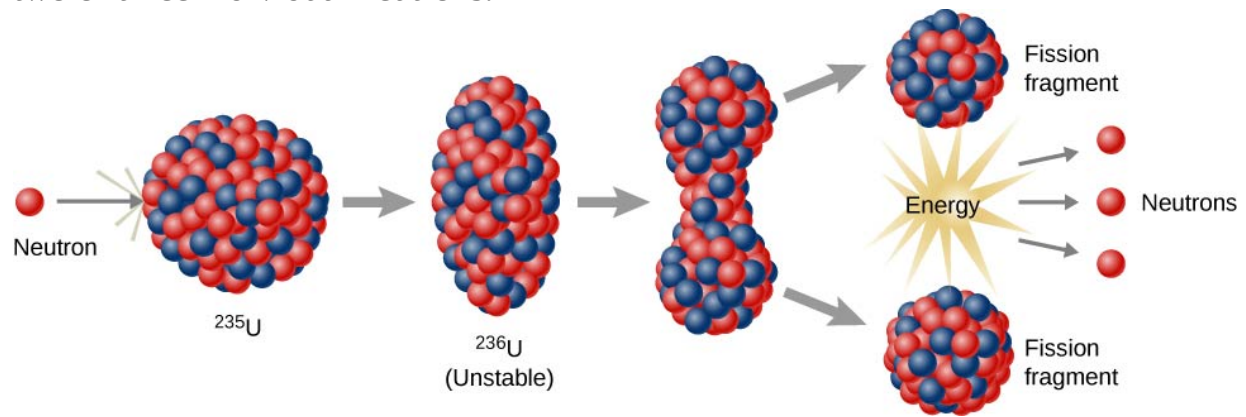


In this graph of fission fragments from U-235, the peaks in the graph indicate nuclei that are produced in the greatest abundance by the fission process.

Energy changes in a nuclear fission reaction can be understood in terms of the binding energy per nucleon curve ([\[link\]](#)). The BEN value for uranium ($A = 236$) is slightly lower than its daughter nuclei, which lie closer to the iron (Fe) peak. This means that nucleons in the nuclear fragments are more tightly bound than those in the U-235 nucleus. Therefore, a fission reaction results in a drop in the average energy of a nucleon. This energy is carried away by high-energy neutrons.

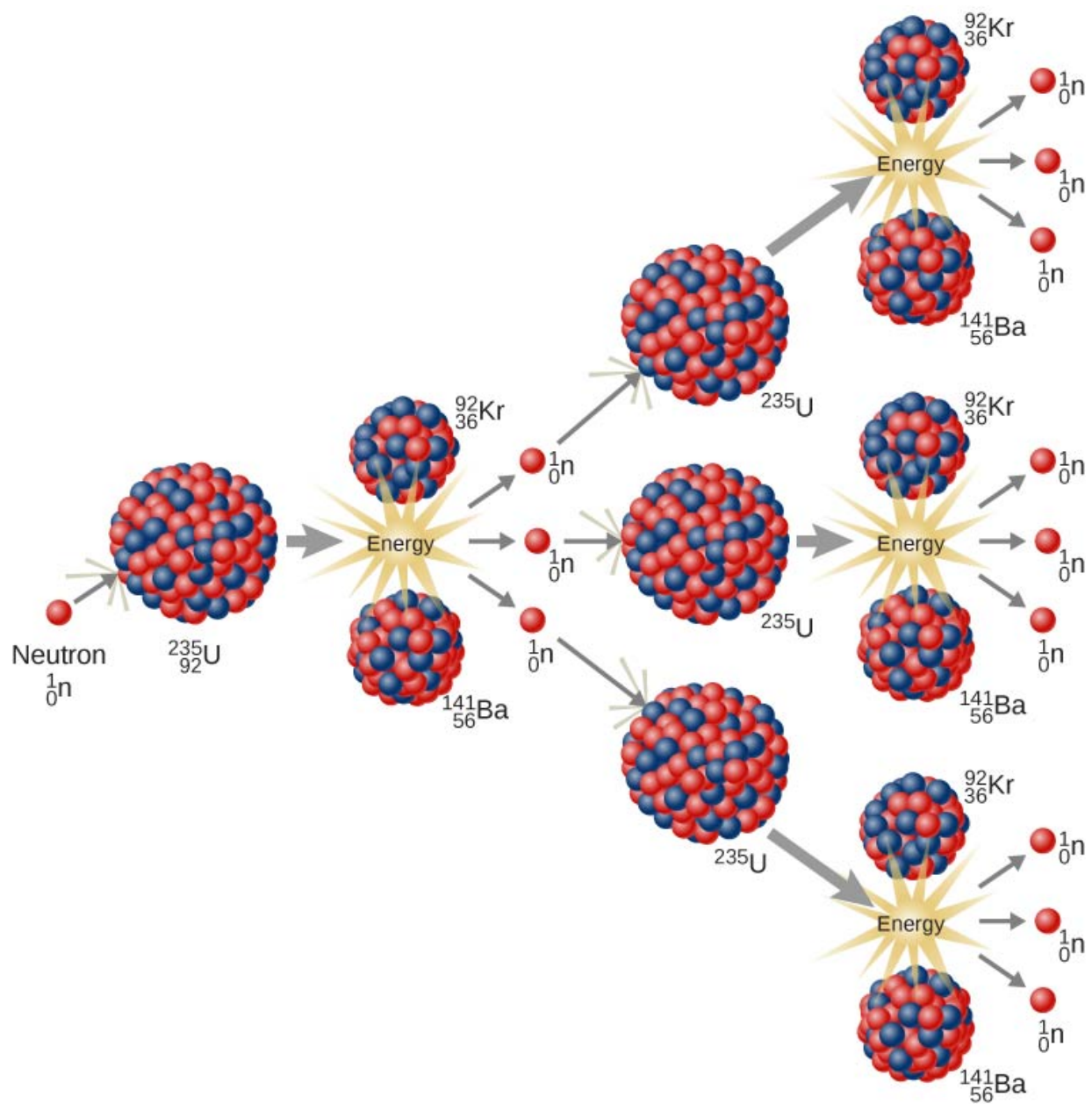
Niels Bohr and John Wheeler developed the **liquid drop model** to understand the fission process. According to this model, firing a neutron at a nucleus is analogous to disturbing a droplet of water ([\[link\]](#)). The analogy works because short-range forces between nucleons in a nucleus are similar to the attractive forces between

water molecules in a water droplet. In particular, forces between nucleons at the surface of the nucleus result in a surface tension similar to that of a water droplet. A neutron fired into a uranium nucleus can set the nucleus into vibration. If this vibration is violent enough, the nucleus divides into smaller nuclei and also emits two or three individual neutrons.



In the liquid drop model of nuclear fission, the uranium nucleus is split into two lighter nuclei by a high-energy neutron.

U-235 fission can produce a chain reaction. In a compound consisting of many U-235 nuclei, neutrons in the decay of one U-235 nucleus can initiate the fission of additional U-235 nuclei ([\[link\]](#)). This chain reaction can proceed in a controlled manner, as in a nuclear reactor at a power plant, or proceed uncontrollably, as in an explosion.



In a U-235 fission chain reaction, the fission of the uranium nucleus produces high-energy neutrons that go on to split more nuclei. The energy released in this process can be used to produce electricity.

Note:

View a simulation on [nuclear fission](#) to start a chain reaction, or introduce nonradioactive isotopes to prevent one. Control energy production in a nuclear reactor.

The Atomic Bomb

The possibility of a chain reaction in uranium, with its extremely large energy release, led nuclear scientists to conceive of making a bomb—an atomic bomb. (These discoveries were taking place in the years just prior to the Second World War and many of the European physicists involved in these discoveries came from countries that were being overrun.) Natural uranium contains 99.3% U-238 and only 0.7% U-235, and does not produce a chain reaction. To produce a controlled, sustainable chain reaction, the percentage of U-235 must be increased to about 50%. In addition, the uranium sample must be massive enough so a typical neutron is more likely to induce fission than it is to escape. The minimum mass needed for the chain reaction to occur is called the **critical mass**. When the critical mass reaches a point at which the chain reaction becomes self-sustaining, this is a condition known as **criticality**. The original design required two pieces of U-235 below the critical mass. When one piece in the form of a bullet is fired into the second piece, the critical mass is exceeded and a chain reaction is produced.

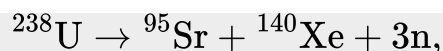
An important obstacle to the U-235 bomb is the production of a critical mass of fissionable material. Therefore, scientists developed a plutonium-239 bomb because Pu-239 is more fissionable than U-235 and thus requires a smaller critical mass. The bomb was made in the form of a sphere with pieces of plutonium, each below the critical mass, at the edge of the sphere. A series of chemical explosions fired the plutonium pieces toward the center of the sphere simultaneously. When all these pieces of plutonium came together, the combination exceeded the critical mass and produced a chain reaction. Both the U-235 and Pu-239 bombs were used in World War II. Whether to develop and use atomic weapons remain two of the most important questions faced by human civilization.

Example:

Calculating Energy Released by Fission

Calculate the energy released in the following rare spontaneous fission reaction:

Equation:



The atomic masses are $m(^{238}\text{U}) = 238.050784 \text{ u}$, $m(^{95}\text{Sr}) = 94.919388 \text{ u}$, $m(^{140}\text{Xe}) = 139.921610 \text{ u}$, and $m(\text{n}) = 1.008665 \text{ u}$.

Strategy

As always, the energy released is equal to the mass destroyed times c^2 , so we must find the difference in mass between the parent ^{238}U and the fission products.

Solution

The products have a total mass of

Equation:

$$\begin{aligned} m_{\text{products}} &= 94.919388 \text{ u} + 139.921610 \text{ u} + 3(1.008665 \text{ u}) \\ &= 237.866993 \text{ u}. \end{aligned}$$

The mass lost is the mass of $^{238}\text{U} - m_{\text{products}}$ or

Equation:

$$\Delta m = 238.050784 \text{ u} - 237.866993 \text{ u} = 0.183791 \text{ u}.$$

Therefore, the energy released is

Equation:

$$E = (\Delta m)c^2 = (0.183791 \text{ u}) \frac{931.5 \text{ MeV}/c^2}{\text{u}} c^2 = 171.2 \text{ MeV}.$$

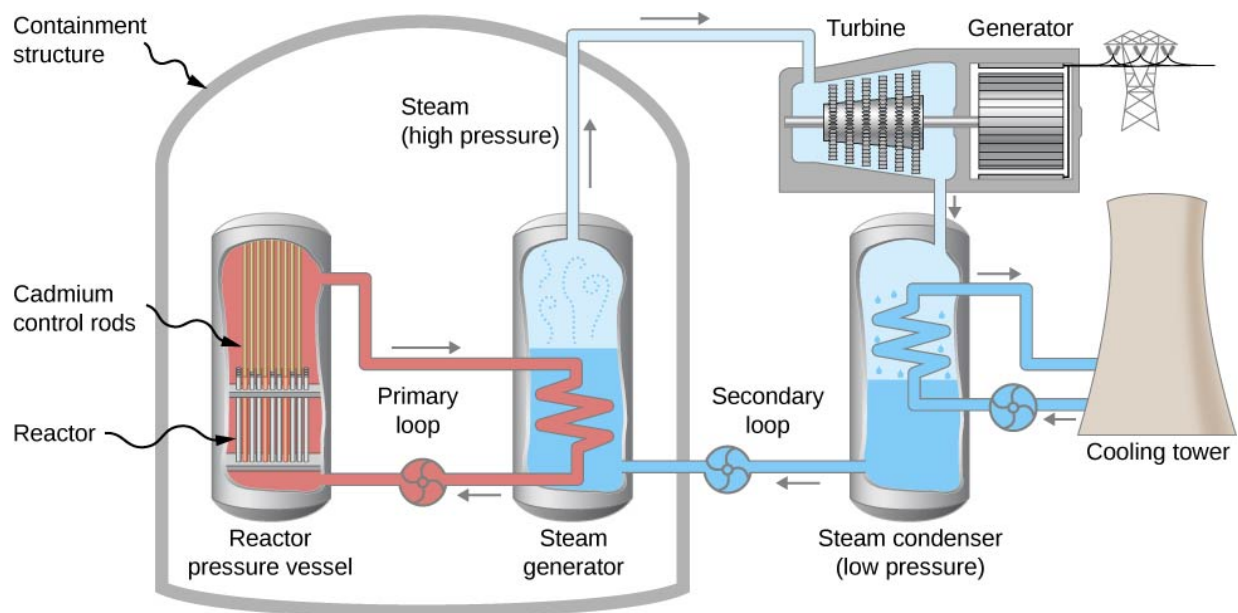
Significance

Several important things arise in this example. The energy release is large but less than it would be if the nucleus split into two equal parts, since energy is carried away by neutrons. However, this fission reaction produces neutrons and does not split the nucleus into two equal parts. Fission of a given nuclide, such as ^{238}U , does not always produce the same products. Fission is a statistical process in which an entire range of products are produced with various probabilities. Most fission produces neutrons, although the number varies. This is an extremely important aspect of fission, because *neutrons can induce more fission*, enabling self-sustaining chain reactions.

Fission Nuclear Reactors

The first nuclear reactor was built by Enrico Fermi on a squash court on the campus of the University of Chicago on December 2, 1942. The reactor itself contained U-238 enriched with 3.6% U-235. Neutrons produced by the chain reaction move too fast to initiate fission reactions. One way to slow them down is to enclose the entire reactor in a water bath under high pressure. The neutrons collide with the water molecules and are slowed enough to be used in the fission process. The slowed neutrons split more U-235 nuclei and a chain reaction occurs. The rate at which the chain reaction proceeds is controlled by a series of “control” rods made of cadmium inserted into the reactor. Cadmium is capable of absorbing a large number of neutrons without becoming unstable.

A nuclear reactor design, called a pressurized water reactor, can also be used to generate electricity ([link](#)). A pressurized water reactor (on the left in the figure) is designed to control the fission of large amounts of ^{235}U . The energy released in this process is absorbed by water flowing through pipes in the system (the “primary loop”) and steam is produced. Cadmium control rods adjust the neutron flux (the rate of flow of neutrons passing through the system) and therefore control the reaction. In case the reactor overheats and the water boils away, the chain reaction terminates, because water is used to thermalize the neutrons. (This safety feature can be overwhelmed in extreme circumstances.) The hot, high-pressure water then passes through a pipe to a second tank of water at normal pressure in the steam generator. The steam produced at one end of the steam generator fills a chamber that contains a turbine. This steam is at a very high pressure. Meanwhile, a steam condenser connected to the other side of the turbine chamber maintains steam at low pressure. The pressure differences force steam through the chamber, which turns the turbine. The turbine, in turn, powers an electric generator.



A nuclear reactor uses the energy produced in the fission of U-235 to produce electricity. Energy from a nuclear fission reaction produces hot, high-pressure steam that turns a turbine. As the turbine turns, electricity is produced.

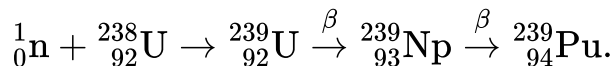
The major drawback to a fission reactor is nuclear waste. U-235 fission produces nuclei with long half-lives such as ^{236}U that must be stored. These products cannot be dumped into oceans or left in any place where they will contaminate the environment, such as through the soil, air, or water. Many scientists believe that the best place to store nuclear waste is the bottom of old salt mines or inside of stable mountains.

Many people are fearful that a nuclear reactor may explode like an atomic bomb. However, a nuclear reactor does not contain enough U-235 to do this. Also, a nuclear reactor is designed so that failure of any mechanism of the reactor causes the cadmium control rods to fall fully into the reactor, stopping the fission process. As evidenced by the Fukushima and Chernobyl disasters, such systems can fail. Systems and procedures to avoid such disasters is an important priority for advocates of nuclear energy.

If all electrical power were produced by nuclear fission of U-235, Earth's known reserves of uranium would be depleted in less than a century. However, Earth's supply of fissionable material can be expanded considerably using a **breeder reactor**. A breeder reactor operates for the first time using the fission of U-235 as

just described for the pressurized water reactor. But in addition to producing energy, some of the fast neutrons originating from the fission of U-235 are absorbed by U-238, resulting in the production of Pu-239 via the set of reactions

Equation:



The Pu-239 is itself highly fissionable and can therefore be used as a nuclear fuel in place of U-235. Since 99.3% of naturally occurring uranium is the U-238 isotope, the use of breeder reactors should increase our supply of nuclear fuel by roughly a factor of 100. Breeder reactors are now in operation in Great Britain, France, and Russia. Breeder reactors also have drawbacks. First, breeder reactors produce plutonium, which can, if leaked into the environment, produce serious public health problems. Second, plutonium can be used to build bombs, thus increasing significantly the risk of nuclear proliferation.

Example:

Calculating Energy of Fissionable Fuel

Calculate the amount of energy produced by the fission of 1.00 kg of ${}^{235}\text{U}$ given that the average fission reaction of ${}^{235}\text{U}$ produces 200 MeV.

Strategy

The total energy produced is the number of ${}^{235}\text{U}$ atoms times the given energy per ${}^{235}\text{U}$ fission. We should therefore find the number of ${}^{235}\text{U}$ atoms in 1.00 kg.

Solution

The number of ${}^{235}\text{U}$ atoms in 1.00 kg is Avogadro's number times the number of moles. One mole of ${}^{235}\text{U}$ has a mass of 235.04 g; thus, there are $(1000 \text{ g}) / (235.04 \text{ g/mol}) = 4.25 \text{ mol}$. The number of ${}^{235}\text{U}$ atoms is therefore

Equation:

$$(4.25 \text{ mol}) (6.02 \times 10^{23} {}^{235}\text{U}/\text{mol}) = 2.56 \times 10^{24} {}^{235}\text{U}.$$

Thus, the total energy released is

Equation:

$$E = (2.56 \times 10^{24} {}^{235}\text{U}) \left(\frac{200 \text{ MeV}}{{}^{235}\text{U}} \right) \left(\frac{1.60 \times 10^{-13} \text{ J}}{\text{MeV}} \right) = 8.21 \times 10^{13} \text{ J}.$$

Significance

This is another impressively large amount of energy, equivalent to about 14,000 barrels of crude oil or 600,000 gallons of gasoline. However, it is only one-fourth the energy produced by the fusion of a kilogram mixture of deuterium and tritium. Even though each fission reaction yields about 10 times the energy of a fusion reaction, the energy per kilogram of fission fuel is less, because there are far fewer moles per kilogram of the heavy nuclides. Fission fuel is also much scarcer than fusion fuel, and less than 1% of uranium (the ^{235}U) is readily usable.

Note:

Exercise:

Problem:

Check Your Understanding Which has a larger energy yield per fission reaction, a large or small sample of pure ^{235}U ?

Solution:

the same

Summary

- Nuclear fission is a process in which the sum of the masses of the product nuclei are less than the masses of the reactants.
- Energy changes in a nuclear fission reaction can be understood in terms of the binding energy per nucleon curve.
- The production of new or different isotopes by nuclear transformation is called breeding, and reactors designed for this purpose are called breeder reactors.

Conceptual Questions

Exercise:

Problem: Should an atomic bomb really be called *nuclear* bomb?

Solution:

Yes. An atomic bomb is a fission bomb, and a fission bomb occurs by splitting the *nucleus* of atom.

Exercise:

Problem: Why does a chain reaction occur during a fission reaction?

Exercise:

Problem: In what way is an atomic nucleus like a liquid drop?

Solution:

Short-range forces between nucleons in a nucleus are analogous to the forces between water molecules in a water droplet. In particular, the forces between nucleons at the surface of the nucleus produce a surface tension similar to that of a water droplet.

Problems**Exercise:****Problem:**

A large power reactor that has been in operation for some months is turned off, but residual activity in the core still produces 150 MW of power. If the average energy per decay of the fission products is 1.00 MeV, what is the core activity?

Exercise:**Problem:**

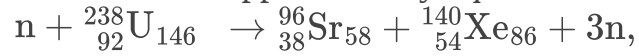
(a) Calculate the energy released in this rare neutron-induced fission $n + {}^{238}\text{U} \rightarrow {}^{96}\text{Sr} + {}^{140}\text{Xe} + 3n$, given $m({}^{96}\text{Sr}) = 95.921750 \text{ u}$ and $m({}^{140}\text{Xe}) = 139.92164$.

(b) This result is about 6 MeV greater than the result for spontaneous fission. Why?

(c) Confirm that the total number of nucleons and total charge are conserved in this reaction.

Solution:

a. 177.1 MeV; b. This value is approximately equal to the average BEN for



heavy nuclei. c.

$$A_i = 239 = A_f,$$

$$Z_i = 92 = 38 + 54 = Z_f$$

Exercise:

Problem:

(a) Calculate the energy released in the neutron-induced fission reaction $n + {}^{235}\text{U} \rightarrow {}^{92}\text{Kr} + {}^{142}\text{Ba} + 2n$, given $m({}^{92}\text{Kr}) = 91.926269 \text{ u}$ and $m({}^{142}\text{Ba}) = 141.916361 \text{ u}$. (b) Confirm that the total number of nucleons and total charge are conserved in this reaction.

Exercise:

Problem:

The electrical power output of a large nuclear reactor facility is 900 MW. It has a 35.0% efficiency in converting nuclear power to electrical power.

(a) What is the thermal nuclear power output in megawatts?

(b) How many ${}^{235}\text{U}$ nuclei fission each second, assuming the average fission produces 200 MeV?

(c) What mass of ${}^{235}\text{U}$ is fissioned in 1 year of full-power operation?

Solution:

a. $2.57 \times 10^3 \text{ MW}$; b. $8.04 \times 10^{19} \text{ fissions/s}$; c. 991 kg

Exercise:

Problem:

Find the total energy released if 1.00 kg of ${}^{235}_{92}\text{U}$ were to undergo fission.

Glossary

breeder reactor

reactor that is designed to make plutonium

criticality

condition in which a chain reaction easily becomes self-sustaining

critical mass

minimum mass required of a given nuclide in order for self-sustained fission to occur

fission

splitting of a nucleus

liquid drop model

model of nucleus (only to understand some of its features) in which nucleons in a nucleus act like atoms in a drop

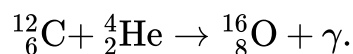
Nuclear Fusion

By the end of this section, you will be able to:

- Describe the process of nuclear fusion in terms of its product and reactants
- Calculate the energies of particles produced by a fusion reaction
- Explain the fission concept in the context of fusion bombs, the production of energy by the Sun, and nucleosynthesis

The process of combining lighter nuclei to make heavier nuclei is called **nuclear fusion**. As with fission reactions, fusion reactions are exothermic—they release energy. Suppose that we fuse a carbon and helium nuclei to produce oxygen:

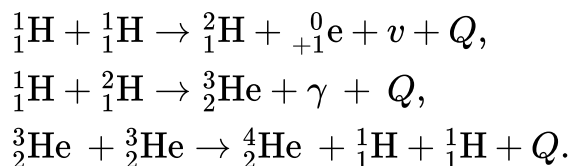
Equation:



The energy changes in this reaction can be understood using a graph of binding energy per nucleon ([\[link\]](#)). Comparing the binding energy per nucleon for oxygen, carbon, and helium, the oxygen nucleus is much more tightly bound than the carbon and helium nuclei, indicating that the reaction produces a drop in the energy of the system. This energy is released in the form of gamma radiation. Fusion reactions are said to be exothermic when the amount of energy released (known as the *Q value*) in each reaction is greater than zero ($Q > 0$).

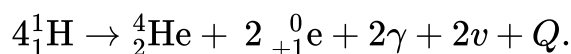
An important example of nuclear fusion in nature is the production of energy in the Sun. In 1938, Hans Bethe proposed that the Sun produces energy when hydrogen nuclei (${}^1\text{H}$) fuse into stable helium nuclei (${}^4\text{He}$) in the Sun's core ([\[link\]](#)). This process, called the **proton-proton chain**, is summarized by three reactions:

Equation:

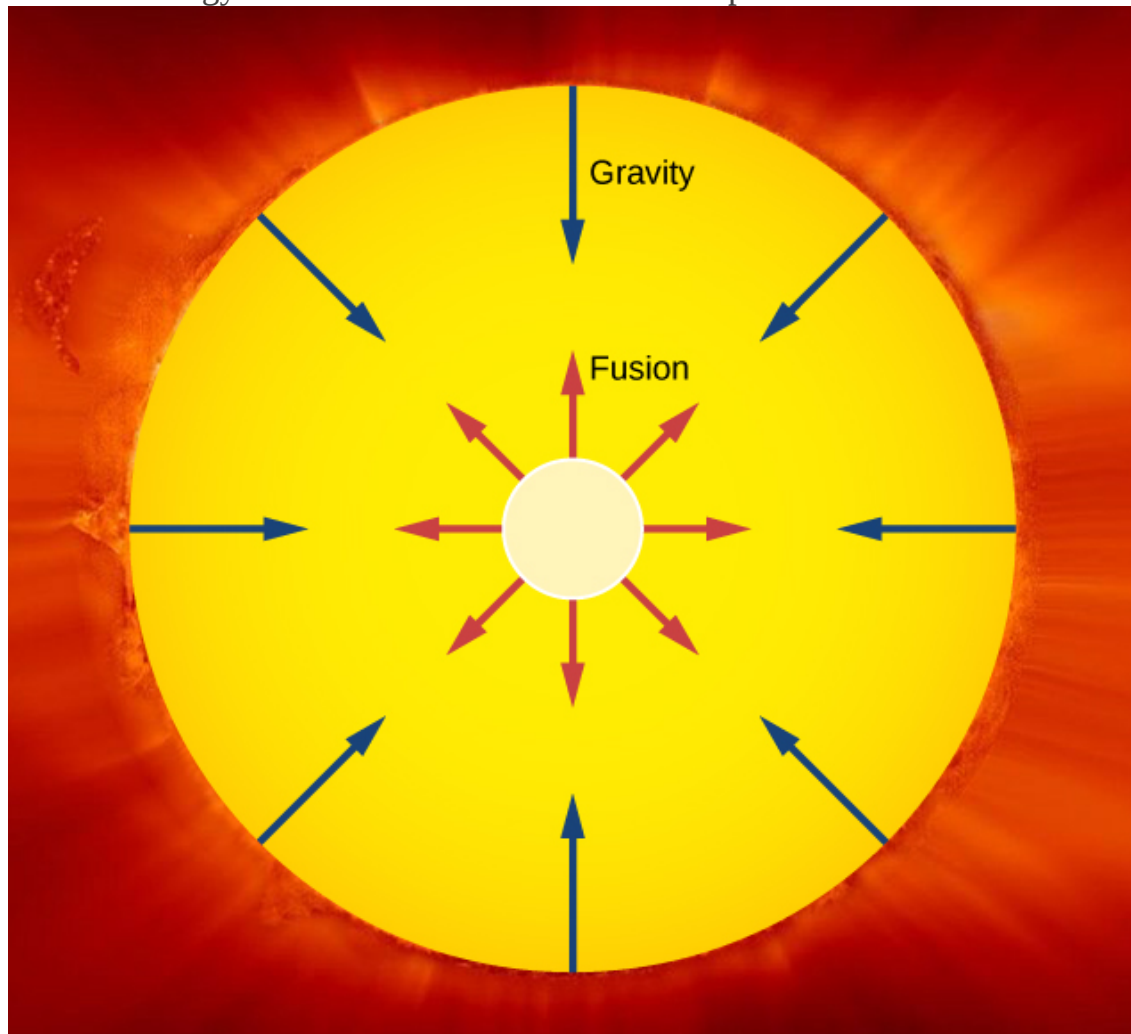


Thus, a stable helium nucleus is formed from the fusion of the nuclei of the hydrogen atom. These three reactions can be summarized by

Equation:



The net Q value is about 26 MeV. The release of this energy produces an outward thermal gas pressure that prevents the Sun from gravitational collapse. Astrophysicists find that hydrogen fusion supplies the energy stars require to maintain energy balance over most of a star's life span.

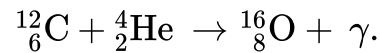
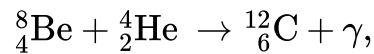
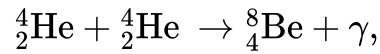


The Sun produces energy by fusing hydrogen into helium at the Sun's core. The red arrows show outward pressure due to thermal gas, which tends to make the Sun expand. The blue arrows show inward pressure due to gravity, which tends to make the Sun contract. These two influences balance each other.

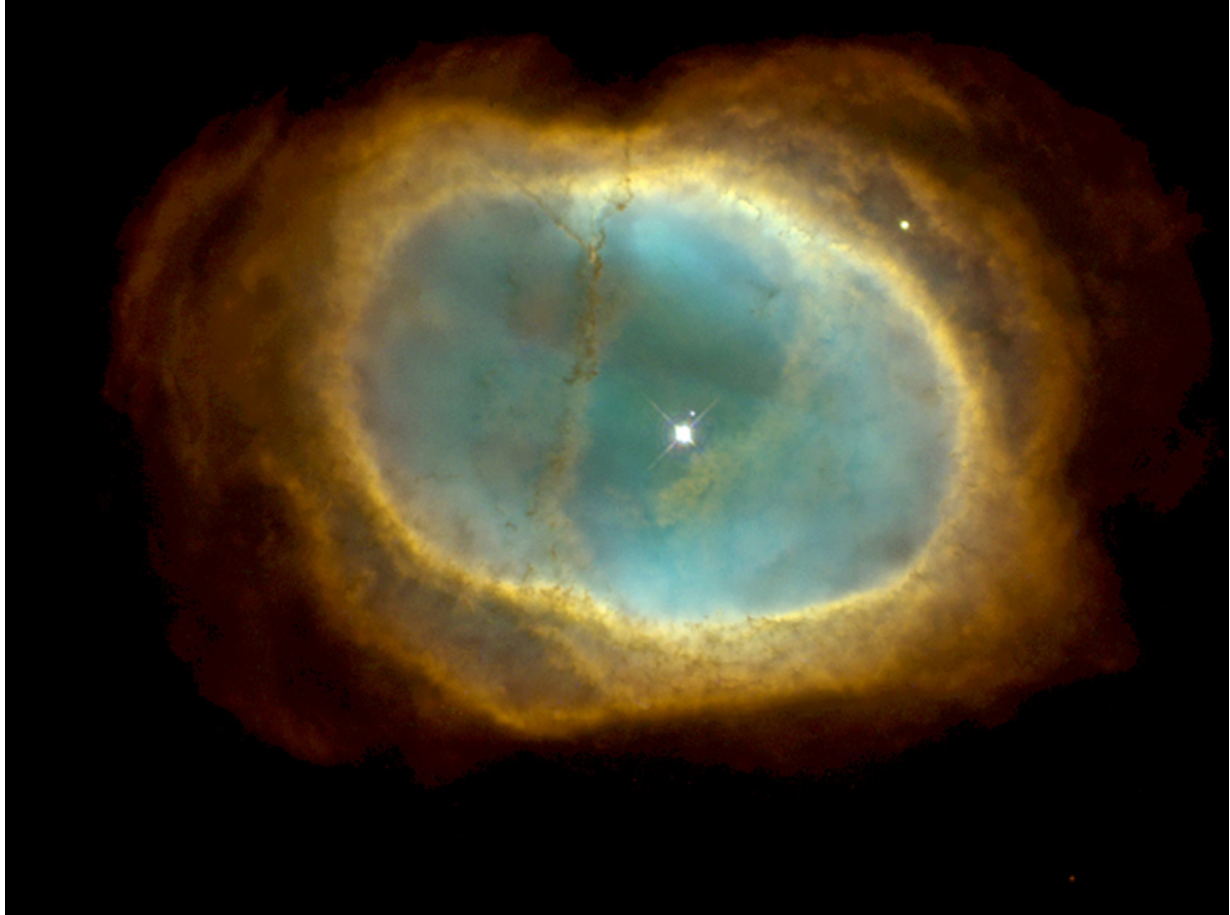
Nucleosynthesis

Scientists now believe that many heavy elements found on Earth and throughout the universe were originally synthesized by fusion within the hot cores of the stars. This process is known as **nucleosynthesis**. For example, in lighter stars, hydrogen combines to form helium through the proton-proton chain. Once the hydrogen fuel is exhausted, the star enters the next stage of its life and fuses helium. An example of a nuclear reaction chain that can occur is:

Equation:



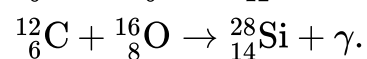
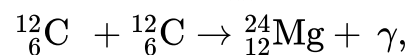
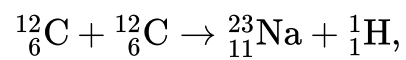
Carbon and oxygen nuclei produced in such processes eventually reach the star's surface by convection. Near the end of its lifetime, the star loses its outer layers into space, thus enriching the interstellar medium with the nuclei of heavier elements ([link](#)).



A planetary nebula is produced at the end of the life of a star. The greenish color of this planetary nebula comes from oxygen ions. (credit: Hubble Heritage Team (STScI/AURA/NASA/ESA))

Stars similar in mass to the Sun do not become hot enough to fuse nuclei as heavy (or heavier) than oxygen nuclei. However, in massive stars whose cores become much hotter ($T > 6 \times 10^8 \text{ K}$), even more complex nuclei are produced. Some representative reactions are

Equation:



Nucleosynthesis continues until the core is primarily iron-nickel metal. Now, iron has the peculiar property that any fusion or fission reaction involving the iron nucleus is endothermic, meaning that energy is absorbed rather than produced. Hence, nuclear energy cannot be generated in an iron-rich core. Lacking an outward pressure from fusion reactions, the star begins to contract due to gravity. This process heats the core to a temperature on the order of $5 \times 10^9 \text{K}$. Expanding shock waves generated within the star due to the collapse cause the star to quickly explode. The luminosity of the star can increase temporarily to nearly that of an entire galaxy. During this event, the flood of energetic neutrons reacts with iron and the other nuclei to produce elements heavier than iron. These elements, along with much of the star, are ejected into space by the explosion. Supernovae and the formation of planetary nebulae together play a major role in the dispersal of chemical elements into space.

Eventually, much of the material lost by stars is pulled together through the gravitational force, and it condenses into a new generation of stars and accompanying planets. Recent images from the Hubble Space Telescope provide a glimpse of this magnificent process taking place in the constellation Serpens ([link](#)). The new generation of stars begins the nucleosynthesis process anew, with a higher percentage of heavier elements. Thus, stars are “factories” for the chemical elements, and many of the atoms in our bodies were once a part of stars.



This image taken by NASA's Spitzer Space Telescope and the Two Micron All Sky Survey (2MASS), shows the Serpens Cloud Core, a star-forming region in the constellation Serpens (the “Serpent”). Located about 750 light-years away, this cluster of stars is formed

700 light years away, this cluster of stars is formed from cooling dust and gases. Infrared light has been used to reveal the youngest stars in orange and yellow.
(credit: NASA/JPL-Caltech/2MASS)

Example:**Energy of the Sun**

The power output of the Sun is approximately 3.8×10^{26} J/s. Most of this energy is produced in the Sun's core by the proton-proton chain. This energy is transmitted outward by the processes of convection and radiation. (a) How many of these fusion reactions per second must occur to supply the power radiated by the Sun? (b) What is the rate at which the mass of the Sun decreases? (c) In about five billion years, the central core of the Sun will be depleted of hydrogen. By what percentage will the mass of the Sun have decreased from its present value when the core is depleted of hydrogen?

Strategy

The total energy output per second is given in the problem statement. If we know the energy released in each fusion reaction, we can determine the rate of the fusion reactions. If the mass loss per fusion reaction is known, the mass loss rate is known. Multiplying this rate by five billion years gives the total mass lost by the Sun. This value is divided by the original mass of the Sun to determine the percentage of the Sun's mass that has been lost when the hydrogen fuel is depleted.

Solution

- a. The decrease in mass for the fusion reaction is

Equation:

$$\begin{aligned}\Delta m &= 4m({}_1^1\text{H}) - m({}_2^4\text{He}) - 2m({}_{+1}^0\text{e}) \\ &= 4(1.007825 \text{ u}) - 4.002603 \text{ u} - 2(0.000549 \text{ u}) \\ &= 0.0276 \text{ u}.\end{aligned}$$

The energy released per fusion reaction is

Equation:

$$Q = (0.0276 \text{ u})(931.49 \text{ MeV/u}) = 25.7 \text{ MeV}.$$

Thus, to supply 3.8×10^{26} J/s = 2.38×10^{39} MeV/s, there must be

Equation:

$$\frac{2.38 \times 10^{39} \text{ MeV/s}}{25.7 \text{ MeV/reaction}} = 9.26 \times 10^{37} \text{ reaction/s.}$$

- b. The Sun's mass decreases by $0.0276 \text{ u} = 4.58 \times 10^{-29} \text{ kg}$ per fusion reaction, so the rate at which its mass decreases is

Equation:

$$(9.26 \times 10^{37} \text{ reaction/s}) (4.58 \times 10^{-29} \text{ kg/reaction}) = 4.24 \times 10^9 \text{ kg/s.}$$

- c. In $5 \times 10^9 \text{ y} = 1.6 \times 10^{17} \text{ s}$, the Sun's mass will therefore decrease by

Equation:

$$\Delta M = (4.24 \times 10^9 \text{ kg/s}) (1.6 \times 10^{17} \text{ s}) = 6.8 \times 10^{26} \text{ kg.}$$

The current mass of the Sun is about $2.0 \times 10^{30} \text{ kg}$, so the percentage decrease in its mass when its hydrogen fuel is depleted will be

Equation:

$$\left(\frac{6.8 \times 10^{26} \text{ kg}}{2.0 \times 10^{30} \text{ kg}} \right) \times 100\% = 0.034\%.$$

Significance

After five billion years, the Sun is very nearly the same mass as it is now. Hydrogen burning does very little to change the mass of the Sun. This calculation assumes that only the proton-proton decay change is responsible for the power output of the Sun.

Note:

Exercise:

Problem:

Check Your Understanding Where does the energy from the Sun originate?

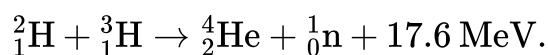
Solution:

the conversion of mass to energy

The Hydrogen Bomb

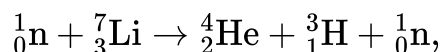
In 1942, Robert Oppenheimer suggested that the extremely high temperature of an atomic bomb could be used to trigger a fusion reaction between deuterium and tritium, thus producing a fusion (or hydrogen) bomb. The reaction between deuterium and tritium, both isotopes of hydrogen, is given by

Equation:



Deuterium is relatively abundant in ocean water but tritium is scarce. However, tritium can be generated in a nuclear reactor through a reaction involving lithium. The neutrons from the reactor cause the reaction

Equation:

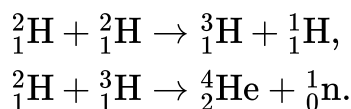


to produce the desired tritium. The first hydrogen bomb was detonated in 1952 on the remote island of Eniwetok in the Marshall Islands. A hydrogen bomb has never been used in war. Modern hydrogen bombs are approximately 1000 times more powerful than the fission bombs dropped on Hiroshima and Nagasaki in World War II.

The Fusion Reactor

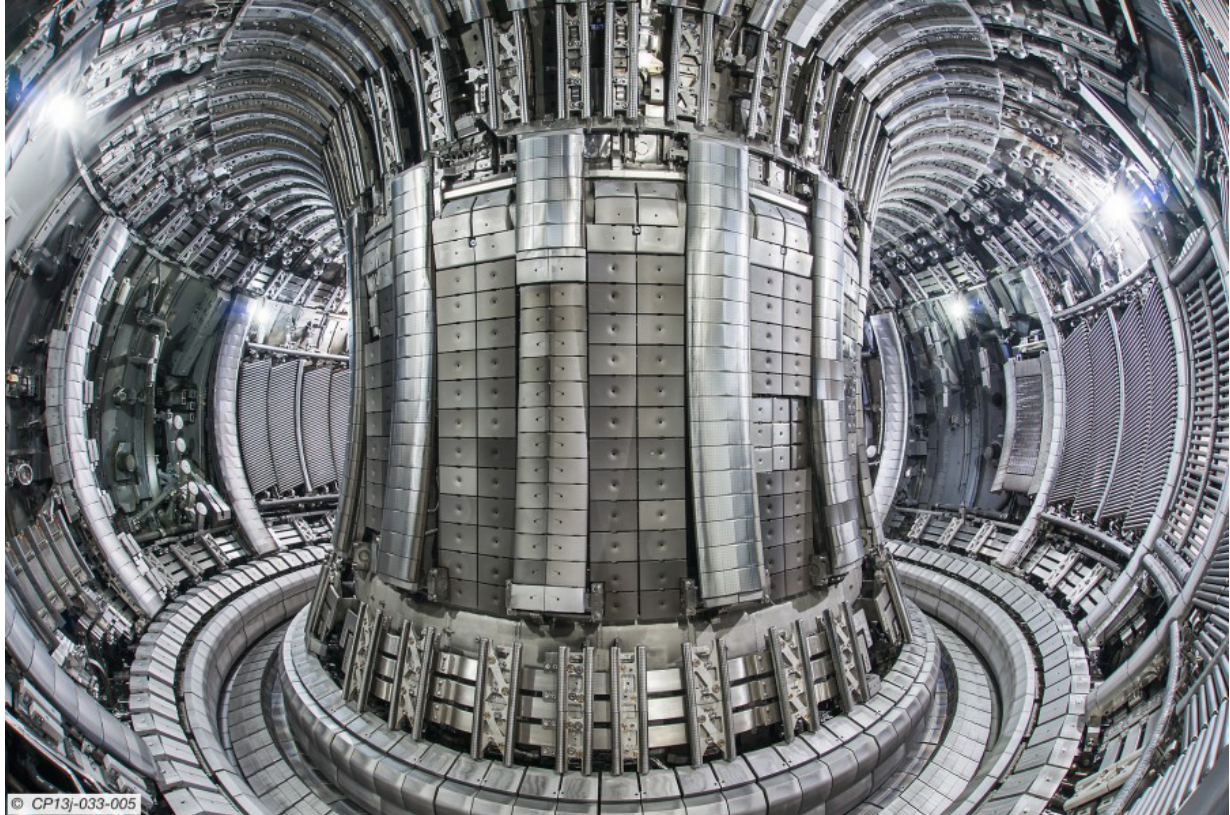
The fusion chain believed to be the most practical for use in a **nuclear fusion reactor** is the following two-step process:

Equation:



This chain, like the proton-proton chain, produces energy without any radioactive by-product. However, there is a very difficult problem that must be overcome before fusion can be used to produce significant amounts of energy: Extremely high temperatures ($\sim 10^7\text{K}$) are needed to drive the fusion process. To meet this challenge, test fusion reactors are being developed to withstand temperatures 20 times greater than the Sun's core temperature. An example is the Joint European Torus (JET) shown in [\[link\]](#). A great deal of work still has to be done on fusion reactor

technology, but many scientists predict that fusion energy will power the world's cities by the end of the twentieth century.



The Joint European Torus (JET) tokamak fusion reactor uses magnetic fields to fuse deuterium and tritium nuclei (credit: EUROfusion).

Summary

- Nuclear fusion is a reaction in which two nuclei are combined to form a larger nucleus; energy is released when light nuclei are fused to form medium-mass nuclei.
- The amount of energy released by a fusion reaction is known as the Q value.
- Nuclear fusion explains the reaction between deuterium and tritium that produces a fusion (or hydrogen) bomb; fusion also explains the production of energy in the Sun, the process of nucleosynthesis, and the creation of the heavy elements.

Conceptual Questions

Exercise:

Problem: Explain the difference between nuclear fission and nuclear fusion.

Exercise:

Problem:

Why does the fusion of light nuclei into heavier nuclei release energy?

Solution:

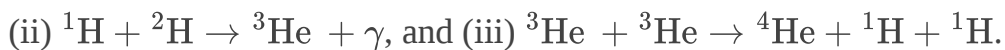
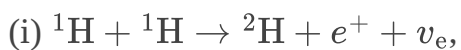
The nuclei produced in the fusion process have a larger binding energy per nucleon than the nuclei that are fused. That is, nuclear fusion decreases average energy of the nucleons in the system. The energy difference is carried away as radiation.

Problems

Exercise:

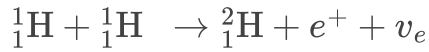
Problem:

Verify that the total number of nucleons, and total charge are conserved for each of the following fusion reactions in the proton-proton chain.



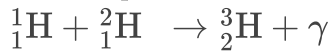
(List the value of each of the conserved quantities before and after each of the reactions.)

Solution:



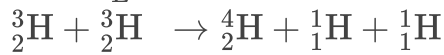
$$\text{i.} \quad A_i = 1 + 1 = 2; A_f = 2 \quad Z_i = 1 + 1 = 2;$$

$$Z_f = 1 + 1 = 2$$



$$\text{ii.} \quad A_i = 1 + 2 = 3; A_f = 3 + 0 = 3 \quad Z_i = 1 + 1 = 2;$$

$$Z_f = 1 + 1 = 2$$



$$\text{iii.} \quad A_i = 3 + 3 = 6; A_f = 4 + 1 + 1 = 6 \quad Z_i = 2 + 2 = 4$$

$$Z_f = 2 + 1 + 1 = 4$$

Exercise:

Problem:

Calculate the energy output in each of the fusion reactions in the proton-proton chain, and verify the values determined in the preceding problem.

Exercise:

Problem:

Show that the total energy released in the proton-proton chain is 26.7 MeV, considering the overall effect in ${}^1_1\text{H} + {}^1_1\text{H} \rightarrow {}^2_1\text{H} + e^+ + \nu_e$, ${}^1_1\text{H} + {}^2_1\text{H} \rightarrow {}^3_2\text{He} + \gamma$, and ${}^3_2\text{He} + {}^3_2\text{He} \rightarrow {}^4_2\text{He} + {}^1_1\text{H} + {}^1_1\text{H}$. Be sure to include the annihilation energy.

Solution:

26.73 MeV

Exercise:

Problem:

Two fusion reactions mentioned in the text are $n + {}^3_2\text{He} \rightarrow {}^4_2\text{He} + \gamma$ and $n + {}^1_1\text{H} \rightarrow {}^2_1\text{H} + \gamma$. Both reactions release energy, but the second also creates more fuel. Confirm that the energies produced in the reactions are 20.58 and 2.22 MeV, respectively. Comment on which product nuclide is most tightly bound, ${}^4_2\text{He}$ or ${}^2_1\text{H}$.

Exercise:

Problem:

The power output of the Sun is $4 \times 10^{26} \text{ W}$. (a) If 90% of this energy is supplied by the proton-proton chain, how many protons are consumed per second? (b) How many neutrinos per second should there be per square meter at the surface of Earth from this process?

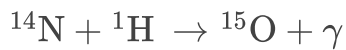
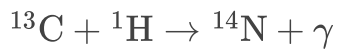
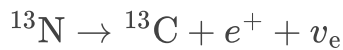
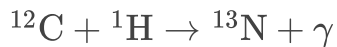
Solution:

a. 3×10^{38} protons/s; b. 6×10^{14} neutrinos/ $\text{m}^2 \cdot \text{s}$;

This huge number is indicative of how rarely a neutrino interacts, since large detectors observe very few per day.

Exercise:**Problem:**

Another set of reactions that fuses hydrogen into helium in the Sun and especially in hotter stars is called the CNO cycle:



This process is a “cycle” because ^{12}C appears at the beginning and end of these reactions. Write down the overall effect of this cycle (as done for the proton-proton chain in $2e^- + 4^1\text{H} \rightarrow ^4\text{He} + 2\nu_e + 6\gamma$). Assume that the positrons annihilate electrons to form more γ rays.

Exercise:**Problem:**

(a) Calculate the energy released by the fusion of a 1.00-kg mixture of deuterium and tritium, which produces helium. There are equal numbers of deuterium and tritium nuclei in the mixture.

(b) If this process takes place continuously over a period of a year, what is the average power output?

Solution:

a. The atomic mass of deuterium (^2H) is 2.014102 u, while that of tritium (^3H) is 3.016049 u, for a total of 5.032151 u per reaction. So a mole of reactants has a mass of 5.03 g, and in 1.00 kg, there are $(1000 \text{ g}) / (5.03 \text{ g/mol}) = 198.8 \text{ mol}$ of reactants. The number of reactions that take place is therefore $(198.8 \text{ mol}) (6.02 \times 10^{23} \text{ mol}^{-1}) = 1.20 \times 10^{26}$ reactions.

The total energy output is the number of reactions times the energy per reaction:
 $E = 3.37 \times 10^{14} \text{ J};$

b. Power is energy per unit time. One year has $3.16 \times 10^7 \text{ s}$, so
 $P = 10.7 \text{ MW}.$

We expect nuclear processes to yield large amounts of energy, and this is certainly the case here. The energy output of $3.37 \times 10^{14} \text{ J}$ from fusing 1.00 kg of deuterium and tritium is equivalent to 2.6 million gallons of gasoline and about eight times the energy output of the bomb that destroyed Hiroshima. Yet the average backyard swimming pool has about 6 kg of deuterium in it, so that fuel is plentiful if it can be utilized in a controlled manner.

Glossary

nuclear fusion

process of combining lighter nuclei to make heavier nuclei

nuclear fusion reactor

nuclear reactor that uses the fusion chain to produce energy

nucleosynthesis

process of fusion by which all elements on Earth are believed to have been created

proton-proton chain

combined reactions that fuse hydrogen nuclei to produce He nuclei

Medical Applications and Biological Effects of Nuclear Radiation

By the end of this section, you will be able to:

- Describe two medical uses of nuclear technology
- Explain the origin of biological effects due to nuclear radiation
- List common sources of radiation and their effects
- Estimate exposure for nuclear radiation using common dosage units

Nuclear physics is an integral part of our everyday lives ([link](#)). Radioactive compounds are used to identify cancer, study ancient artifacts, and power our cities. Nuclear fusion also powers the Sun, the primary source of energy on Earth. The focus of this chapter is nuclear radiation. In this section, we ask such questions as: How is nuclear radiation used to benefit society? What are its health risks? How much nuclear radiation is the average person exposed to in a lifetime?

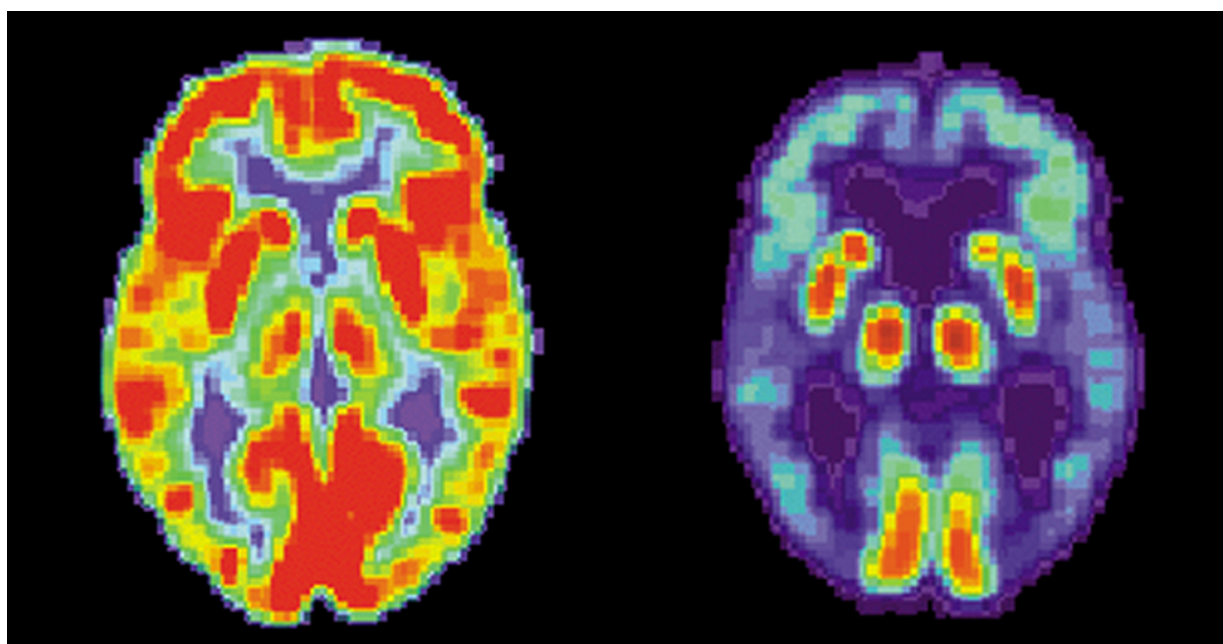


Dr. Tori Randall, a curator at the San Diego Museum of Man, uses nuclear radiation to study a 500-year-old Peruvian child mummy. The origin of this radiation is the transformation of one nucleus to another. (credit: Samantha A. Lewis, U.S. Navy)

Medical Applications

Medical use of nuclear radiation is quite common in today's hospitals and clinics. One of the most important uses of nuclear radiation is the location and study of diseased tissue. This

application requires a special drug called a **radiopharmaceutical**. A radiopharmaceutical contains an unstable radioactive isotope. When the drug enters the body, it tends to concentrate in inflamed regions of the body. (Recall that the interaction of the drug with the body does not depend on whether a given nucleus is replaced by one of its isotopes, since this interaction is determined by chemical interactions.) Radiation detectors used outside the body use nuclear radiation from the radioisotopes to locate the diseased tissue. Radiopharmaceuticals are called **radioactive tags** because they allow doctors to track the movement of drugs in the body. Radioactive tags are for many purposes, including the identification of cancer cells in the bones, brain tumors, and Alzheimer's disease ([\[link\]](#)). Radioactive tags are also used to monitor the function of body organs, such as blood flow, heart muscle activity, and iodine uptake in the thyroid gland.



These brain images are produced using a radiopharmaceutical. The colors indicate relative metabolic or biochemical activity (red indicates high activity and blue indicates low activity). The figure on the left shows the normal brain of an individual and the figure on the right shows the brain of someone diagnosed with Alzheimer's disease. The brain image of the normal brain indicates much greater metabolic activity (a larger fraction of red and orange areas). (credit: modification of works by National Institutes of Health)

[\[link\]](#) lists some medical diagnostic uses of radiopharmaceuticals, including isotopes and typical activity (A) levels. One common diagnostic test uses iodine to image the thyroid, since iodine is concentrated in that organ. Another common nuclear diagnostic is the thallium scan for the cardiovascular system, which reveals blockages in the coronary arteries and

examines heart activity. The salt TlCl can be used because it acts like NaCl and follows the blood. Note that [\[link\]](#) lists many diagnostic uses for $^{99\text{m}}\text{Tc}$, where “m” stands for a metastable state of the technetium nucleus. This isotope is used in many compounds to image the skeleton, heart, lungs, and kidneys. About 80% of all radiopharmaceuticals employ $^{99\text{m}}\text{Tc}$ because it produces a single, easily identified, 0.142-MeV γ ray and has a short 6.0-h half-life, which reduces radiation exposure.

Procedure, Isotope	Activity (mCi), where $1 \text{ mCi} = 3.7 \times 10^7 \text{ Bq}$	Procedure, Isotope	Activity (mCi), where $1 \text{ mCi} = 3.7 \times 10^7 \text{ Bq}$
<i>Brain scan</i>		<i>Thyroid scan</i>	
$^{99\text{m}}\text{Tc}$	7.5	^{131}I	0.05
^{15}O (PET)	50	^{123}I	0.07
<i>Lung scan</i>		<i>Liver scan</i>	
^{133}Xe	7.5	^{198}Au (colloid)	0.1
$^{99\text{m}}\text{Tc}$	2	$^{99\text{m}}\text{Tc}$ (colloid)	2
<i>Cardiovascular blood pool</i>		<i>Bone scan</i>	
^{131}I	0.2	^{85}Sr	0.1
$^{99\text{m}}\text{Tc}$	2	$^{99\text{m}}\text{Tc}$	10
<i>Cardiovascular arterial flow</i>		<i>Kidney scan</i>	
^{201}Tl	3	^{197}Hg	0.1
^{24}Na	7.5	$^{99\text{m}}\text{Tc}$	1.5

Diagnostic Uses of Radiopharmaceuticals

The first radiation detectors produced two-dimensional images, like a photo taken from a camera. However, a circular array of detectors that can be rotated can be used to produce three-dimensional images. This technique is similar to that used in X-ray computed

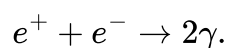
tomography (CT) scans. One application of this technique is called **single-photon-emission CT (SPECT)** ([\[link\]](#)). The spatial resolution of this technique is about 1 cm.



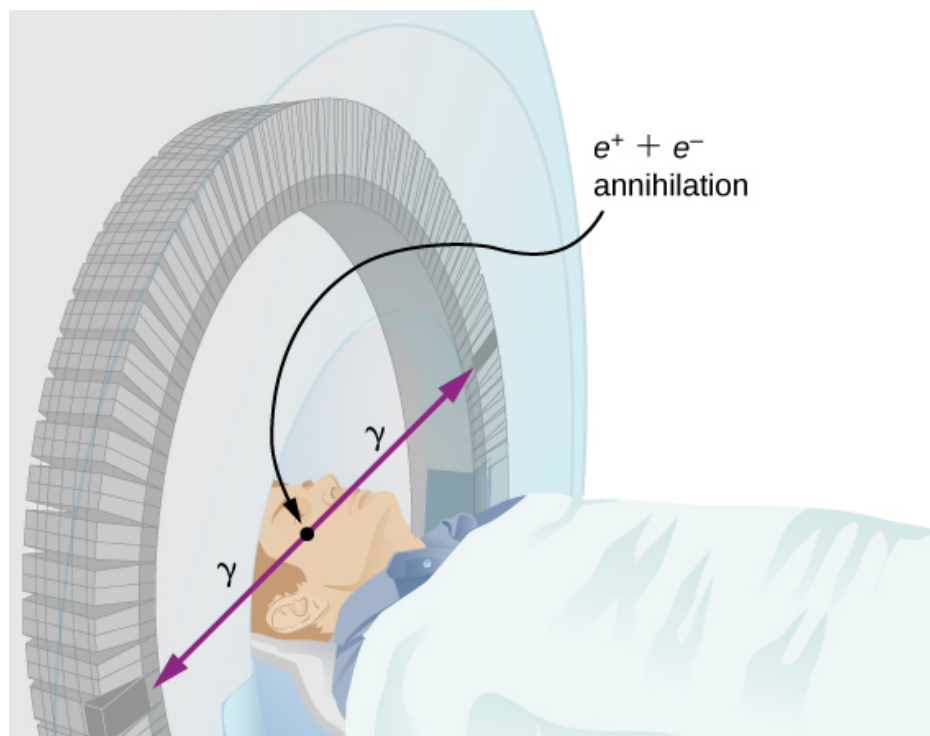
The SPECT machine uses radiopharmaceutical compounds to produce an image of the human body. The machine takes advantage of the physics of nuclear beta decays and electron-positron collisions. (credit: “Woldo”/Wikimedia Commons)

Improved image resolution is achieved by a technique known as **positron emission tomography (PET)**. This technique uses radioisotopes that decay by β^+ radiation. When a positron encounters an electron, these particles annihilate to produce two gamma-ray photons. This reaction is represented by

Equation:



These γ -ray photons have identical 0.511-MeV energies and move directly away from one another ([\[link\]](#)). This easily identified decay signature can be used to identify the location of the radioactive isotope. Examples of β^+ -emitting isotopes used in PET include ^{11}C , ^{13}N , ^{15}O , and ^{18}F . The nuclei have the advantage of being able to function as tags for natural body compounds. Its resolution of 0.5 cm is better than that of SPECT.



A PET system takes advantage of the two identical γ -ray photons produced by positron-electron annihilation. These γ rays are emitted in opposite directions, so that the line along which each pair is emitted is determined.

PET scans are especially useful to examine the brain's anatomy and function. For example, PET scans can be used to monitor the brain's use of oxygen and water, identify regions of decreased metabolism (linked to Alzheimer's disease), and locate different parts of the brain responsible for sight, speech, and fine motor activity

Note:

Is it a tumor? View an [animation](#) of simplified magnetic resonance imaging (MRI) to see if you can tell. Your head is full of tiny radio transmitters (the nuclear spins of the hydrogen

nuclei of your water molecules). In an MRI unit, these little radios can be made to broadcast their positions, giving a detailed picture of the inside of your head.

Biological Effects

Nuclear radiation can have both positive and negative effects on biological systems. However, it can also be used to treat and even cure cancer. How do we understand these effects? To answer this question, consider molecules within cells, particularly DNA molecules.

Cells have long, double-helical DNA molecules containing chemical codes that govern the function and processes of the cell. Nuclear radiation can alter the structural features of the DNA chain, leading to changes in the genetic code. In human cells, we can have as many as a million individual instances of damage to DNA per cell per day. DNA contains codes that check whether the DNA is damaged and can repair itself. This repair ability of DNA is vital for maintaining the integrity of the genetic code and for the normal functioning of the entire organism. It should be constantly active and needs to respond rapidly. The rate of DNA repair depends on various factors such as the type and age of the cell. If nuclear radiation damages the ability of the cell to repair DNA, the cell can

1. Retreat to an irreversible state of dormancy (known as senescence);
2. Commit suicide (known as programmed cell death); or
3. Progress into unregulated cell division, possibly leading to tumors and cancers.

Nuclear radiation can harm the human body in many other ways as well. For example, high doses of nuclear radiation can cause burns and even hair loss.

Biological effects of nuclear radiation are expressed by many different physical quantities and in many different units. A common unit to express the biological effects of nuclear radiation is the **rad** or **radiation dose unit**. One rad is equal to 1/100 of a joule of nuclear energy deposited per kilogram of tissue, written:

Equation:

$$1 \text{ rad} = 0.01 \text{ J/kg}.$$

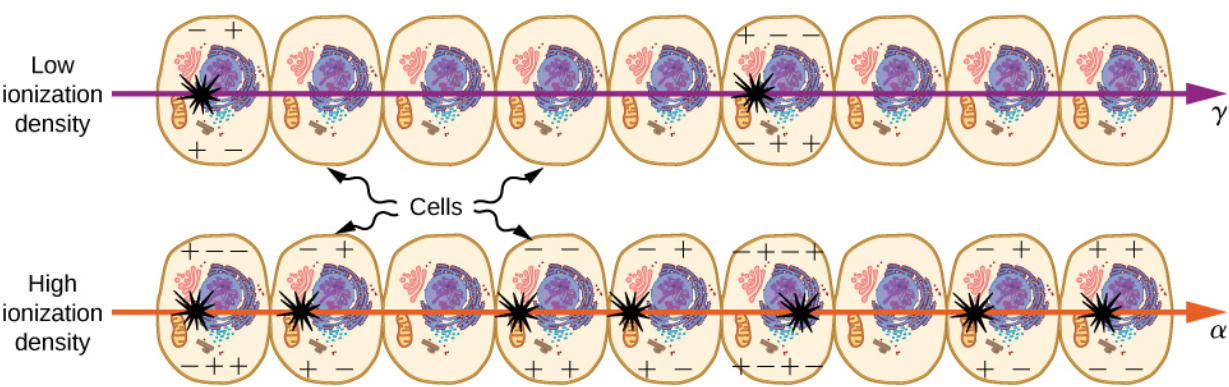
For example, if a 50.0-kg person is exposed to nuclear radiation over her entire body and she absorbs 1.00 J, then her whole-body radiation dose is

Equation:

$$(1.00 \text{ J}) / (50.0 \text{ kg}) = 0.0200 \text{ J/kg} = 2.00 \text{ rad}.$$

Nuclear radiation damages cells by ionizing atoms in the cells as they pass through the cells ([link](#)). The effects of ionizing radiation depend on the dose in rads, but also on the type of

radiation (alpha, beta, gamma, or X-ray) and the type of tissue. For example, if the range of the radiation is small, as it is for α rays, then the ionization and the damage created is more concentrated and harder for the organism to repair. To account for such affects, we define the **relative biological effectiveness (RBE)**. Sample RBE values for several types of ionizing nuclear radiation are given in [\[link\]](#).



The image shows ionization created in cells by α and γ radiation. Because of its shorter range, the ionization and damage created by α rays is more concentrated and harder for the organism to repair. Thus, the RBE for α rays is greater than the RBE for γ rays, even though they create the same amount of ionization at the same energy.

Type and Energy of Radiation	RBE ^[1]
X-rays	1
γ rays	1
β rays greater than 32 keV	1
β rays less than 32 keV	1.7
Neutrons, thermal to slow (<20 keV)	2–5
Neutrons, fast (1–10 MeV)	10 (body), 32 (eyes)

Type and Energy of Radiation	RBE ^[1]
Protons (1–10 MeV)	10 (body), 32 (eyes)
α rays from radioactive decay	10–20
Heavy ions from accelerators	10–20

Relative Biological Effectiveness^[1] Values approximate. Difficult to determine.

A dose unit more closely related to effects in biological tissue is called the **roentgen equivalent man (rem)** and is defined to be the dose (in rads) multiplied by the relative biological effectiveness (RBE). Thus, if a person had a whole-body dose of 2.00 rad of γ radiation, the dose in rem would be $(2.00 \text{ rad}) (1) = 2.00 \text{ rem}$ for the whole body. If the person had a whole-body dose of 2.00 rad of α radiation, then the dose in rem would be $(2.00 \text{ rad}) (20) = 40.0 \text{ rem}$ for the whole body. The α rays would have 20 times the effect on the person than the γ rays for the same deposited energy. The SI equivalent of the rem, and the more standard term, is the **sievert (Sv)** is

Equation:

$$1 \text{ Sv} = 100 \text{ rem.}$$

The RBEs given in [\[link\]](#) are approximate but reflect an understanding of nuclear radiation and its interaction with living tissue. For example, neutrons are known to cause more damage than γ rays, although both are neutral and have large ranges, due to secondary radiation. Any dose less than 100 mSv (10 rem) is called a **low dose**, 0.1 Sv to 1 Sv (10 to 100 rem) is called a **moderate dose**, and anything greater than 1 Sv (100 rem) is called a **high dose**. It is difficult to determine if a person has been exposed to less than 10 mSv.

Biological effects of different levels of nuclear radiation on the human body are given in [\[link\]](#). The first clue that a person has been exposed to radiation is a change in blood count, which is not surprising since blood cells are the most rapidly reproducing cells in the body. At higher doses, nausea and hair loss are observed, which may be due to interference with cell reproduction. Cells in the lining of the digestive system also rapidly reproduce, and their destruction causes nausea. When the growth of hair cells slows, the hair follicles become thin and break off. High doses cause significant cell death in all systems, but the lowest doses that cause fatalities do so by weakening the immune system through the loss of white blood cells.

Dose in Sv^[1]	Effect
0–0.10	No observable effect.
0.1–1	Slight to moderate decrease in white blood cell counts.
0.5	Temporary sterility; 0.35 for women, 0.50 for men.
1–2	Significant reduction in blood cell counts, brief nausea and vomiting. Rarely fatal.
2–5	Nausea, vomiting, hair loss, severe blood damage, hemorrhage, fatalities.
4.5	Lethal to 50% of the population within 32 days after exposure if not treated.
5–20	Worst effects due to malfunction of small intestine and blood systems. Limited survival.
>20	Fatal within hours due to collapse of central nervous system.

Immediate Effects of Radiation (Adults, Whole-Body, Single Exposure)^[1] Multiply by 100 to obtain dose in rem.

Sources of Radiation

Human are also exposed to many sources of nuclear radiation. A summary of average radiation doses for different sources by country is given in [\[link\]](#). Earth emits radiation due to the isotopes of uranium, thorium, and potassium. Radiation levels from these sources depend on location and can vary by a factor of 10. Fertilizers contain isotopes of potassium and uranium, which we digest in the food we eat. Fertilizers have more than 3000 Bq/kg radioactivity, compared to just 66 Bq/kg for Carbon-14.

Source	Dose (mSv/y)^[1]			
	Australia	Germany	US	World

Source	Dose (mSv/y) ^[1]			
Natural radiation – external				
Cosmic rays	0.30	0.28	0.30	0.39
Soil, building materials	0.40	0.40	0.30	0.48
Radon gas	0.90	1.1	2.0	1.2
Natural radiation – internal				
⁴⁰ K, ¹⁴ C, ²²⁶ Ra	0.24	0.28	0.40	0.29
Artificial radiation				
Medical and dental	0.80	0.90	0.53	0.40
TOTAL	2.6	3.0	3.5	2.8

Background Radiation Sources and Average Doses^[1] Multiply by 100 to obtain does in mrem/y.

Medical visits are also a source of nuclear radiation. A sample of common nuclear radiation doses is given in [\[link\]](#). These doses are generally low and can be lowered further with improved techniques and more sensitive detectors. With the possible exception of routine dental X-rays, medical use of nuclear radiation is used only when the risk-benefit is favorable. Chest X-rays give the lowest doses—about 0.1 mSv to the tissue affected, with less than 5% scattering into tissues that are not directly imaged. Other X-ray procedures range upward to about 10 mSv in a CT scan, and about 5 mSv (0.5 rem) per dental X-ray, again both only affecting the tissue imaged. Medical images with radiopharmaceuticals give doses ranging from 1 to 5 mSv, usually localized.

Procedure	Effective Dose (mSv)
Chest	0.02
Dental	0.01
Skull	0.07

Procedure	Effective Dose (mSv)
Leg	0.02
Mammogram	0.40
Barium enema	7.0
Upper GI	3.0
CT head	2.0
CT abdomen	10.0

Typical Doses Received During Diagnostic X-Ray Exams

Example:

What Mass of ^{137}Cs Escaped Chernobyl?

The Chernobyl accident in Ukraine (formerly in the Soviet Union) exposed the surrounding population to a large amount of radiation through the decay of ^{137}Cs . The initial radioactivity level was approximately $A = 6.0 \text{ MCi}$. Calculate the total mass of ^{137}Cs involved in this accident.

Strategy

The total number of nuclei, N , can be determined from the known half-life and activity of ^{137}Cs (30.2 y). The mass can be calculated from N using the concept of a mole.

Solution

Solving the equation $A = \frac{0.693 N}{t_{1/2}}$ for N gives

Equation:

$$N = \frac{A t_{1/2}}{0.693}.$$

Entering the given values yields

Equation:

$$N = \frac{(6.0 \text{ MCi}) (30.2 \text{ y})}{0.693}.$$

To convert from curies to becquerels and years to seconds, we write

Equation:

$$N = \frac{(6.0 \times 10^6 \text{ Ci}) (3.7 \times 10^{10} \text{ Bq/Ci}) (30.2 \text{ y}) (3.16 \times 10^7 \text{ s/y})}{0.693} = 3.1 \times 10^{26}.$$

One mole of a nuclide AX has a mass of A grams, so that one mole of ${}^{137}\text{Cs}$ has a mass of 137 g. A mole has 6.02×10^{23} nuclei. Thus the mass of ${}^{137}\text{Cs}$ released was

Equation:

$$m = \left(\frac{137 \text{ g}}{6.02 \times 10^{23}} \right) (3.1 \times 10^{26}) = 70 \times 10^3 \text{ g} = 70 \text{ kg}.$$

Significance

The mass of ${}^{137}\text{Cs}$ involved in the Chernobyl accident is a small material compared to the typical amount of fuel used in a nuclear reactor. However, approximately 250 people were admitted to local hospitals immediately after the accident, and diagnosed as suffering acute radiation syndrome. They received external radiation dosages between 1 and 16 Sv. Referring to biological effects in [\[link\]](#), these dosages are extremely hazardous. The eventual death toll is estimated to be around 4000 people, primarily due to radiation-induced cancer.

Note:

Exercise:

Problem:

Check Your Understanding Radiation propagates in all directions from its source, much as electromagnetic radiation from a light bulb. Is *activity* concept more analogous to power, intensity, or brightness?

Solution:

power

Summary

- Nuclear technology is used in medicine to locate and study diseased tissue using special drugs called radiopharmaceuticals. Radioactive tags are used to identify cancer cells in the bones, brain tumors, and Alzheimer's disease, and to monitor the function of body organs, such as blood flow, heart muscle activity, and iodine uptake in the thyroid gland.
- The biological effects of ionizing radiation are due to two effects it has on cells: interference with cell reproduction and destruction of cell function.
- Common sources of radiation include that emitted by Earth due to the isotopes of uranium, thorium, and potassium; natural radiation from cosmic rays, soils, and building materials, and artificial sources from medical and dental diagnostic tests.

- Biological effects of nuclear radiation are expressed by many different physical quantities and in many different units, including the rad or radiation dose unit.

Key Equations

Atomic mass number	$A = Z + N$
Standard format for expressing an isotope	${}^A_Z\text{X}$
Nuclear radius, where r_0 is the radius of a single proton	$r = r_0 A^{1/3}$
Mass defect	$\Delta m = Zm_p + (A - Z)m_n - m_{\text{nuc}}$
Binding energy	$E = (\Delta m)c^2$
Binding energy per nucleon	$BEN = \frac{E_b}{A}$
Radioactive decay rate	$-\frac{dN}{dt} = \lambda N$
Radioactive decay law	$N = N_0 e^{-\lambda t}$
Decay constant	$\lambda = \frac{0.693}{T_{1/2}}$
Lifetime of a substance	$T = \frac{1}{\lambda}$
Activity of a radioactive substance	$A = A_0 e^{-\lambda t}$
Activity of a radioactive substance (linear form)	$\ln A = -\lambda t + \ln A_0$
Alpha decay	${}^A_Z\text{X} \rightarrow {}^{A-4}_{Z-2}\text{X} + {}^4_2\text{He}$
Beta decay	${}^A_Z\text{X} \rightarrow {}^A_{Z+1}\text{X} + {}^0_{-1}\text{e} + \nu$
Positron emission	${}^A_Z\text{X} \rightarrow {}^A_{Z-1}\text{X} + {}^0_{+1}\text{e} + \nu$
Gamma decay	${}^A_Z\text{X}^* \rightarrow {}^A_Z\text{X} + \gamma$

Conceptual Questions

Exercise:

Problem: Why is a PET scan more accurate than a SPECT scan?

Exercise:

Problem:

Isotopes that emit α radiation are relatively safe outside the body and exceptionally hazardous inside. Explain why.

Solution:

Alpha particles do not penetrate materials such as skin and clothes easily. (Recall that alpha radiation is barely able to pass through a thin sheet of paper.) However, when produce inside the body, neighboring cells are vulnerable.

Exercise:

Problem:

Ionizing radiation can impair the ability of a cell to repair DNA. What are the three ways the cell can respond?

Problems

Exercise:

Problem:

What is the dose in mSv for: (a) a 0.1-Gy X-ray? (b) 2.5 mGy of neutron exposure to the eye? (c) 1.5m Gy of α exposure?

Exercise:

Problem:

Find the radiation dose in Gy for: (a) A 10-mSv fluoroscopic X-ray series. (b) 50 mSv of skin exposure by an α emitter. (c) 160 mSv of β^- and γ rays from the ^{40}K in your body.

Solution:

$$\text{Gy} = \frac{\text{Sv}}{\text{RBE}}: \text{a. } 0.01 \text{ Gy; b. } 0.0025 \text{ Gy; c. } 0.16 \text{ Gy}$$

Exercise:

Problem: Find the mass of ^{239}Pu that has an activity of $1.00\ \mu\text{Ci}$.

Exercise:

Problem:

In the 1980s, the term picowave was used to describe food irradiation in order to overcome public resistance by playing on the well-known safety of microwave radiation. Find the energy in MeV of a photon having a wavelength of a picometer.

Solution:

1.24 MeV

Exercise:

Problem:

What is the dose in Sv in a cancer treatment that exposes the patient to 200 Gy of γ rays?

Exercise:

Problem:

One half the γ rays from $^{99\text{m}}\text{Tc}$ are absorbed by a 0.170-mm-thick lead shielding. Half of the γ rays that pass through the first layer of lead are absorbed in a second layer of equal thickness. What thickness of lead will absorb all but one in 1000 of these γ rays?

Solution:

1.69 mm

Exercise:

Problem:

How many Gy of exposure is needed to give a cancerous tumor a dose of 40 Sv if it is exposed to α activity?

Exercise:

Problem:

A plumber at a nuclear power plant receives a whole-body dose of 30 mSv in 15 minutes while repairing a crucial valve. Find the radiation-induced yearly risk of death from cancer and the chance of genetic defect from this maximum allowable exposure.

Solution:

For cancer: $(3 \text{ rem}) \left(\frac{10}{10^6 \text{ rem}\cdot\text{y}} \right) = \frac{30}{10^6 \text{ y}}$, The risk each year of dying from induced cancer is 30 in a million. For genetic defect: $(3 \text{ rem}) \left(\frac{3.3}{10^6 \text{ rem}\cdot\text{y}} \right) = \frac{9.9}{10^6 \text{ y}}$, The chance each year of an induced genetic defect is 10 in a million.

Exercise:

Problem:

Calculate the dose in rem/y for the lungs of a weapons plant employee who inhales and retains an activity of $1.00 \mu\text{Ci } ^{239}\text{Pu}$ in an accident. The mass of affected lung tissue is 2.00 kg and the plutonium decays by emission of a 5.23-MeV α particle. Assume a RBE value of 20.

Additional Problems

Exercise:

Problem:

The wiki-phony site states that the atomic mass of chlorine is 40 g/mol. Check this result. *Hint:* The two, most common stable isotopes of chlorine are: $^{35}_{17}\text{Cl}$ and $^{37}_{17}\text{Cl}$. (The abundance of Cl-35 is 75.8%, and the abundance of Cl-37 is 24.2%.)

Solution:

atomic mass (Cl) = 35.5 g/mol

Exercise:

Problem:

A particle physicist discovers a neutral particle with a mass of 2.02733 u that he assumes is two neutrons bound together.

- (a) Find the binding energy.
- (b) What is unreasonable about this result?

Exercise:

Problem:

A nuclear physicist finds $1.0 \mu\text{g}$ of ^{236}U in a piece of uranium ore ($T_{1/2} = 2.348 \times 10^7 \text{ y}$). (a) Use the decay law to determine how much ^{236}U would had to have been on Earth when it formed $4.543 \times 10^9 \text{ y}$ ago for $1.0 \mu\text{g}$ to be left today. (b) What is unreasonable about this result? (c) How is this unreasonable result resolved?

Solution:

a. 1.71×10^{58} kg; b. This mass is impossibly large; it is greater than the mass of the entire Milky Way galaxy. c. ^{236}U is not produced through natural processes operating over long times on Earth, but through artificial processes in a nuclear reactor.

Exercise:

Problem:

A group of scientists use carbon dating to date a piece of wood to be 3 billion years old. Why doesn't this make sense?

Exercise:

Problem:

According to your lab partner, a 2.00-cm-thick sodium-iodide crystal absorbs all but 10% of rays from a radioactive source and a 4.00-cm piece of the same material absorbs all but 5%. Is this result reasonable?

Solution:

If 10% of rays are left after 2.00 cm, then only $(0.100)^2 = 0.01 = 1\%$ are left after 4.00 cm. This is much smaller than your lab partner's result (5%).

Exercise:

Problem:

In the science section of the newspaper, an article reports the efforts of a group of scientists to create a new nuclear reactor based on the fission of iron (Fe). Is this a good idea?

Exercise:

Problem:

The ceramic glaze on a red-orange "Fiestaware" plate is U_2O_3 and contains 50.0 grams of ^{238}U , but very little ^{235}U . (a) What is the activity of the plate? (b) Calculate the total energy that will be released by the ^{238}U decay. (c) If energy is worth 12.0 cents per $\text{kW} \cdot \text{h}$, what is the monetary value of the energy emitted? (These brightly-colored ceramic plates went out of production some 30 years ago, but are still available as collectibles.)

Solution:

a. 1.68×10^{-5} Ci; (b) From [Appendix B](#), the energy released per decay is 4.27 MeV, so 8.65×10^{10} J; (c) The monetary value of the energy is $\$2.9 \times 10^3$

Exercise:

Problem:

Large amounts of depleted uranium (^{238}U) are available as a by-product of uranium processing for reactor fuel and weapons. Uranium is very dense and makes good counter weights for aircraft. Suppose you have a 4000-kg block of ^{238}U . (a) Find its activity. (b) How many calories per day are generated by thermalization of the decay energy? (c) Do you think you could detect this as heat? Explain.

Exercise:**Problem:**

A piece of wood from an ancient Egyptian tomb is tested for its carbon-14 activity. It is found to have an activity per gram of carbon of $A = 10 \text{ decay/min} \cdot \text{g}$. What is the age of the wood?

Solution:

We know that $\lambda = 3.84 \times 10^{-12} \text{ s}^{-1}$ and
 $A_0 = 0.25 \text{ decays/s} \cdot \text{g} = 15 \text{ decays/min} \cdot \text{g}$.

Thus, the age of the tomb is

$$t = -\frac{1}{3.84 \times 10^{-12} \text{ s}^{-1}} \ln \frac{10 \text{ decays/min} \cdot \text{g}}{15 \text{ decays/min} \cdot \text{g}} = 1.06 \times 10^{11} \text{ s} \approx 3350 \text{ y}.$$

Challenge Problems**Exercise:****Problem:**

This problem demonstrates that the binding energy of the electron in the ground state of a hydrogen atom is much smaller than the rest mass energies of the proton and electron.

(a) Calculate the mass equivalent in u of the 13.6-eV binding energy of an electron in a hydrogen atom, and compare this with the known mass of the hydrogen atom.

(b) Subtract the known mass of the proton from the known mass of the hydrogen atom.

(c) Take the ratio of the binding energy of the electron (13.6 eV) to the energy equivalent of the electron's mass (0.511 MeV).

(d) Discuss how your answers confirm the stated purpose of this problem.

Exercise:

Problem:

The *Galileo* space probe was launched on its long journey past Venus and Earth in 1989, with an ultimate goal of Jupiter. Its power source is 11.0 kg of ^{238}Pu , a by-product of nuclear weapons plutonium production. Electrical energy is generated thermoelectrically from the heat produced when the 5.59-MeV α particles emitted in each decay crash to a halt inside the plutonium and its shielding. The half-life of ^{238}Pu is 87.7 years.

- (a) What was the original activity of the ^{238}Pu in becquerels?
 - (b) What power was emitted in kilowatts?
 - (c) What power was emitted 12.0 y after launch? You may neglect any extra energy from daughter nuclides and any losses from escaping γ rays.
-

Solution:

a. 6.97×10^{15} Bq; b. 6.24 kW; c. 5.67 kW

Exercise:

Problem: Find the energy emitted in the β^- decay of ^{60}Co .

Exercise:**Problem:**

Engineers are frequently called on to inspect and, if necessary, repair equipment in nuclear power plants. Suppose that the city lights go out. After inspecting the nuclear reactor, you find a leaky pipe that leads from the steam generator to turbine chamber. (a) How do the pressure readings for the turbine chamber and steam condenser compare? (b) Why is the nuclear reactor *not* generating electricity?

Solution:

- a. Due to the leak, the pressure in the turbine chamber has dropped significantly. The pressure difference between the turbine chamber and steam condenser is now very low.
- b. A large pressure difference is required for steam to pass through the turbine chamber and turn the turbine.

Exercise:

Problem:

If two nuclei are to fuse in a nuclear reaction, they must be moving fast enough so that the repulsive Coulomb force between them does not prevent them from getting within $R \approx 10^{-14}\text{m}$ of one another. At this distance or nearer, the attractive nuclear force can overcome the Coulomb force, and the nuclei are able to fuse.

(a) Find a simple formula that can be used to estimate the minimum kinetic energy the nuclei must have if they are to fuse. To keep the calculation simple, assume the two nuclei are identical and moving toward one another with the same speed v . (b) Use this minimum kinetic energy to estimate the minimum temperature a gas of the nuclei must have before a significant number of them will undergo fusion. Calculate this minimum temperature first for hydrogen and then for helium. (*Hint:* For fusion to occur, the minimum kinetic energy when the nuclei are far apart must be equal to the Coulomb potential energy when they are a distance R apart.)

Exercise:**Problem:**

For the reaction, $n + {}^3\text{He} \rightarrow {}^4\text{He} + \gamma$, find the amount of energy transferred to ${}^4\text{He}$ and γ (on the right side of the equation). Assume the reactants are initially at rest. (*Hint:* Use conservation of momentum principle.)

Solution:

The energies are

$$E_{\gamma} = 20.6 \text{ MeV}$$

$E_{{}^4\text{He}} = 5.68 \times 10^{-2} \text{ MeV}$. Notice that most of the energy goes to the γ ray.

Exercise:**Problem:**

Engineers are frequently called on to inspect and, if necessary, repair equipment in medical hospitals. Suppose that the PET system malfunctions. After inspecting the unit, you suspect that one of the PET photon detectors is misaligned. To test your theory you position one detector at the location $(r, \theta, \varphi) = (1.5, 45, 30)$ relative to a radioactive test sample at the center of the patient bed. (a) If the second photon detector is properly aligned where should it be located? (b) What energy reading is expected?

Glossary

high dose

dose of radiation greater than 1 Sv (100 rem)

low dose

dose of radiation less than 100 mSv (10 rem)

moderate dose

dose of radiation from 0.1 Sv to 1 Sv (10 to 100 rem)

positron emission tomography (PET)

tomography technique that uses β^+ emitters and detects the two annihilation γ rays, aiding in source localization

radiation dose unit (rad)

ionizing energy deposited per kilogram of tissue

radioactive tags

special drugs (radiopharmaceuticals) that allow doctors to track movement of other drugs in the body

radiopharmaceutical

compound used for medical imaging

relative biological effectiveness (RBE)

number that expresses the relative amount of damage that a fixed amount of ionizing radiation of a given type can inflict on biological tissues

roentgen equivalent man (rem)

dose unit more closely related to effects in biological tissue

sievert (Sv)

SI equivalent of the rem

single-photon-emission computed tomography (SPECT)

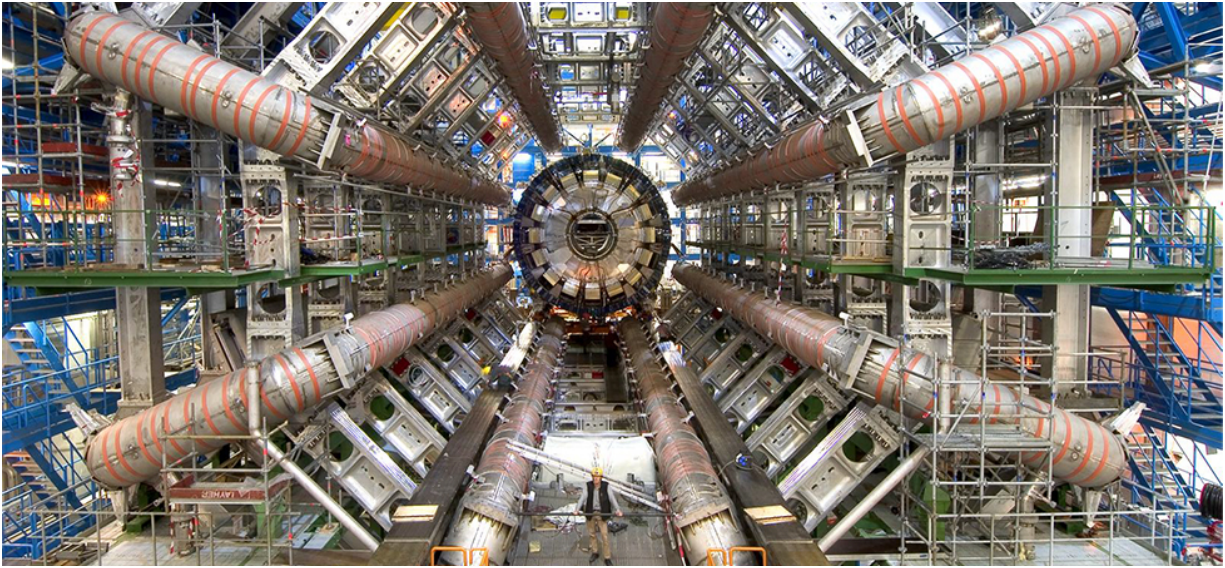
tomography performed with γ -emitting radiopharmaceuticals

Introduction

class="introduction"

The Large Hadron Collider (LHC) is located over 150 meters (500 feet) underground on the border of Switzerland and France near Geneva, Switzerland. The LHC is the most powerful machine ever developed to test our understanding of elementary particle interactions. Shown here is the ATLAS detector, which helps identify new particles formed in collisions. (credit:

modification
of work by
Maximilien
Brice,
(CERN)



At the very beginning of this text we discussed the wide range of scales that physics encompasses, from the very smallest particles to the largest scale possible—the universe itself. In this final chapter we examine some of the frontiers of research at these extreme scales. Particle physics deals with the most basic building blocks of matter and the forces that hold them together. Cosmology is the study of the stars, galaxies, and galactic structures that populate our universe, as well as their past history and future evolution.

These two areas of physics are not as disconnected as you might think. The study of elementary particles requires enormous energies to produce isolated particles, involving some of the largest machines humans have ever built. But such high energies were present in the earliest stages of the universe and the universe we see around us today was shaped in part by the nature and interactions of the elementary particles created then. Bear in mind that particle physics and cosmology are both areas of intense current research, subject to much speculation on the part of physicists (as well as science-fiction writers). In this chapter we try to emphasize what is known

on the basis of deductions from observational evidence, and identify ideas that are conjectured but still unproven.

Introduction to Particle Physics

By the end of this section, you will be able to:

- Describe the four fundamental forces and what particles participate in them
- Identify and describe fermions and bosons
- Identify and describe the quark and lepton families
- Distinguish between particles and antiparticles, and describe their interactions

Elementary particle physics is the study of fundamental particles and their interactions in nature. Those who study elementary particle physics—the particle physicists—differ from other physicists in the scale of the systems that they study. A particle physicist is not content to study the microscopic world of cells, molecules, atoms, or even atomic nuclei. They are interested in physical processes that occur at scales even smaller than atomic nuclei. At the same time, they engage the most profound mysteries in nature: How did the universe begin? What explains the pattern of masses in the universe? Why is there more matter than antimatter in the universe? Why are energy and momentum conserved? How will the universe evolve?

Four Fundamental Forces

An important step to answering these questions is to understand particles and their interactions. Particle interactions are expressed in terms of four **fundamental forces**. In order of decreasing strength, these forces are the **strong nuclear force**, the electromagnetic force, the **weak nuclear force**, and the gravitational force.

1. **Strong nuclear force.** The strong nuclear force is a very strong attractive force that acts only over very short distances (about 10^{-15} m). The strong nuclear force is responsible for binding protons and neutrons together in atomic nuclei. Not all particles participate in the strong nuclear force; for instance, electrons and neutrinos are not affected by it. As the name suggests, this force is much stronger than the other forces.
2. **Electromagnetic force.** The electromagnetic force can act over very large distances (it has an infinite range) but is only 1/100 the strength of the strong nuclear force. Particles that interact through this force are said to have “charge.” In the classical theory of static electricity (Coulomb’s law), the electric force varies as the product of the charges of the interacting particles, and as the inverse square of the distances between them. In contrast to the strong force, the electromagnetic force can be attractive or repulsive (opposite charges attract and like charges repel). The magnetic force depends in a more complicated way on the charges and their motions. The unification of the electric and magnetic force into a single electromagnetic force (an achievement of James Clerk Maxwell) stands as one of the greatest intellectual achievements of the nineteenth century. This force is central to scientific models of atomic structure and molecular bonding.
3. **Weak nuclear force.** The weak nuclear force acts over very short distances (10^{-15} m) and, as its name suggest, is very weak. It is roughly 10^{-6} the strength of the strong nuclear force. This force is manifested most notably in decays of elementary particles and neutrino interactions. For example, the neutron can decay to a proton, electron, and electron neutrino through the weak force. The weak force is vitally important because it is essential

for understanding stellar nucleosynthesis—the process that creates new atomic nuclei in the cores of stars.

4. **Gravitational force.** Like the electromagnetic force, the gravitational force can act over infinitely large distances; however, it is only 10^{-38} as strong as the strong nuclear force. In Newton’s classical theory of gravity, the force of gravity varies as the product of the masses of the interacting particles and as the inverse square of the distance between them. This force is an attractive force that acts between all particles with mass. In modern theories of gravity, this force behavior is considered a special case for low-energy macroscopic interactions. Compared with the other forces of nature, gravity is by far the weakest.

The fundamental forces may not be truly “fundamental” but may actually be different aspects of the same force. Just as the electric and magnetic forces were unified into an electromagnetic force, physicists in the 1970s unified the electromagnetic force with the weak nuclear force into an **electroweak force**. Any scientific theory that attempts to unify the electroweak force and strong nuclear force is called a **grand unified theory**, and any theory that attempts to unify all four forces is called a **theory of everything**. We will return to the concept of unification later in this chapter.

Classifications of Elementary Particles

A large number of subatomic particles exist in nature. These particles can be classified in two ways: the property of spin and participation in the four fundamental forces. Recall that the spin of a particle is analogous to the rotation of a macroscopic object about its own axis. These types of classification are described separately below.

Classification by spin

Particles of matter can be divided into **fermions** and **bosons**. Fermions have half-integral spin ($\frac{1}{2}\hbar, \frac{3}{2}\hbar, \dots$) and bosons have integral spin ($0\hbar, 1\hbar, 2\hbar, \dots$). Familiar examples of fermions are electrons, protons, and neutrons. A familiar example of a boson is a photon. Fermions and bosons behave very differently in groups. For example, when electrons are confined to a small region of space, Pauli’s exclusion principle states that no two electrons can occupy the same quantum-mechanical state. However, when photons are confined to a small region of space, there is no such limitation.

The behavior of fermions and bosons in groups can be understood in terms of the property of indistinguishability. Particles are said to be “indistinguishable” if they are identical to one another. For example, electrons are indistinguishable because every electron in the universe has exactly the same mass and spin as all other electrons—“when you’ve seen one electron, you’ve seen them all.” If you switch two indistinguishable particles in the same small region of space, the square of the wave function that describes this system and can be measured ($|\psi|^2$) is unchanged. If this were not the case, we could tell whether or not the particles had been switched and the particle would not be truly indistinguishable. Fermions and bosons differ by whether the sign of the wave function (ψ)—not directly observable—flips:

Equation:

$$\begin{aligned}\psi &\rightarrow -\psi \text{ (indistinguishable fermions),} \\ \psi &\rightarrow +\psi \text{ (indistinguishable bosons).}\end{aligned}$$

Fermions are said to be “antisymmetric on exchange” and bosons are “symmetric on exchange.” Pauli’s exclusion principle is a consequence of **exchange symmetry** of fermions—a connection developed in a more advanced course in modern physics. The electronic structure of atoms is predicated on Pauli’s exclusion principle and is therefore directly related to the indistinguishability of electrons.

Classification by force interactions

Fermions can be further divided into **quarks** and **leptons**. The primary difference between these two types of particles is that quarks interact via the strong force and leptons do not. Quarks and leptons (as well as bosons to be discussed later) are organized in [\[link\]](#). The upper two rows (first three columns in purple) contain six quarks. These quarks are arranged into two particle families: up, charm, and top (u, c, t), and down, strange, and bottom (d, s, b). Members of the same particle family share the same properties but differ in mass (given in MeV/c^2). For example, the mass of the top quark is much greater than the charm quark, and the mass of the charm quark is much greater than the up quark. All quarks interact with one another through the strong nuclear force.

The families of subatomic particles, categorized by the types of forces with which they interact. (credit: modification of work by “MissMJ”/Wikimedia Commons)

The lower two rows in the figure (in green) contain six leptons arranged into two particle families: electron, muon, and tau (e, μ, τ), and electron neutrino, muon neutrino, and tau neutrino (ν_e, ν_μ, ν_τ). The muon is over 200 times heavier than an electron, but is otherwise similar to the electron. The tau is about 3500 times heavier than the electron, but is otherwise similar to the muon and electron. Once created, the muon and tau quickly decay to lighter particles via the weak force. Leptons do not participate in the strong force. Quarks and leptons

will be discussed later in this chapter. Leptons participate in the weak, electromagnetic, and gravitational forces, but do not participate in the strong force.

Bosons (shown in red) are the force carriers of the fermions. In this model, leptons and quarks interact with each other by sending and receiving bosons. For example, Coulombic interaction occurs when two positively charged particles send and receive (exchange) photons. The photons are said to “carry” the force between charged particles. Likewise, attraction between two quarks in an atomic nucleus occurs when two quarks send and receive **gluons**. Additional examples include **W and Z bosons** (which carry weak nuclear force) and gravitons (which carry gravitational force). The Higgs boson is a special particle: When it interacts with other particles, it endows them not with force but with mass. In other words, the Higgs boson helps to explain *why* particles have mass. These assertions are part of a tentative but very productive scientific model (the Standard Model) discussed later.

Particles and Antiparticles

In the late 1920s, the special theory of relativity and quantum mechanics were combined into a relativistic quantum theory of the electron. A surprising result of this theory was the prediction of two energy states for each electron: One is associated with the electron, and the other is associated with another particle with the same mass of an electron but with a charge of e^+ . This particle is called the antielectron or **positron**. The positron was discovered experimentally in the 1930s.

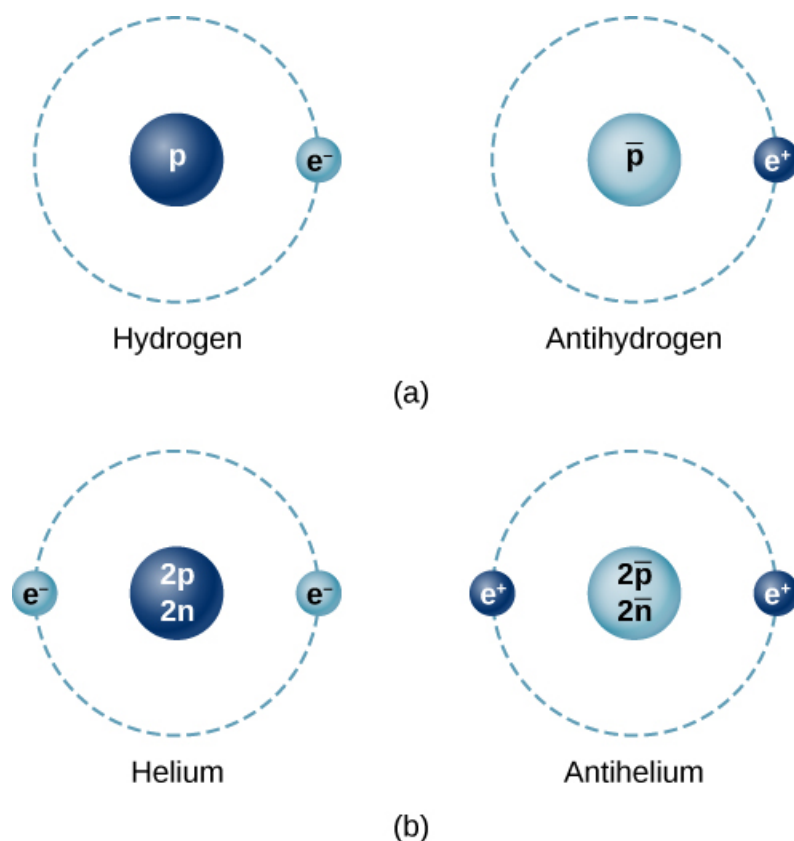
Soon it was discovered that for every particle in nature, there is a corresponding **antiparticle**. An antiparticle has the same mass and lifetime as its associated particle, and the opposite sign of electric charge. These particles are produced in high-energy reactions. Examples of high-energy particles include the antimuon (μ^+), anti-up quark (\bar{u}), and anti-down quark (\bar{d}). (Note that antiparticles for quarks are designated with an over-bar.) Many mesons and baryons contain antiparticles. For example, the antiproton (\bar{p}) is $\bar{u}\bar{u}\bar{d}$ and the positively charged pion (π^+) is $u\bar{d}$. Some neutral particles, such as the photon and the π^0 meson, are their own antiparticles. Sample particles, antiparticles, and their properties are listed in [\[link\]](#).

	Particle name	Symbol	Antiparticle	Mass (MeV/ c^2)	Average lifetime (s)
Leptons					
	Electron	e^-	e^+	0.511	Stable

	Particle name	Symbol	Antiparticle	Mass (MeV/ c^2)	Average lifetime (s)
	Electron neutrino	ν_e	$\bar{\nu}_e$	≈ 0	Stable
	Muon	μ^-	μ^+	105.7	2.20×10^{-6}
	Muon neutrino	ν_μ	$\bar{\nu}_\mu$	≈ 0	Stable
	Tau	τ^-	τ^+	1784	$< 4 \times 10^{-13}$
	Tau neutrino	ν_τ	$\bar{\nu}_\tau$	≈ 0	Stable
Hadrons					
Baryons	Proton	p	\bar{p}	938.3	Stable
	Neutron	n	\bar{n}	939.6	920
	Lambda	Λ^0	$\bar{\Lambda}^0$	1115.6	2.6×10^{-10}
	Sigma	Σ^+	Σ^-	1189.4	0.80×10^{-10}
	Xi	Ξ^+	Ξ^-	1315	2.9×10^{-10}
	Omega	Ω^+	Ω^-	1672	0.82×10^{-10}
Mesons	Pion	π^+	π^-	139.6	2.60×10^{-8}
	π -Zero	π^0	π^0	135.0	0.83×10^{-16}
	Kaon	K^+	K^-	493.7	1.24×10^{-8}
	k-Short	K_S^0	\bar{K}_S^0	497.6	0.89×10^{-10}
	k-Long	K_L^0	\bar{K}_L^0	497.6	5.2×10^{-8}
	J/ ψ	J/ ψ	J/ ψ	3100	7.1×10^{-21}
	Upsilon	Υ	Υ	9460	1.2×10^{-20}

Particles and their Properties

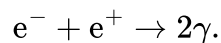
The same forces that hold ordinary matter together also hold antimatter together. Under the right conditions, it is possible to create antiatoms such as antihydrogen, antioxygen, and even antiwater. In antiatoms, positrons orbit a negatively charged nucleus of antiprotons and antineutrons. [\[link\]](#) compares atoms and antiatoms.



A comparison of the simplest atoms of matter and antimatter. (a) In the Bohr model, an antihydrogen atom consists of a positron that orbits an antiproton. (b) An antihelium atom consists of two positrons that orbit a nucleus of two antiprotons and two antineutrons.

Antimatter cannot exist for long in nature because particles and antiparticles annihilate each other to produce high-energy radiation. A common example is electron-positron annihilation. This process proceeds by the reaction

Equation:



The electron and positron vanish completely and two photons are produced in their place. (It turns out that the production of a single photon would violate conservation of energy and momentum.) This reaction can also proceed in the reverse direction: Two photons can annihilate each other to produce an electron and positron pair. Or, a single photon can produce an electron-positron pair in the field of a nucleus, a process called pair production. Reactions of this kind are measured routinely in modern particle detectors. The existence of antiparticles in nature is not science fiction.

Note:

Watch this [video](#) to learn more about matter and antimatter particles.

Summary

- The four fundamental forces of nature are, in order of strength: strong nuclear, electromagnetic, weak nuclear, and gravitational. Quarks interact via the strong force, but leptons do not. Both quark and leptons interact via the electromagnetic, weak, and gravitational forces.
- Elementary particles are classified into fermions and boson. Fermions have half-integral spin and obey the exclusion principle. Bosons have integral spin and do not obey this principle. Bosons are the force carriers of particle interactions.
- Quarks and leptons belong to particle families composed of three members each. Members of a family share many properties (charge, spin, participation in forces) but not mass.
- All particles have antiparticles. Particles share the same properties as their antimatter particles, but carry opposite charge.

Conceptual Questions

Exercise:

Problem: What are the four fundamental forces? Briefly describe them.

Solution:

Strong nuclear force: interaction between quarks, mediated by gluons. Electromagnetic force: interaction between charge particles, mediated photons. Weak nuclear force: interactions between fermions, mediated by heavy bosons. Gravitational force: interactions between material (massive) particle, mediate by hypothetical gravitons.

Exercise:**Problem:**

Distinguish fermions and bosons using the concepts of indistinguishability and exchange symmetry.

Exercise:

Problem: List the quark and lepton families

Solution:

electron, muon, tau; electron neutrino, muon neutrino, tau neutrino; down quark, strange quark, bottom quark; up quark, charm quark, top quark

Exercise:

Problem:

Distinguish between elementary particles and antiparticles. Describe their interactions.

Problems**Exercise:**

Problem:

How much energy is released when an electron and a positron at rest annihilate each other? (For particle masses, see [\[link\]](#).)

Solution:

1.022 MeV

Exercise:

Problem:

If 1.0×10^{30} MeV of energy is released in the annihilation of a sphere of matter and antimatter, and the spheres are equal mass, what are the masses of the spheres?

Exercise:

Problem:

When both an electron and a positron are at rest, they can annihilate each other according to the reaction

$$e^- + e^+ \rightarrow \gamma + \gamma.$$

In this case, what are the energy, momentum, and frequency of each photon?

Solution:

0.511 MeV, 2.73×10^{-22} kg · m/s, 1.23×10^{20} Hz

Exercise:

Problem:

What is the *total kinetic energy* carried away by the particles of the following decays?

(a) $\pi^0 \rightarrow \gamma + \gamma$

(b) $K^0 \rightarrow \pi^+ + \pi^-$

(c) $\Sigma^+ \rightarrow n + \pi^+$

(d) $\Sigma^0 \rightarrow \Lambda^0 + \gamma$.

Glossary

antiparticle

subatomic particle with the same mass and lifetime as its associated particle, but opposite electric charge

baryons

group of three quarks

boson

particle with integral spin that are symmetric on exchange

electroweak force

unification of electromagnetic force and weak-nuclear force interactions

exchange symmetry

property of a system of indistinguishable particles that requires the exchange of any two particles to be unobservable

fermion

particle with half-integral spin that is antisymmetric on exchange

fundamental force

one of four forces that act between bodies of matter: the strong nuclear, electromagnetic, weak nuclear, and gravitational forces

gluon

particle that carry the strong nuclear force between quarks within an atomic nucleus

grand unified theory

theory of particle interactions that unifies the strong nuclear, electromagnetic, and weak nuclear forces

hadron

a meson or baryon

lepton

a fermion that participates in the electroweak force

mesons

a group of two quarks

quark

a fermion that participates in the electroweak and strong nuclear force

positron

antielectron

strong nuclear force

relatively strong attractive force that acts over short distances (about 10^{-15} m) responsible for binding protons and neutrons together in atomic nuclei

theory of everything

a theory of particle interactions that unifies all four fundamental forces

W and Z boson

particle with a relatively large mass that carries the weak nuclear force between leptons and quarks

weak nuclear force

relative weak force (about 10^{-6} the strength of the strong nuclear force) responsible for decays of elementary particles and neutrino interactions

Particle Conservation Laws

By the end of this section, you will be able to:

- Distinguish three conservation laws: baryon number, lepton number, and strangeness
- Use rules to determine the total baryon number, lepton number, and strangeness of particles before and after a reaction
- Use baryon number, lepton number, and strangeness conservation to determine if particle reactions or decays occur

Conservation laws are critical to an understanding of particle physics. Strong evidence exists that energy, momentum, and angular momentum are all conserved in all particle interactions. The annihilation of an electron and positron at rest, for example, cannot produce just one photon because this violates the conservation of linear momentum. As discussed in [Relativity](#), the special theory of relativity modifies definitions of momentum, energy, and other familiar quantities. In particular, the relativistic momentum of a particle differs from its classical momentum by a factor $\gamma = 1/\sqrt{1 - (v/c)^2}$ that varies from 1 to ∞ , depending on the speed of the particle.

In previous chapters, we encountered other conservation laws as well. For example, charge is conserved in all electrostatic phenomena. Charge lost in one place is gained in another because charge is carried by particles. No known physical processes violate charge conservation. In the next section, we describe three less-familiar conservation laws: baryon number, lepton number, and strangeness. These are by no means the only conservation laws in particle physics.

Baryon Number Conservation

No conservation law considered thus far prevents a neutron from decaying via a reaction such as **Equation:**

$$n \rightarrow e^+ + e^-.$$

This process conserves charge, energy, and momentum. However, it does not occur because it violates the law of baryon number conservation. This law requires that the total baryon number of a reaction is the same before and after the reaction occurs. To determine the total baryon number, every elementary particle is assigned a **baryon number** B . The baryon number has the value $B = +1$ for baryons, -1 for antibaryons, and 0 for all other particles. Returning to the above case (the decay of the neutron into an electron-positron pair), the neutron has a value $B = +1$, whereas the electron and the positron each has a value of 0. Thus, the decay does not occur because the total baryon number changes from 1 to 0. However, the proton-antiproton collision process

Equation:

$$p + \bar{p} \rightarrow p + p + \bar{p} + \bar{p},$$

does satisfy the law of conservation of baryon number because the baryon number is zero before and after the interaction. The baryon number for several common particles is given in [\[link\]](#).

Particle name	Symbol	Lepton number (L_e)	Lepton number (L_μ)	Lepton number (L_τ)	Baryon number (B)	Strange-ness number
Electron	e^-	1	0	0	0	0
Electron neutrino	ν_e	1	0	0	0	0
Muon	μ^-	0	1	0	0	0
Muon neutrino	ν_μ	0	1	0	0	0
Tau	τ^-	0	0	1	0	0
Tau neutrino	ν_τ	0	0	1	0	0
Pion	π^+	0	0	0	0	0
Positive kaon	K^+	0	0	0	0	1
Negative kaon	K^-	0	0	0	0	-1
Proton	p	0	0	0	1	0
Neutron	n	0	0	0	1	0
Lambda zero	Λ^0	0	0	0	1	-1
Positive sigma	Σ^+	0	0	0	1	-1
Negative sigma	Σ^-	0	0	0	1	-1

Particle name	Symbol	Lepton number (L_e)	Lepton number (L_μ)	Lepton number (L_τ)	Baryon number (B)	Strange-ness number
Xi zero	Ξ^0	0	0	0	1	-2
Negative xi	Ξ^-	0	0	0	1	-2
Omega	Ω^-	0	0	0	1	-3

Conserved Properties of Particles

Example:

Baryon Number Conservation

Based on the law of conservation of baryon number, which of the following reactions can occur?

Equation:

$$(a) \pi^- + p \rightarrow \pi^0 + n + \pi^- + \pi^+$$

$$(b) p + \bar{p} \rightarrow p + p + \bar{p}$$

Strategy

Determine the total baryon number for the reactants and products, and require that this value does not change in the reaction. **Solution**

For reaction (a), the net baryon number of the two reactants is $0 + 1 = 1$ and the net baryon number of the four products is $0 + 1 + 0 + 0 = 1$. Since the net baryon numbers of the reactants and products are equal, this reaction is allowed on the basis of the baryon number conservation law.

For reaction (b), the net baryon number of the reactants is $1 + (-1) = 0$ and the net baryon number of the proposed products is $1 + 1 + (-1) = 1$. Since the net baryon numbers of the reactants and proposed products are not equal, this reaction cannot occur.

Significance

Baryon number is conserved in the first reaction, but not in the second. Baryon number conservation constrains what reactions can and cannot occur in nature.

Note:

Exercise:

Problem: Check Your Understanding What is the baryon number of a hydrogen nucleus?

Solution:

Lepton Number Conservation

Lepton number conservation states that the sum of lepton numbers before and after the interaction must be the same. There are three different **lepton numbers**: the electron-lepton number L_e , the muon-lepton number L_μ , and the tau-lepton number L_τ . In any interaction, each of these quantities must be conserved *separately*. For electrons and electron neutrinos, $L_e = 1$; for their antiparticles, $L_e = -1$; all other particles have $L_e = 0$. Similarly, $L_\mu = 1$ for muons and muon neutrinos, $L_\mu = -1$ for their antiparticles, and $L_\mu = 0$ for all other particles. Finally, $L_\tau = 1, -1$, or 0 , depending on whether we have a tau or tau neutrino, their antiparticles, or any other particle, respectively. Lepton number conservation guarantees that the number of electrons and positrons in the universe stays relatively constant. (*Note*: The total lepton number is, as far as we know, conserved in nature. However, observations have shown variations of family lepton number (for example, L_e) in a phenomenon called *neutrino oscillations*.)

To illustrate the lepton number conservation law, consider the following known two-step decay process:

Equation:

$$\begin{aligned}\pi^+ &\rightarrow \mu^+ + \nu_\mu \\ \mu^+ &\rightarrow e^+ + \nu_e + \bar{\nu}_\mu.\end{aligned}$$

In the first decay, all of the lepton numbers for π^+ are 0. For the products of this decay, $L_\mu = -1$ for μ^+ and $L_\mu = 1$ for ν_μ . Therefore, muon-lepton number is conserved. Neither electrons nor tau are involved in this decay, so $L_e = 0$ and $L_\tau = 0$ for the initial particle and all decay products. Thus, electron-lepton and tau-lepton numbers are also conserved. In the second decay, μ^+ has a muon-lepton number $L_\mu = -1$, whereas the net muon-lepton number of the decay products is $0 + 0 + (-1) = -1$. Thus, the muon-lepton number is conserved. Electron-lepton number is also conserved, as $L_e = 0$ for μ^+ , whereas the net electron-lepton number of the decay products is $(-1) + 1 + 0 = 0$. Finally, since no taus or tau-neutrinos are involved in this decay, the tau-lepton number is also conserved.

Example:

Lepton Number Conservation

Based on the law of conservation of lepton number, which of the following decays can occur?

Equation:

$$\begin{aligned}\text{(a)} \quad n &\rightarrow p + e^- + \bar{\nu}_e \\ \text{(b)} \quad \pi^- &\rightarrow \mu^- + \nu_\mu + \bar{\nu}_\mu\end{aligned}$$

Strategy

Determine the total lepton number for the reactants and products, and require that this value does not change in the reaction.

Solution

For decay (a), the electron-lepton number of the neutron is 0, and the net electron-lepton number of the decay products is $0 + 1 + (-1) = 0$. Since the net electron-lepton numbers before and after the decay are the same, the decay is possible on the basis of the law of conservation of electron-lepton number. Also, since there are no muons or taus involved in this decay, the muon-lepton and tauon-lepton numbers are conserved.

For decay (b), the muon-lepton number of the π^- is 0, and the net muon-lepton number of the proposed decay products is $1 + 1 + (-1) = 1$. Thus, on the basis of the law of conservation of muon-lepton number, this decay cannot occur.

Significance

Lepton number is conserved in the first reaction, but not in the second. Lepton number conservation constrains what reactions can and cannot occur in nature.

Note:**Exercise:****Problem:**

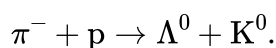
Check Your Understanding What is the lepton number of an electron-positron pair?

Solution:

0

Strangeness Conservation

In the late 1940s and early 1950s, cosmic-ray experiments revealed the existence of particles that had never been observed on Earth. These particles were produced in collisions of pions with protons or neutrons in the atmosphere. Their production and decay were unusual. They were produced in the strong nuclear interactions of pions and nucleons, and were therefore inferred to be hadrons; however, their decay was mediated by the much more slowly acting weak nuclear interaction. Their lifetimes were on the order of 10^{-10} to 10^{-8} s, whereas a typical lifetime for a particle that decays via the strong nuclear reaction is 10^{-23} s. These particles were also unusual because they were always produced in pairs in the pion-nucleon collisions. For these reasons, these newly discovered particles were described as *strange*. The production and subsequent decay of a pair of strange particles is illustrated in [\[link\]](#) and follows the reaction

Equation:

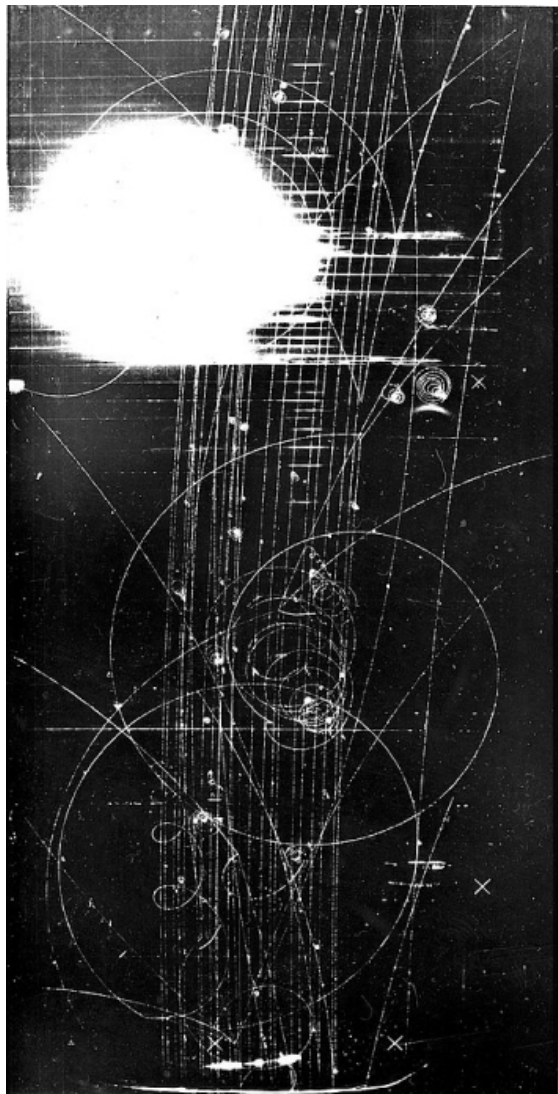
The lambda particle then decays through the weak nuclear interaction according to
Equation:

$$\Lambda^0 \rightarrow \pi^- + p,$$

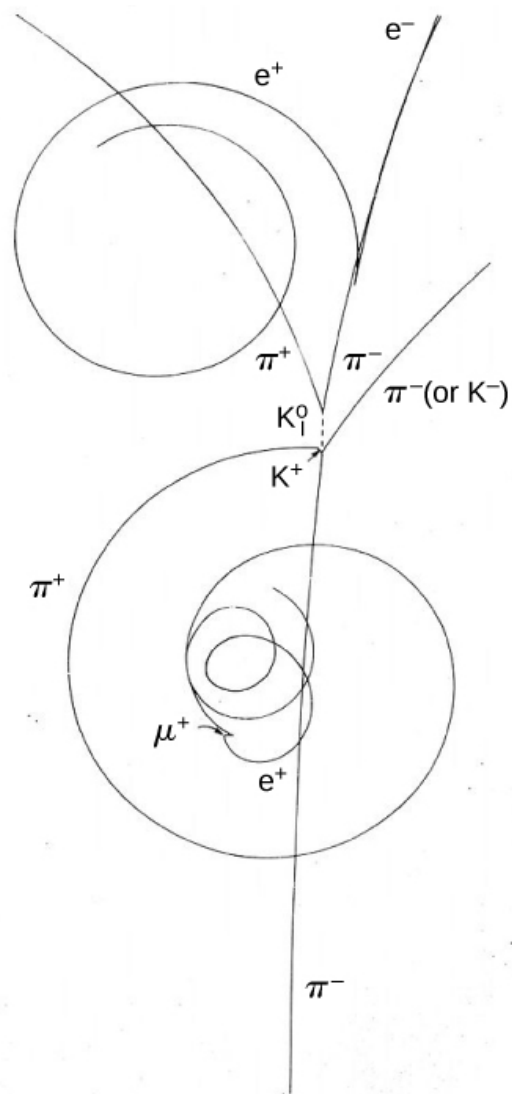
and the kaon decays via the weak interaction

Equation:

$$K^0 \rightarrow \pi^+ + \pi^-.$$



(a)



(b)

The interactions of hadrons. (a) Bubble chamber photograph; (b) sketch that

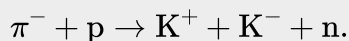
represents the photograph.

To rationalize the behavior of these strange particles, particle physicists invented a particle property conserved in strong interactions but not in weak interactions. This property is called **strangeness** and, as the name suggests, is associated with the presence of a strange quark. The strangeness of a particle is equal to the number of strange quarks of the particle. Strangeness conservation requires the total strangeness of a reaction or decay (summing the strangeness of all the particles) is the same before and after the interaction. Strangeness conservation is not absolute: It is conserved in strong interactions and electromagnetic interactions but not in weak interactions. The strangeness number for several common particles is given in [\[link\]](#).

Example:**Strangeness Conservation**

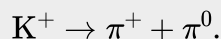
(a) Based on the conservation of strangeness, can the following reaction occur?

Equation:



(b) The following decay is mediated by the weak nuclear force:

Equation:



Does the decay conserve strangeness? If not, can the decay occur?

Strategy

Determine the strangeness of the reactants and products and require that this value does not change in the reaction.

Solution

- The net strangeness of the reactants is $0 + 0 = 0$, and the net strangeness of the products is $1 + (-1) + 0 = 0$. Thus, the strong nuclear interaction between a pion and a proton is not forbidden by the law of conservation of strangeness. Notice that baryon number is also conserved in the reaction.
- The net strangeness before and after this decay is 1 and 0, so the decay does not conserve strangeness. However, the decay may still be possible, because the law of conservation of strangeness does not apply to weak decays.

Significance

Strangeness is conserved in the first reaction, but not in the second. Strangeness conservation constrains what reactions can and cannot occur in nature.

Note:

Exercise:

Problem: Check Your Understanding What is the strangeness number of a muon?

Solution:

0

Summary

- Elementary particle interactions are governed by particle conservation laws, which can be used to determine what particle reactions and decays are possible (or forbidden).
- The baryon number conservation law and the three lepton number conservation law are valid for all physical processes. However, conservation of strangeness is valid only for strong nuclear interactions and electromagnetic interactions.

Conceptual Questions

Exercise:

Problem: What are six particle conservation laws? Briefly describe them.

Solution:

Conservation energy, momentum, and charge (familiar to classical and relativistic mechanics). Also, conservation of baryon number, lepton number, and strangeness—numbers that do not change before and after a collision or decay.

Exercise:

Problem: In general, how do we determine if a particle reaction or decay occurs?

Exercise:**Problem:**

Why might the detection of particle interaction that violates an established particle conservation law be considered a *good* thing for a scientist?

Solution:

It means that the theory that requires the conservation law is not understood. The failure of a long-established theory often leads to a deeper understanding of nature.

Problems

Exercise:**Problem:**

Which of the following decays cannot occur because the law of conservation of lepton number is violated?

- | | |
|---|---|
| (a) $n \rightarrow p + e^-$ | (e) $\pi^- \rightarrow e^- + \bar{\nu}_e$ |
| (b) $\mu^+ \rightarrow e^+ + \nu_e$ | (f) $\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$ |
| (c) $\pi^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$ | (g) $\Lambda^0 \rightarrow \pi^- + p$ |
| (d) $p \rightarrow n + e^+ + \nu_e$ | (h) $K^+ \rightarrow \mu^+ + \nu_\mu$ |

Solution:

a, b, and c

Exercise:**Problem:**

Which of the following reactions cannot because the law of conservation of strangeness is violated?

- | | |
|--|---|
| (a) $p + n \rightarrow p + p + \pi^-$ | (e) $K^- + p \rightarrow \Xi^0 + K^+ + \pi^-$ |
| (b) $p + n \rightarrow p + p + K^-$ | (f) $K^- + p \rightarrow \Xi^0 + \pi^- + \pi^-$ |
| (c) $K^- + p \rightarrow K^- + \Sigma^+$ | (g) $\pi^+ + p \rightarrow \Sigma^+ + K^+$ |
| (d) $\pi^- + p \rightarrow K^+ + \Sigma^-$ | (h) $\pi^- + n \rightarrow K^- + \Lambda^0$ |

Exercise:

Problem: Identify one possible decay for each of the following antiparticles:

- (a) \bar{n} , (b) $\bar{\Lambda}^0$, (c) Ω^+ , (d) K^- , and (e) $\bar{\Sigma}^-$.

Solution:

- a. $\bar{p}_e^+ \nu_e$; b. $\bar{p}\pi^+$ or $\bar{p}\pi^0$; c. $\bar{\Xi}^0\pi^0$ or $\bar{\Lambda}^0 K^+$; d. $\mu^-\bar{\nu}_\mu$ or $\pi^-\pi^0$; e. $\bar{p}\pi^0$ or $\bar{n}\pi^-$

Exercise:**Problem:**

Each of the following strong nuclear reactions is forbidden. Identify a conservation law that is violated for each one.

(a) $p + \bar{p} \rightarrow p + n + \bar{p}$

(b) $p + n \rightarrow p + \bar{p} + n + \pi^+$

(c) $\pi^- + p \rightarrow \Sigma^+ + K^-$

(d) $K^- + p \rightarrow \Lambda^0 + n$

Glossary

baryon number

baryon number has the value $B = +1$ for baryons, -1 for antibaryons, and 0 for all other particles and is conserved in particle interactions

lepton number

electron-lepton number L_e , the muon-lepton number L_μ , and the tau-lepton number L_τ are conserved separately in every particle interaction

strangeness

particle property associated with the presence of a strange quark

Quarks

By the end of this section, you will be able to:

- Compare and contrast the six known quarks
- Use quark composition of hadrons to determine the total charge of these particles
- Explain the primary evidence for the existence of quarks

In the 1960s, particle physicists began to realize that hadrons are not elementary particles but are made of particles called *quarks*. (The name ‘quark’ was coined by the physicist Murray Gell-Mann, from a phrase in the James Joyce novel *Finnegans Wake*.) Initially, it was believed there were only three types of quarks, called *up* (*u*), *down* (*d*), and *strange* (*s*). However, this number soon grew to six—interestingly, the same as the number of leptons—to include *charmed* (*c*), *bottom* (*b*), and *top* (*t*).

All quarks are spin-half fermions ($s = 1/2$), have a fractional charge ($1/3$ or $2/3e$), and have baryon number $B = 1/3$. Each quark has an antiquark with the same mass but opposite charge and baryon number. The names and properties of the six quarks are listed in [\[link\]](#).

Quark	Charge (units of e)	Spin (s)	Baryon number	Strangeness number
Down (d)	$-1/3$	$1/2$	$1/3$	0
Up (u)	$+2/3$	$1/2$	$1/3$	0
Strange (s)	$-1/3$	$1/2$	$1/3$	-1

Quark	Charge (units of e)	Spin (s)	Baryon number	Strangeness number
Charm (c)	$+2/3$	$1/2$	$1/3$	0
Bottom (b)	$-1/3$	$1/2$	$1/3$	0
Top (t)	$+2/3$	$1/2$	$1/3$	0

Quarks

Quark Combinations

As mentioned earlier, quarks bind together in groups of two or three to form hadrons. Baryons are formed from three quarks. Sample baryons, including quark content and properties, are given in [\[link\]](#). Interestingly, the delta plus (Δ^+) baryon is formed from the same three quarks as the proton, but the total spin of the particle is $3/2$ rather than $1/2$. Similarly, the mass of Δ^+ with spin $3/2$ is 1.3 times the mass of the proton, and the delta zero (Δ^0) baryon with a spin $3/2$ is 1.3 times the neutron mass. Evidently, the energy associated with the spin (or angular momentum) of the particle contributes to its mass energy. It is also interesting that no baryons are believed to exist with top quarks, because top quarks decay too quickly to bind to the other quarks in their production.

Name	Symbol	Quarks	Charge (unit of e)	Spin (s)	Mass (GeV/c^2)
Proton	p	$u\ u\ d$	1	$1/2$	0.938
Neutron	n	$u\ d\ d$	0	$1/2$	0.940

Name	Symbol	Quarks	Charge (unit of e)	Spin (s)	Mass (GeV/c^2)
Delta plus plus	Δ^{++}	$u\ u\ u$	2	3/2	1.232
Delta plus	Δ^+	$u\ u\ d$	1	3/2	1.232
Delta zero	Δ^0	$u\ d\ d$	0	3/2	1.232
Delta minus	Δ^-	$d\ d\ d$	-1	3/2	1.232
Lambda zero	Λ^0	$u\ d\ s$	0	1/2	1.116
Positive sigma	Σ^+	$u\ u\ s$	1	1/2	1.189
Neutral sigma	Σ^0	$u\ d\ s$	0	1/2	1.192
Negative xi	Ξ^-	$s\ d\ s$	-1	1/2	1.321
Neutral xi	Ξ^0	$s\ u\ s$	0	1/2	1.315
Omega minus	Ω^-	$s\ s\ s$	-1	3/2	1.672
Charmed lambda	Λ_C^+	$u\ d\ c$	1	1/2	2.281

Name	Symbol	Quarks	Charge (unit of e)	Spin (s)	Mass (GeV/c^2)
Bottom lambda	Λ_b^0	$u d b$	0	1/2	5.619

Baryon Quarks

Mesons are formed by two quarks—a quark-antiquark pair. Sample mesons, including quark content and properties, are given in [\[link\]](#). Consider the

formation of the pion ($\pi^+ = u\bar{d}$). Based on its quark content, the charge of the pion is

Equation:

$$\frac{2}{3}e + \frac{1}{3}e = e.$$

Both quarks are spin-half ($s = 1/2$), so the resultant spin is either 0 or 1. The spin of the π^+ meson is 0. The same quark-antiquark combination gives the rho (ρ) meson with spin 1. This meson has a mass approximately 5.5 times that of the π^+ meson.

Example:

Quark Structure

Show that the quark composition given in [\[link\]](#) for Ξ^0 is consistent with the known charge, spin, and strangeness of this baryon.

Strategy

Ξ^0 is composed of two strange quarks and an up quark ($s u s$). We can add together the properties of quarks to predict the resulting properties of the Ξ^0 baryon.

Solution

The charge of the s quark is $-e/3$ and the charge of the u quark is $2e/3$. Thus, the combination ($s u s$) has no net charge, in agreement with the known charge of Ξ^0 . Since three spin $-1/2$ quarks can combine to produce a particle with spin of either $1/2$ or $3/2$, the quark composition is consistent with the known

spin ($s = 1/2$) of Ξ^0 . Finally, the net strangeness of the ($s u s$) combination is $(-1) + 0 + (-1) = -2$, which also agrees with experiment.

Significance

The charge, spin, and strangeness of the Ξ^0 particle can be determined from the properties of its constituent quarks. The great diversity of baryons and mesons can be traced to the properties of just six quarks: up, down, charge, strange, top, and bottom.

Note:

Exercise:

Problem:

Check Your Understanding What is the baryon number of a pion?

Solution:

0

Name	Symbol	Quarks	Charge (e)	Spin	Mass (GeV/ c^2)
Positive pion	π^+	$u\bar{d}$	1	0	0.140
Positive rho	ρ^+	$u\bar{d}$	1	1	0.768
Negative pion	π^-	$\bar{u}d$	-1	0	0.140

Name	Symbol	Quarks	Charge (e)	Spin	Mass (GeV/c ²)
Negative rho	ρ^-	$\bar{u}d$	-1	1	0.768
Neutral Pion	π^0	$\bar{u}u$ or $\bar{d}d$	0	0	0.135
Neutral eta	η^0	$\bar{u}u, \bar{d}d$ or $\bar{s}s$	0	0	0.547
Positive kaon	K^+	$u\bar{s}$	1	0	0.494
Neutral kaon	K^0	$d\bar{s}$	0	0	0.498
Negative kaon	K^-	$\bar{u}s$	-1	0	0.494
J/Psi	J/ψ	$\bar{c}c$	0	1	3.10
Charmed eta	η^0	$\bar{c}c$	0	0	2.98
Neutral D	D^0	$\bar{u}c$	0	0	1.86
Neutral D	D^{*0}	$\bar{u}c$	0	1	2.01
Positive D	D^+	$\bar{d}c$	1	0	1.87

Name	Symbol	Quarks	Charge (e)	Spin	Mass (GeV/ c^2)
Neutral B	B^0	$\bar{d}b$	0	0	5.26
Upsilon	Υ	$b\bar{b}$	0	1	9.46

Meson Quarks

Color

Quarks are fermions that obey Pauli’s exclusion principle, so it might be surprising to learn that three quarks can bind together within a nucleus. For example, how can two up quarks exist in the same small region of space within a proton? The solution is to invent a third new property to distinguish them. This property is called **color**, and it plays the same role in the strong nuclear interaction as charge does in electromagnetic interactions. For this reason, quark color is sometimes called “strong charge.”

Quarks come in three colors: red, green, and blue. (These are just labels—quarks are not actually colored.) Each type of quark (u, d, c, s, b, t) can possess any other colors. For example, three strange quarks exist: a red strange quark, a green strange quark, and a blue strange quark. Antiquarks have anticolor. Quarks that bind together to form hadrons (baryons and mesons) must be color neutral, colorless, or “white.” Thus, a baryon must contain a red, blue, and green quark. Likewise, a meson contains either a red-antired, blue-antiblue, or green-antigreen quark pair. Thus, two quarks can be found in the same spin state in a hadron, without violating Pauli’s exclusion principle, because their colors are different.

Quark Confinement

The first strong evidence for the existence of quarks came from a series of experiments performed at the Stanford Linear Accelerator Center (SLAC) and at CERN around 1970. This experiment was designed to probe the structure of the proton, much like Rutherford studied structure inside the atom with his α -

particle scattering experiments. Electrons were collided with protons with energy in excess of 20 GeV. At this energy, $E \approx pc$, so the de Broglie wavelength of an electron is

Equation:

$$\lambda = \frac{h}{p} = \frac{hc}{E} \approx 6 \times 10^{-17} \text{ m.}$$

The wavelength of the electron is much smaller than the diameter of the proton (about 10^{-15} m). Thus, like an automobile traveling through a rocky mountain range, electrons can be used to probe the structure of the nucleus.

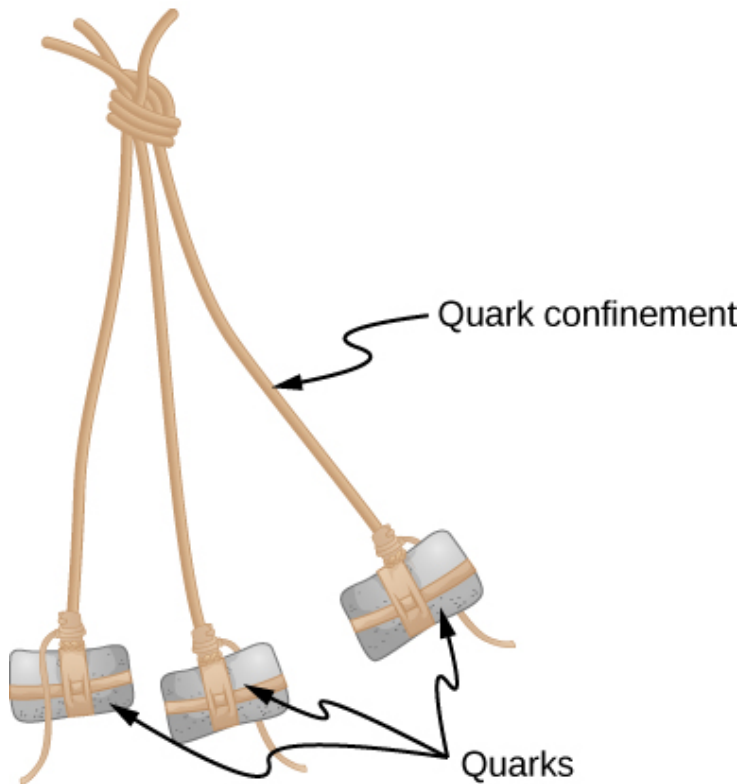
The SLAC experiments found that some electrons were deflected at very large angles, indicating small scattering centers within the proton. The scattering distribution was consistent with electrons being scattered from sites with spin 1/2, the spin of quarks. The experiments at CERN used neutrinos instead of electrons. This experiment also found evidence for the tiny scattering centers. In both experiments, the results suggested that the charges of the scattering particles were either $+2/3e$ or $-1/3e$, in agreement with the quark model.

Note:

Watch this [video](#) to learn more about quarks.

The quark model has been extremely successful in organizing the complex world of subatomic particles. Interestingly, however, no experiment has ever produced an isolated quark. All quarks have fractional charge and should therefore be easily distinguishable from the known elementary particles, whose charges are all an integer multiple of e . Why are isolated quarks not observed? In current models of particle interactions, the answer is expressed in terms of quark confinement. Quark confinement refers to the confinement of quarks in groups of two or three in a small region of space. The quarks are completely free to move about in this space, and send and receive gluons (the carriers of the strong force). However, if these quarks stray too far from one another, the strong force pulls them back it. This action is likened to a bola, a weapon used for hunting ([\[link\]](#)). The stones are tied to a central point by a string, so none of the

rocks can move too far from the others. The bola corresponds to a baryon, the stones correspond to quarks, and the string corresponds to the gluons that hold the system together.



A baryon is analogous to a bola, a weapon used for hunting. The rocks in this image correspond to the baryon quarks. The quarks are free to move about but must remain close to the other quarks.

Summary

- Six known quarks exist: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). These particles are fermions with half-integral spin and fractional charge.

- Baryons consist of three quarks, and mesons consist of a quark-antiquark pair. Due to the strong force, quarks cannot exist in isolation.
- Evidence for quarks is found in scattering experiments.

Conceptual Questions

Exercise:

Problem: What are the six known quarks? Summarize their properties.

Exercise:

Problem: What is the general quark composition of a baryon? Of a meson?

Solution:

3 quarks, 2 quarks (a quark-antiquark pair)

Exercise:

Problem: What evidence exists for the existence of quarks?

Exercise:

Problem:

Why do baryons with the same quark composition sometimes differ in their rest mass energies?

Solution:

Baryons with the same quark composition differ in rest energy because this energy depends on the internal energy of the quarks $m = E/c^2$. So, a baryon that contains a quark with a large angular momentum is expected to be more massive than the same baryon with less angular momentum.

Problems

Exercise:

Problem:

Based on quark composition of a proton, show that its charge is $+1$.

Solution:

A proton consists of two up quarks and one down quark. The total charge of a proton is therefore $+\frac{2}{3} + \frac{2}{3} + -\frac{1}{3} = +1$.

Exercise:**Problem:**

Based on the quark composition of a neutron, show that its charge is 0 .

Exercise:**Problem:**

Argue that the quark composition given in [\[link\]](#) for the positive kaon is consistent with the known charge, spin, and strangeness of this baryon.

Solution:

The K^+ meson is composed of an up quark and a strange antiquark ($u\bar{s}$). Since the charges of this quark and antiquark are $2e/3$ and $e/3$, respectively, the net charge of the K^+ meson is e , in agreement with its known value. Two spin $-1/2$ particles can combine to produce a particle with spin of either 0 or 1 , consistent with the K^+ meson's spin of 0 . The net strangeness of the up quark and strange antiquark is $0 + 1 = 1$, in agreement with the measured strangeness of the K^+ meson.

Exercise:**Problem:**

Mesons are formed from the following combinations of quarks (subscripts indicate color and $AR = \text{antired}$): (d_R, \bar{d}_{AR}) , (s_G, \bar{u}_{AG}) , and (s_R, \bar{s}_{AR}) .

(a) Determine the charge and strangeness of each combination. (b) Identify one or more mesons formed by each quark-antiquark combination.

Exercise:

Problem: Why can't either set of quarks shown below form a hadron?



(a)



(b)

Solution:

a. color; b. quark-antiquark

Exercise:

Problem:

Experimental results indicate an isolated particle with charge $+2/3$ —an isolated quark. What quark could this be? Why would this discovery be important?

Exercise:

Problem:

Express the β decays $n \rightarrow p + e^- + \bar{\nu}$ and $p \rightarrow n + e^+ + \nu$ in terms of β decays of quarks. Check to see that the conservation laws for charge, lepton number, and baryon number are satisfied by the quark β decays.

Solution:

$$d \rightarrow u + e^- + \bar{\nu}_e; u \rightarrow d + e^+ + \nu_e$$

Glossary

color

property of particles and that plays the same role in strong nuclear interactions as electric charge does in electromagnetic interactions

Particle Accelerators and Detectors

By the end of this section, you will be able to:

- Compare and contrast different types of particle accelerators
- Describe the purpose, components, and function of a typical colliding beam machine
- Explain the role of each type of subdetector of a typical multipurpose particle detector
- Use the curvature of a charge track to determine the momentum of a particle

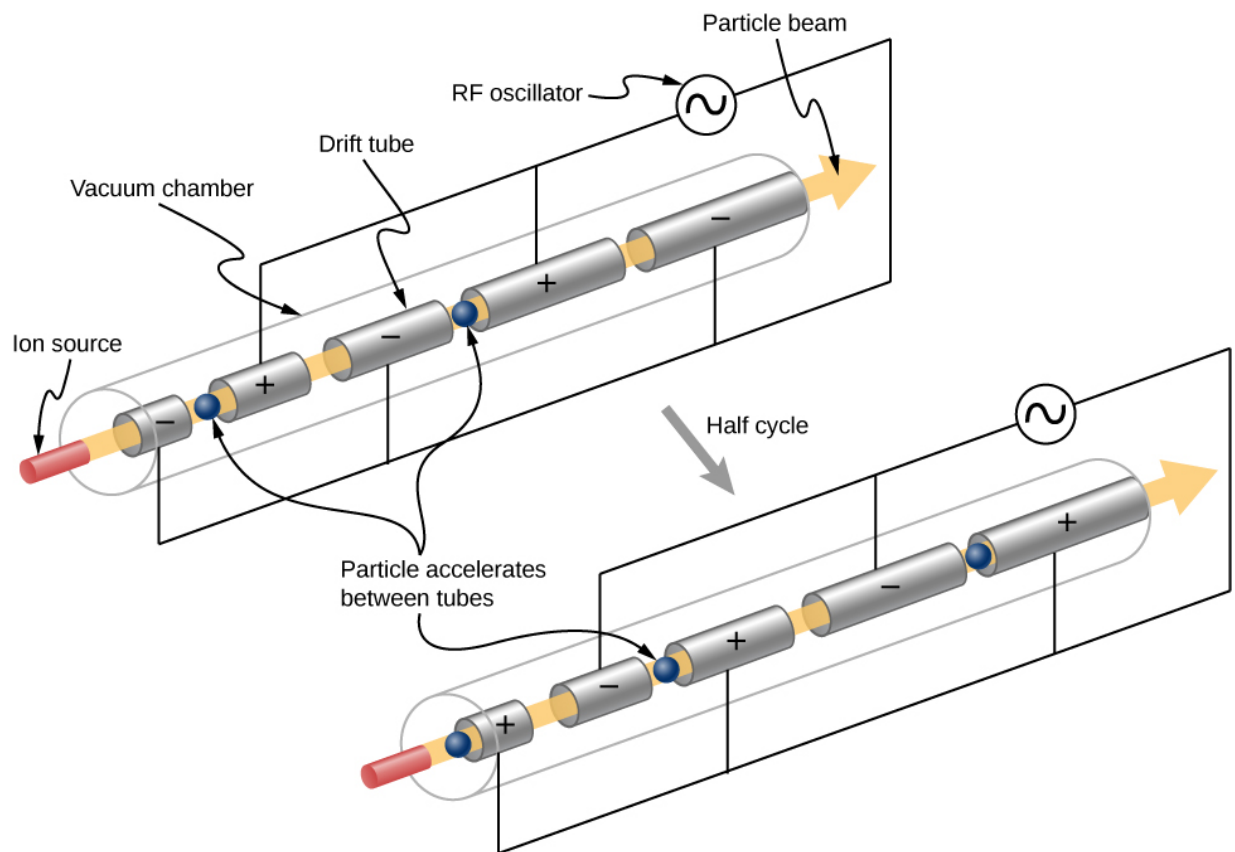
The goal of experimental particle physics is to accurately measure elementary particles. The primary method used to achieve this end is to produce these particles in high-energy collisions and then measure the products of using highly sensitive particle detectors. These experiments are used to test and revise scientific models of particle interactions. The purpose of this section is to describe particle accelerators and detectors. Modern machines are based on earlier ones, so it is helpful to present a brief history of accelerators and detectors.

Early Particle Accelerators

A **particle accelerator** is a machine designed to accelerate charged particles. This acceleration is usually achieved with strong electric fields, magnetic fields, or both. A simple example of a particle accelerator is the Van de Graaff accelerator (see [Electric Potential](#)). This type of accelerator collects charges on a hollow metal sphere using a moving belt. When the electrostatic potential difference of the sphere is sufficiently large, the field is used to accelerate particles through an evacuated tube. Energies produced by a Van de Graaff accelerator are not large enough to create new particles, but the machine was important for early exploration of the atomic nucleus.

Larger energies can be produced by a linear accelerator (called a “linac”). Charged particles produced at the beginning of the linac are accelerated by a continuous line of charged hollow tubes. The voltage between a given pair of tubes is set to draw the charged particle in, and once the particle arrives,

the voltage between the next pair of tubes is set to push the charged particle out. In other words, voltages are applied in such a way that the tubes deliver a series of carefully synchronized electric kicks ([\[link\]](#)). Modern linacs employ radio frequency (RF) cavities that set up oscillating electromagnetic fields, which propel the particle forward like a surfer on an ocean wave. Linacs can accelerate electrons to over 100 MeV. (Electrons with kinetic energies greater than 2 MeV are moving very close to the speed of light.) In modern particle research, linear accelerators are often used in the first stage of acceleration.



In a linear accelerator, charged tubes accelerate particles in a series of electromagnetic kicks. Each tube is longer than the preceding tube because the particle is moving faster as it accelerates.

Example:**Accelerating Tubes**

A linear accelerator designed to produce a beam of 800-MeV protons has 2000 accelerating tubes separated by gaps. What average voltage must be applied between tubes to achieve the desired energy? (*Hint: $U = qV$.*)

Strategy

The energy given to the proton in each gap between tubes is $U = qV$, where q is the proton's charge and V is the potential difference (voltage) across the gap. Since $q = q_e = 1.6 \times 10^{-19} \text{ C}$ and $1 \text{ eV} = (1 \text{ V}) (1.6 \times 10^{-19} \text{ C})$, the proton gains 1 eV in energy for each volt across the gap that it passes through. The ac voltage applied to the tubes is timed so that it adds to the energy in each gap. The effective voltage is the sum of the gap voltages and equals 800 MV to give each proton an energy of 800 MeV.

Solution

There are 2000 gaps and the sum of the voltages across them is 800 MV. Therefore, the average voltage applied is 0.4 MV or 400 kV.

Significance

A voltage of this magnitude is not difficult to achieve in a vacuum. Much larger gap voltages would be required for higher energy, such as those at the 50-GeV SLAC facility. Synchrotrons are aided by the circular path of the accelerated particles, which can orbit many times, effectively multiplying the number of accelerations by the number of orbits. This makes it possible to reach energies greater than 1 TeV.

Note:**Exercise:****Problem:**

Check Your Understanding How much energy does an electron receive in accelerating through a 1-V potential difference?

Solution:

1 eV

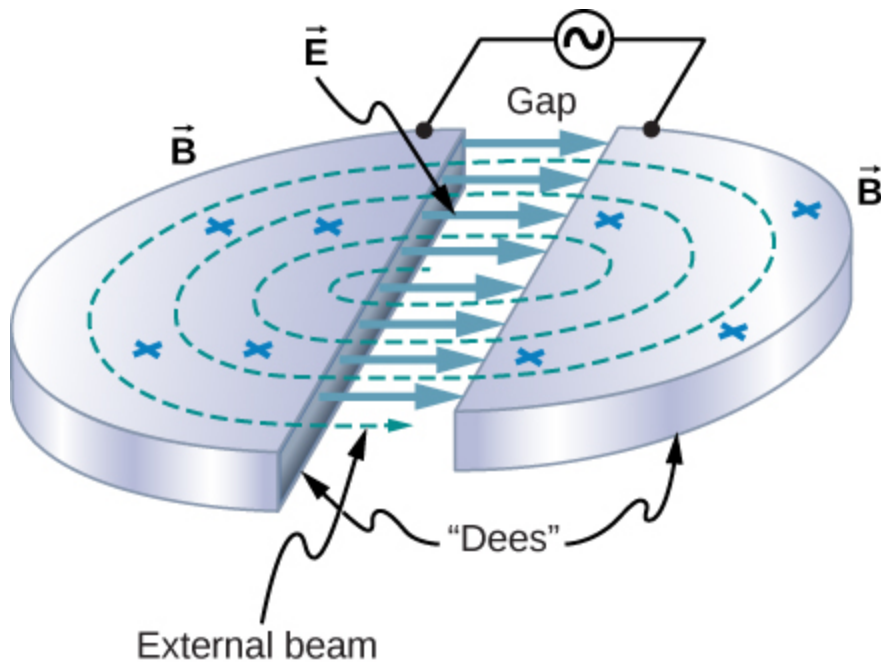
The next-generation accelerator after the linac is the cyclotron ([\[link\]](#)). A cyclotron uses alternating electric fields and fixed magnets to accelerate particles in a circular spiral path. A particle at the center of the cyclotron is first accelerated by an electric field in a gap between two D-shaped magnets (Dees). As the particle crosses over the D-shaped magnet, the particle is bent into a circular path by a Lorentz force. (The Lorentz force was discussed in [Magnetic Forces and Fields](#).) Assuming no energy losses, the momentum of the particle is related to its radius of curvature by

Note:

Equation:

$$p = 0.3Br$$

where p is the momentum in GeV/ c , B is in teslas, and r is the radius of the trajectory (“orbit”) in meters. This expression is valid to classical and relativistic velocities. The circular trajectory returns the particle to the electric field gap, the electric field is reversed, and the process continues. As the particle is accelerated, the radius of curvature gets larger and larger—spirally outward—until the electrons leave the device.



Cyclotrons use a magnetic field to cause particles to move in circular orbits. As the particles pass between the plates of the “Dees,” the voltage across the gap is reversed so the particles are accelerated twice in each orbit.

Note:

Watch this [video](#) to learn more about cyclotrons.

A **synchrotron** is a circular accelerator that uses alternating voltage and increasing magnetic field strength to accelerate particles to higher energies. Charged particles are accelerated by RF cavities, and steered and focused by magnets. RF cavities are *synchronized* to deliver “kicks” to the particles as they pass by, hence the name. Steering high-energy particles requires strong magnetic fields, so superconducting magnets are often used to reduce heat losses. As the charged particles move in a circle, they radiate energy:

According to classical theory, any charged particle that accelerates (and circular motion is an accelerated motion) also radiates. In a synchrotron, such radiation is called **synchrotron radiation**. This radiation is useful for many other purposes, such as medical and materials research.

Example:

The Energy of an Electron in a Cyclotron

An electron is accelerated using a cyclotron. If the magnetic field is 1.5 T and the radius of the “Dees” is 1.2 m, what is the kinetic energy of the outgoing particle?

Strategy

If the radius of orbit of the electron exceeds the radius of the “Dees,” the electron exits the device. So, the radius of the “Dees” places an upper limit on the radius and, therefore, the momentum and energy of the accelerated particle. The exit momentum of the particle is determined using the radius of orbit and strength of the magnetic field. The exit energy of the particle can be determined the particle momentum ([Relativity](#)).

Solution

Assuming no energy losses, the momentum of the particle in the cyclotron is

Equation:

$$p = 0.3Br = 0.3(1.5 \text{ T})(1.2 \text{ m}) = 0.543 \text{ GeV}/c.$$

The momentum energy $pc^2 = 0.543 \text{ GeV} = 543 \text{ MeV}$ is much larger than the rest mass energy of the electron, $mc^2 = 0.511 \text{ MeV}$, so relativistic expression for the energy of the electron must be used (see [Relativity](#)). The total energy of the electron is

Equation:

$$E_{\text{total}} = \sqrt{(pc)^2 + (mc^2)^2} = \sqrt{(543)^2 + (0.511)^2} \approx 543 \text{ MeV and}$$

Equation:

$$K = E_{\text{total}} - mc^2 = 543 \text{ GeV} - 0.511 \text{ GeV} \approx 543 \text{ MeV}.$$

Significance

The total energy of the electron is much larger than its rest mass energy. In other words, the total energy of the electron is almost all in the form of kinetic energy. Cyclotrons can be used to conduct nuclear physics experiments or in particle therapy to treat cancer.

Note:**Exercise:****Problem:**

Check Your Understanding A charged particle of a certain momentum travels in an arc through a uniform magnetic field. What happens if the magnetic field is doubled?

Solution:

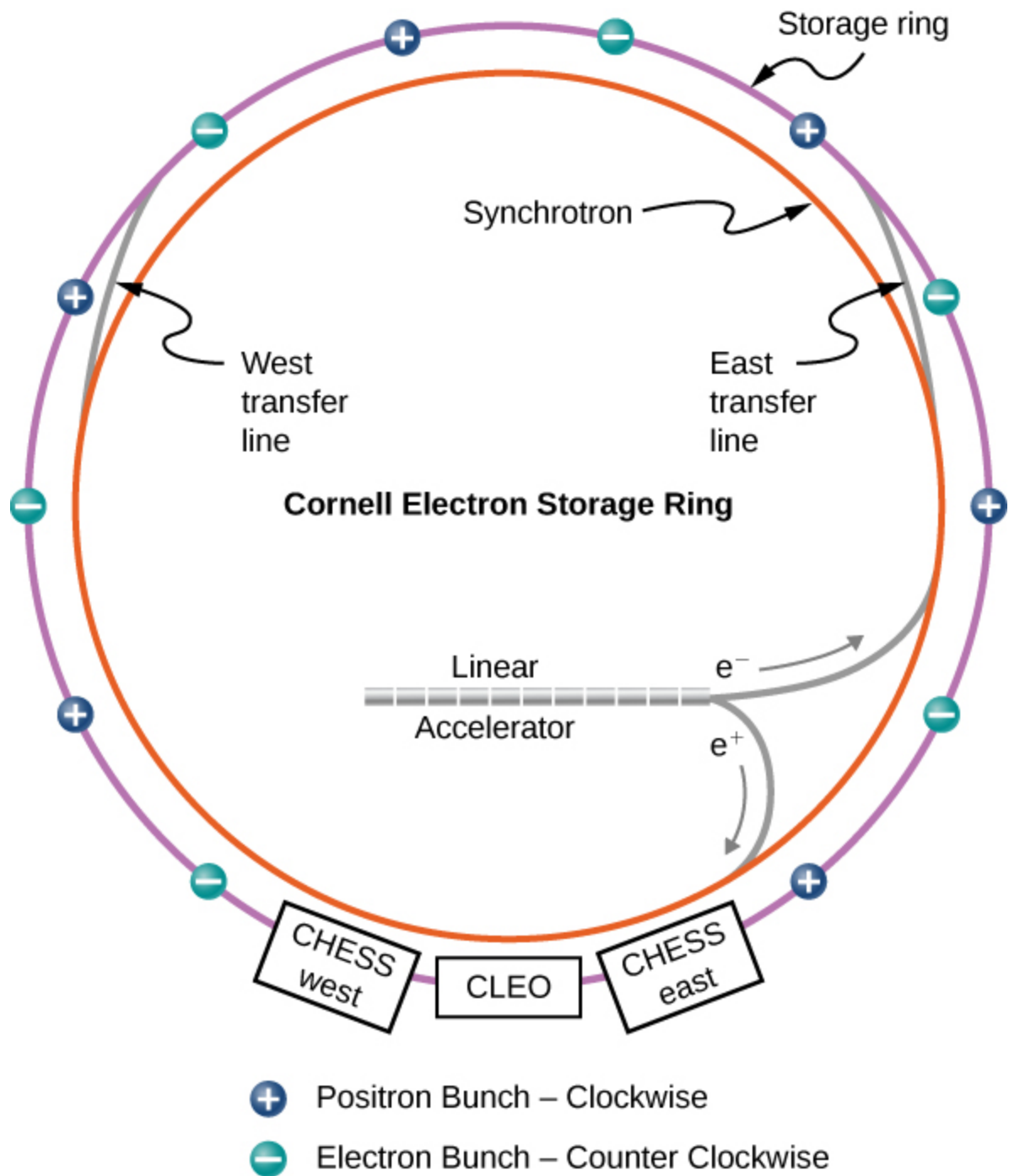
The radius of the track is cut in half.

Colliding Beam Machines

New particles can be created by colliding particles at high energies. According to Einstein's mass-energy relation, the energies of the colliding particles are converted into mass energy of the created particle. The most efficient way to do this is with particle-colliding beam machines. A colliding beam machine creates two counter-rotating beams in a circular accelerator, stores the beams at constant energy, and then at the desired moment, focuses the beams on one another at the center of a sensitive detector.

The prototypical colliding beam machine is the Cornell Electron Storage Ring, located in Ithaca, New York ([\[link\]](#)). Electrons (e^-) and positrons (e^+) are created at the beginning of the linear accelerator and are accelerated up to 150 MeV. The particles are then injected into the inner synchrotron ring,

where they are accelerated by RF cavities to 4.5 to 6 GeV. When the beams are up to speed, they are transferred and “stored” in an outer storage ring at the same energy. The two counter-rotating beams travel through the same evacuated pipe, but are kept apart until collisions are desired. The electrons and positrons circle the machine in bunches 390,000 times every second.



The Cornell Electron Storage Ring uses a linear accelerator and a synchrotron to accelerate electrons and positrons to 4.5–6 GeV. The particles are held in the outer storage ring at that energy until they are made to collide in a particle detector. (credit: modification of work by Laboratory of Nuclear Studies, Cornell Electron Storage Ring)

When an electron and positron collide, they annihilate each other to produce a photon, which exists for too short a time to be detected. The photon produces either a lepton pair (e.g., an electron and positron, muon or antimuon, or tau and antitau) or a quark pair. If quarks are produced, mesons form, such as $c\bar{c}$ and $b\bar{b}$. These mesons are created nearly at rest since the initial total momentum of the electron-positron system is zero. Note, mesons cannot be created at just any colliding energy but only at “resonant” energies that correspond to the unique masses of the mesons ([link](#)). The mesons created in this way are highly unstable and decay quickly into lighter particles, such as electrons, protons, and photons. The collision “fragments” provide valuable information about particle interactions.

As the field of particle physics advances, colliding beam machines are becoming more powerful. The Large Hadron Collider (LHC), currently the largest accelerator in the world, collides protons at beam energies exceeding 6 TeV. The center-of-mass energy (W) refers to the total energy available to create new particles in a colliding machine, or the total energy of incoming particles in the center-of-mass frame. (The concept of a center-of-mass frame of reference is discussed in [Linear Momentum and Collisions](#).) Therefore, the LHC is able to produce one or more particles with a total mass exceeding 12 TeV. The center-of-mass energy is given by:

Note:

Equation:

$$W^2 = 2 [E_1 E_2 + (p_1 c) (p_2 c)] + (m_1 c^2)^2 + (m_2 c^2)^2,$$

where E_1 and E_2 are the total energies of the incoming particles (1 and 2), p_1 and p_2 are the magnitudes of their momenta, and m_1 and m_2 are their rest masses.

Example:**Creating a New Particle**

The mass of the upsilon (Υ) meson ($b\bar{b}$) is created in a symmetric electron-positron collider. What beam energy is required?

Strategy

[The Particle Data Group](#) has stated that the rest mass energy of this meson is approximately 10.58 GeV. The above expression for the center-of-mass energy can be simplified because a symmetric collider implies $\vec{p}_1 = -\vec{p}_2$. Also, the rest masses of the colliding electrons and positrons are identical ($m_e c^2 = 0.511$ MeV) and much smaller than the mass of the energy particle created. Thus, the center-of-mass energy (W) can be expressed completely in terms of the beam energy, $E_{\text{beam}} = E_1 = E_2$.

Solution

Based on the above assumptions, we have

Equation:

$$W^2 \approx 2 [E_1 E_2 + E_1 E_2] = 4E_1 E_2 = 4E_1^2.$$

The beam energy is therefore

Equation:

$$E_{\text{beam}} \approx E_1 = \frac{W}{2}.$$

The rest mass energy of the particle created in the collision is equal to the center-of-mass energy, so

Equation:

$$E_{\text{beam}} \approx \frac{10.58 \text{ GeV}}{2} = 5.29 \text{ GeV}.$$

Significance

Given the energy scale of this problem, the rest mass energy of the upsilon (Υ) meson is due almost entirely due to the initial kinetic energies of the electron and positrons. This meson is highly unstable and quickly decays to lighter and more stable particles. The existence of the upsilon (Υ) particle appears as a dramatic increase of such events at 5.29 GeV.

Note:

Exercise:

Problem:

Check Your Understanding Why is a symmetric collider “symmetric?”

Solution:

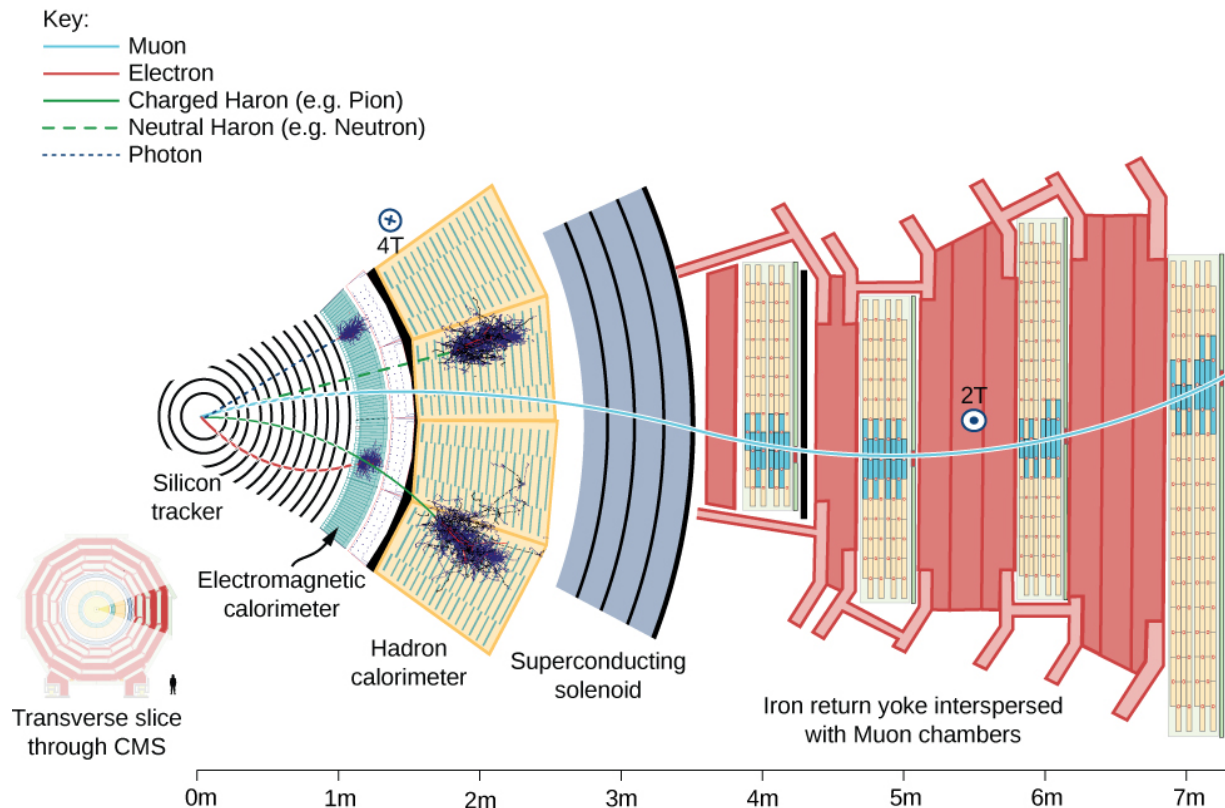
The colliding particles have identical mass but opposite vector momenta.

Higher beam energies require larger accelerators, so modern colliding beam machines are very large. The LHC, for example, is 17 miles in circumference ([link](#)). (In the 1940s, Enrico Fermi envisioned an accelerator that encircled all of Earth!) An important scientific challenge of the twenty-first century is to reduce the size of particle accelerators.

Particle Detectors

The purpose of a **particle detector** is to accurately measure the outcome of collisions created by a particle accelerator. The detectors are multipurpose. In other words, the detector is divided into many subdetectors, each designed to measure a different aspect of the collision event. For example, one detector might be designed to measure photons and another might be

designed to measure muons. To illustrate how subdetectors contribute to an understanding of an entire collision event, we describe the subdetectors of the Compact Muon Solenoid (CMS), which was used to discover the Higgs Boson at the LHC ([\[link\]](#)).

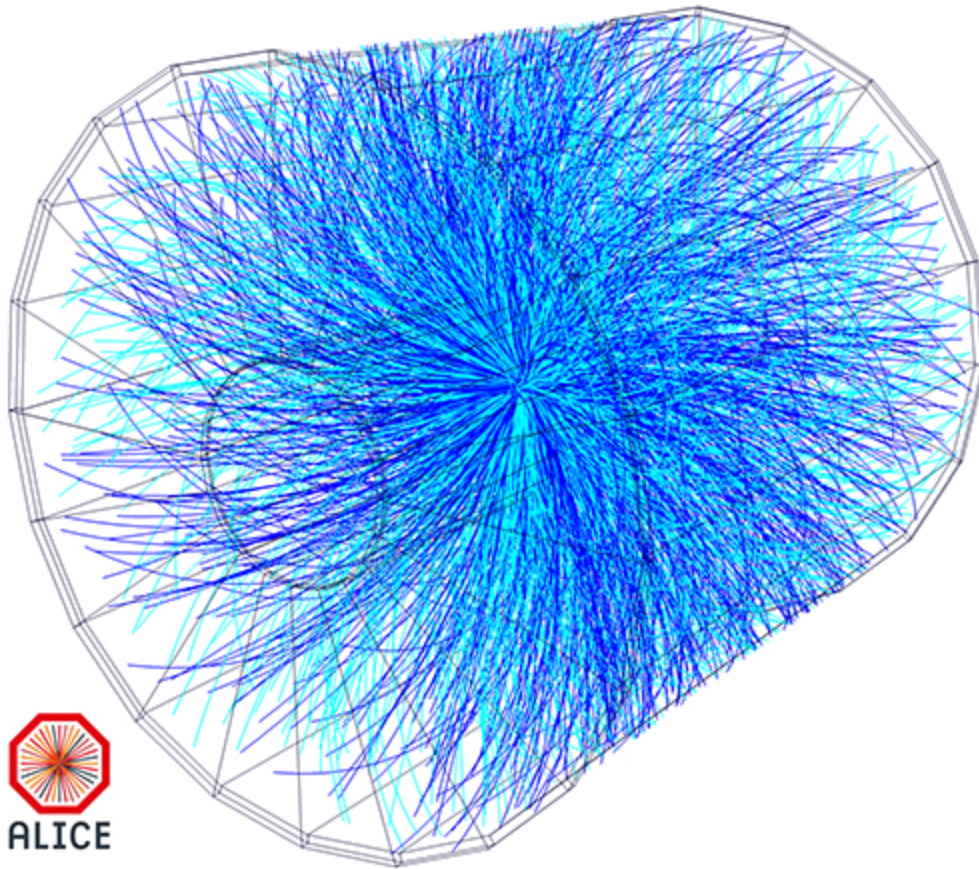


Compact Muon Solenoid detector. The detector consists of several layers, each responsible for measuring different types of particles.
(credit: modification of work by David Barney/CERN)

The beam pipe of the detector is out of (and into) the page at the left. Particles produced by pp collisions (the “collision fragments”) stream out of the detector in all directions. These particles encounter multiple layers of subdetectors. A subdetector is a particle detector within a larger system of detectors designed to measure certain types of particles. There are several main types of subdetectors. Tracking devices determine the path and

therefore momentum of a particle; calorimeters measure a particle's energy; and particle-identification detectors determine a particle's identity (mass).

The first set of subdetectors that particles encounter is the silicon tracking system. This system is designed to measure the momentum of charged particles (such as electrons and protons). The detector is bathed in a uniform magnetic field, so the charged particles are bent in a circular path by a Lorentz force (as for the cyclotron). If the momentum of the particle is large, the radius of the trajectory is large, and the path is almost straight. But if the momentum is small, the radius of the trajectory is small, and the path is tightly curved. As the particles pass through the detector, they interact with silicon microstrip detectors at multiple points. These detectors produce small electrical signals as the charged particles pass near the detector elements. The signals are then amplified and recorded. A series of electrical "hits" is used to determine the trajectory of the particle in the tracking system. A computer-generated "best fit" to this trajectory gives the track radius and therefore the particle momentum. At the LHC, a large number of tracks are recorded for the same collision event. Fits to the tracks are shown by the blue and green lines in [\[link\]](#).



A three-dimensional view of a heavy-ion collision event in the LHC as seen by the ALICE detector. (credit: LHC/CERN)

Beyond the tracking layers is the electromagnetic calorimeter. This detector is made of clear, lead-based crystals. When electrons interact with the crystals, they radiate high-energy photons. The photons interact with the crystal to produce electron-positron pairs. Then, these particles radiate more photons. The process repeats, producing a particle shower (the crystal “glows”). A crude model of this process is as follows.

An electron with energy E_0 strikes the crystal and loses half of its energy in the form of a photon. The photon produces an electron-positron pair, and each particle proceeds away with half the energy of the photon. Meanwhile,

the original electron radiates again. So, we are left with four particles: two electrons, one positron, and one photon, each with an energy $E_0/4$. The number of particles in the shower increases geometrically. After n radiation events, there are $N = 2^n$ particles. Hence, the total energy per particle after n radiation events is

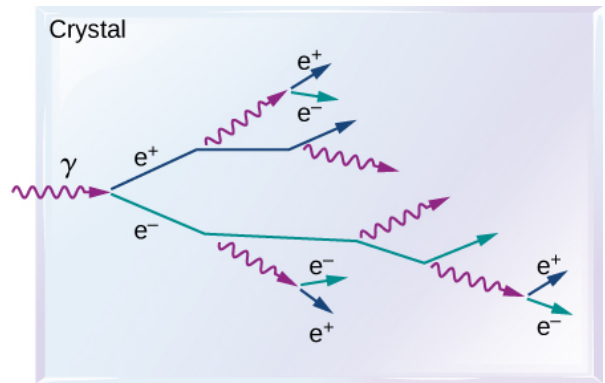
Equation:

$$E(t) = \frac{E_0}{2^n},$$

where E_0 is the incident energy and $E(t)$ is the amount of energy per particle after n events. An incoming photon triggers a similar chain of events ([\[link\]](#)). If the energy per particle drops below a particular threshold value, other types of radiative processes become important and the particle shower ceases. Eventually, the total energy of the incoming particle is absorbed and converted into an electrical signal.



(a)



(b)

(a) A particle shower produced in a crystal calorimeter. (b) A diagram showing a typical sequence of reactions in a particle shower.

Beyond the crystal calorimeter is the hadron calorimeter. As the name suggests, this subdetector measures hadrons such as protons and pions. The

hadron calorimeter consists of layers of brass and steel separated by plastic scintillators. Its purpose is to absorb the particle energy and convert it into an electronic signal. Beyond this detector is a large magnetic coil used to produce a uniform field for tracking.

The last subdetector is the muon detector, which consists of slabs of iron that only muons (and neutrinos) can penetrate. Between the iron slabs are multiple types of muon-tracking elements that accurately measure the momentum of the muon. The muon detectors are important because the Higgs boson (discussed soon) can be detected through its decays to four muons—hence the name of the detector.

Once data is collected from each of the particle subdetectors, the entire collision event can be assessed. The energy of the i th particle is written

Equation:

$$E_i = \sqrt{(p_i c)^2 + (m_i c^2)^2},$$

where p_i is the absolute magnitude of the momentum of the i th particle, and m_i is its rest mass.

The total energy of all particles is therefore

Equation:

$$E_{\text{total}} = \sum_i E_i.$$

If all particles are detected, the total energy should be equal to the center-of-mass energy of the colliding beam machine (W). In practice, not all particles are identified, either because these particles are too difficult to detect (neutrinos) or because these particles “slip through.” In many cases, whole chains of decays can be “reconstructed,” like putting back together a watch that has been smashed to pieces. Information about these decay chains are critical to the evaluation of models of particle interactions.

Summary

- Many types of particle accelerators have been developed to study particles and their interactions. These include linear accelerators, cyclotrons, synchrotrons, and colliding beams.
- Colliding beam machines are used to create massive particles that decay quickly to lighter particles.
- Multipurpose detectors are used to design all aspects of high-energy collisions. These include detectors to measure the momentum and energies of charge particles and photons.
- Charged particles are measured by bending these particles in a circle by a magnetic field.
- Particles are measured using calorimeters that absorb the particles.

Conceptual Questions

Exercise:

Problem:

Briefly compare the Van de Graaff accelerator, linear accelerator, cyclotron, and synchrotron accelerator.

Exercise:

Problem:

Describe the basic components and function of a typical colliding beam machine.

Solution:

the “linac” to accelerate the particles in a straight line, a synchrotron to accelerate and store the moving particles in a circular ring, and a detector to measure the products of the collisions

Exercise:

Problem:

What are the subdetectors of the Compact Muon Solenoid experiment? Briefly describe them.

Exercise:**Problem:**

What is the advantage of a colliding-beam accelerator over one that fires particles into a fixed target?

Solution:

In a colliding beam experiment, the energy of the colliding particles goes into the rest mass energy of the new particle. In a fix-target experiment, some of this energy is lost to the momentum of the new particle since the center-of-mass of colliding particles is not fixed.

Exercise:**Problem:**

An electron appears in the muon detectors of the CMS. How is this possible?

Problems**Exercise:****Problem:**

A charged particle in a 2.0-T magnetic field is bent in a circle of radius 75 cm. What is the momentum of the particle?

Exercise:

Problem:

A proton track passes through a magnetic field with radius of 50 cm. The magnetic field strength is 1.5 T. What is the total energy of the proton?

Solution:

965 GeV

Exercise:**Problem:**

Derive the equation $p = 0.3Br$ using the concepts of centripetal acceleration ([Motion in Two and Three Dimensions](#)) and relativistic momentum ([Relativity](#)).

Exercise:**Problem:**

Assume that beam energy of an electron-positron collider is approximately 4.73 GeV. What is the total mass (W) of a particle produced in the annihilation of an electron and positron in this collider? What meson might be produced?

Solution:

According to [\[link\]](#),
 $W = 2E_{\text{beam}} = 9.46 \text{ GeV}$,
 $M = 9.46 \text{ GeV}/c^2$.

This is the mass of the upsilon ($1S$) meson first observed at Fermi lab in 1977. The upsilon meson consists of a bottom quark and its antiparticle $\left(b\bar{b}\right)$.

Exercise:

Problem:

At full energy, protons in the 2.00-km-diameter Fermilab synchrotron travel at nearly the speed of light, since their energy is about 1000 times their rest mass energy. (a) How long does it take for a proton to complete one trip around? (b) How many times per second will it pass through the target area?

Exercise:**Problem:**

Suppose a W^- created in a particle detector lives for 5.00×10^{-25} s. What distance does it move in this time if it is traveling at $0.900c$? (Note that the time is longer than the given W^- lifetime, which can be due to the statistical nature of decay or time dilation.)

Solution:

0.135 fm; Since this distance is too short to make a track, the presence of the W^- must be inferred from its decay products.

Exercise:**Problem:**

What length track does a π^+ traveling at $0.100c$ leave in a bubble chamber if it is created there and lives for 2.60×10^{-8} s? (Those moving faster or living longer may escape the detector before decaying.)

Exercise:**Problem:**

The 3.20-km-long SLAC produces a beam of 50.0-GeV electrons. If there are 15,000 accelerating tubes, what average voltage must be across the gaps between them to achieve this energy?

Solution:

3.33 MV

Glossary

particle accelerator

machine designed to accelerate charged particles; this acceleration is usually achieved with strong electric fields, magnetic fields, or both

particle detector

detector designed to accurately measure the outcome of collisions created by a particle accelerator; particle detectors are hermetic and multipurpose

synchrotron

circular accelerator that uses alternating voltage and increasing magnetic field strengths to accelerate particles to higher and higher energies

synchrotron radiation

high-energy radiation produced in a synchrotron accelerator by the circular motion of a charged beam

The Standard Model

By the end of this section, you will be able to:

- Describe the Standard Model in terms of the four fundamental forces and exchange particles
- Draw a Feynman diagram for a simple particle interaction
- Use Heisenberg's uncertainty principle to determine the range of forces described by the Standard Model
- Explain the rationale behind grand unification theories

The chief intellectual activity of any scientist is the development and revision of scientific models. A particle physicist seeks to develop models of particle interactions. This work builds directly on work done on gravity and electromagnetism in the seventeenth, eighteenth, and nineteenth centuries. The ultimate goal of physics is a unified “theory of everything” that describes all particle interactions in terms of a single elegant equation and a picture. The equation itself might be complex, but many scientists suspect the *idea* behind the equation will make us exclaim: “How could we have missed it? It was so obvious!”

In this section, we introduce the Standard Model, which is the best current model of particle interactions. We describe the Standard Model in detail in terms of electromagnetic, weak nuclear, and strong forces. At the end of this section, we review unification theories in particle physics.

Introduction to the Standard Model

The **Standard Model** of particle interactions contains two ideas: *electroweak theory* and **quantum chromodynamics (QCD)** (the force acting between color charges). Electroweak theory unifies the theory of **quantum electrodynamics (QED)**, the modern equivalent of classical electromagnetism, and the theory of weak nuclear interactions. The Standard Model combines the theory of relativity and quantum mechanics.

In the Standard Model, particle interactions occur through the exchange of bosons, the “force carriers.” For example, the electrostatic force is communicated between two positively charged particles by sending and receiving massless photons. This can occur at a theoretical infinite range. The

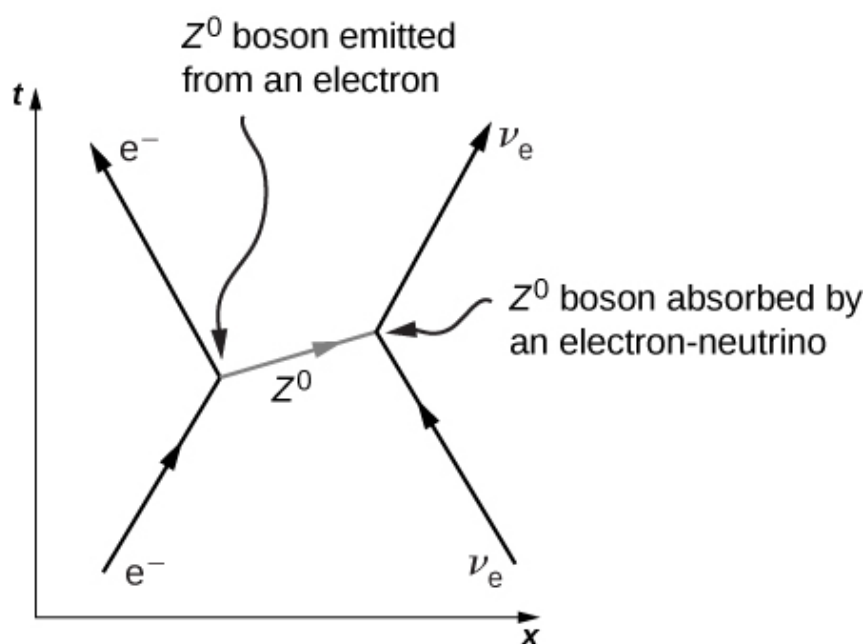
result of these interactions is Coulomb repulsion (or attraction). Similarly, quarks bind together through the exchange of massless gluons. Leptons scatter off other leptons (or decay into lighter particles) through the exchange of massive W and Z bosons. A summary of forces as described by the Standard Model is given in [\[link\]](#). The gravitational force, mediated by the exchange of massless gravitons, is added in this table for completeness but is not part of the Standard Model.

Force	Relative strength	Exchange particle (bosons)	Particles acted upon	Range
Strong	1	Gluon	Quarks	10^{-15} m
Electromagnetic	1/137	photon	Charged particles	∞
Weak	10^{-10}	W^+ , W^- , Z bosons	Quarks, leptons, neutrinos	10^{-18} m
Gravitational	10^{-38}	graviton	All particles	∞

Four Forces and the Standard Model

The Standard Model can be expressed in terms of equations and diagrams. The equations are complex and are usually covered in a more advanced course in modern physics. However, the essence of the Standard Model can be captured using **Feynman diagrams**. A Feynman diagram, invented by American physicist Richard Feynman (1918–1988), is a space-time diagram that describes how particles move and interact. Different symbols are used for different particles. Particle interactions in one dimension are shown as a time-position graph (not a position-time graph). As an example, consider the scattering of an

electron and electron-neutrino ([link](#)). The electron moves toward positive values of x (to the right) and collides with an electron neutrino moving to the left. The electron exchanges a Z boson (charge zero). The electron scatters to the left and the neutrino scatters to the right. This exchange is not instantaneous. The Z boson travels from one particle to the other over a short period of time. The interaction of the electron and neutrino is said to occur via the weak nuclear force. This force cannot be explained by classical electromagnetism because the charge of the neutrino is zero. The weak nuclear force is discussed again later in this section.

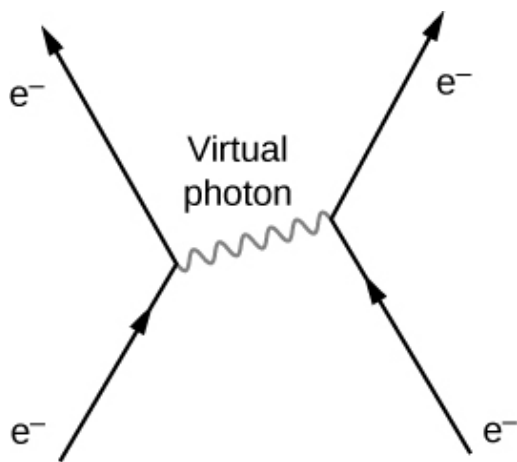


In this Feynman diagram, the exchange of a virtual Z^0 carries the weak nuclear force between an electron and a neutrino.

Electromagnetic Force

According to QED, the electromagnetic force is transmitted between charged particles through the exchange of photons. The theory is based on three basic processes: An electron travels from one place to the next, emits or absorbs a

photon, and travels from one place to another again. When two electrons interact, one electron emits the photon and the other receives it ([\[link\]](#)). Photons transfer energy and momentum from one electron to the other. The net result in this case is a repulsive force. The photons exchanged are virtual. A **virtual particle** is a particle that exists for too short a time to be observable. Since the photon transit time Δt is extremely small, Heisenberg's uncertainty principle states that the uncertainty in the photon's energy, ΔE , may be very large.



Feynman diagram of two electrons interacting through the exchange of a photon.

To estimate the range of the electromagnetic interaction, assume that the uncertainty on the energy is comparable to the energy of the photon itself, written

Equation:

$$\Delta E \approx E.$$

The Heisenberg uncertainty principle states that

Equation:

$$\Delta E \approx \frac{h}{\Delta t}.$$

Combining these equations, we have

Note:

Equation:

$$\Delta t \approx \frac{h}{E}.$$

The energy of a photon is given by $E = hf$, so

Equation:

$$\Delta t \approx \frac{h}{hf} \approx \frac{1}{f} = \frac{\lambda}{c}.$$

The distance d that the photon can move in this time is therefore

Equation:

$$d = c \Delta t \approx c \left(\frac{\lambda}{c} \right) = \lambda.$$

The energy of the virtual photon can be arbitrarily small, so its wavelength can be arbitrarily large—in principle, even infinitely large. The electromagnetic force is therefore a long-range force.

Weak Nuclear Force

The weak nuclear force is responsible for radioactive decay. The range of the weak nuclear force is very short (only about 10^{-18} m) and like the other forces in the Standard Model, the weak force can be described in terms of particle

exchange. (There is no simple function like the Coulomb force to describe these interactions.) The particle exchanged is one of three bosons: W^+ , W^- , and Z^0 . The Standard Model predicts the existence of these spin-1 particles and also predicts their specific masses. In combination with previous experiments, the mass of the charged W bosons was predicted to be $81 \text{ GeV}/c^2$ and that of the Z^0 was predicted to be $90 \text{ GeV}/c^2$. A CERN experiment discovered particles in the 1980s with precisely these masses—an impressive victory for the model.

The weak nuclear force is most frequently associated with scattering and decays of unstable particles to light particles. For example, neutrons decay to protons through the weak nuclear force. This reaction is written

Equation:

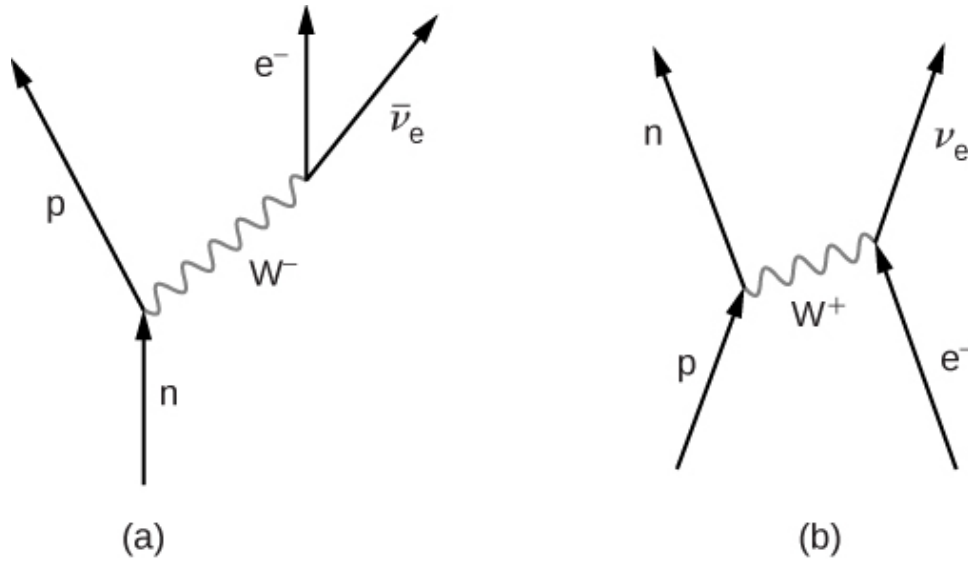
$$n \rightarrow p + e^- + \nu_e,$$

where n is the neutron, p is a proton, e^- is an electron, and ν_e is a nearly massless electron neutrino. This process, called beta decay, is important in many physical processes. A Feynman diagram of beta decay is given in [\[link\]](#) (a). The neutron emits a W^- and becomes a proton, then the W^- produces an electron and an antineutrino. This process is similar to the scattering event

Equation:

$$e^- + p \rightarrow n + \nu_e,$$

In this process, the proton emits a W^+ and is converted into a neutron (b). The W^+ then combines with the electron, forming a neutrino. Other electroweak interactions are considered in the exercises.



Feynman diagram of particles interacting through the exchange of a W boson: (a) beta decay; (b) conversion of a proton into a neutron.

The range of the weak nuclear force can be estimated with an argument similar to the one before. Assuming the uncertainty on the energy is comparable to the energy of the exchange particle by ($E \approx mc^2$), we have

Equation:

$$\Delta t \approx \frac{h}{mc^2}.$$

The maximum distance d that the exchange particle can travel (assuming it moves at a speed close to c) is therefore

Equation:

$$d \approx c\Delta t = \frac{h}{mc}.$$

For one of the charged vector bosons with $mc^2 \approx 80 \text{ GeV} = 1.28 \times 10^{-8} \text{ J}$, we obtain $mc = 4.27 \times 10^{-17} \text{ J} \cdot \text{s/m}$. Hence, the range of the force mediated by this boson is

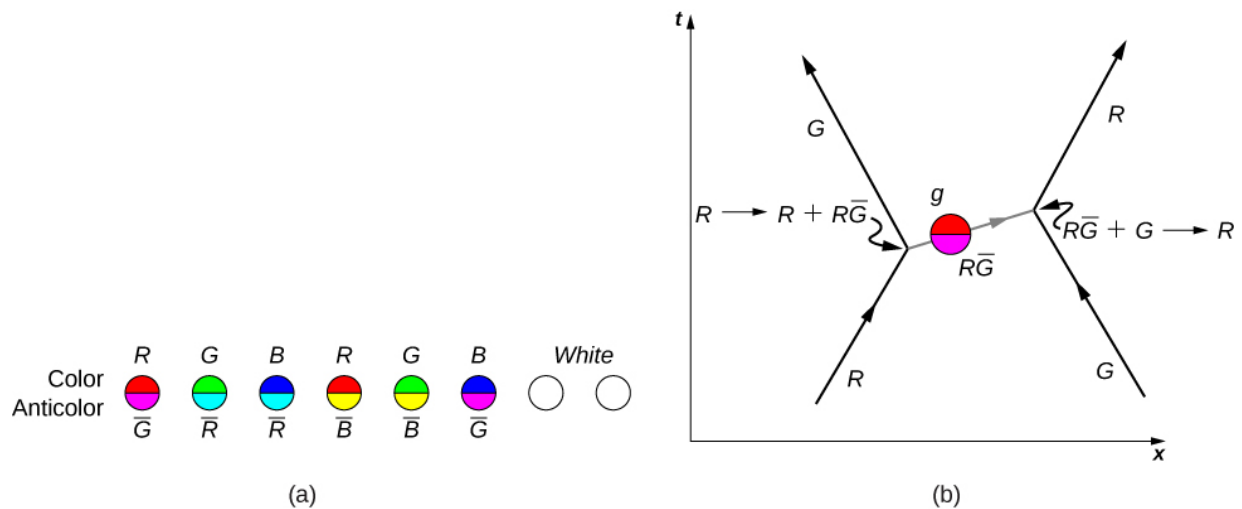
Equation:

$$d \approx \frac{1.05 \times 10^{-34} \text{ J} \cdot \text{s}}{4.27 \times 10^{-17} \text{ J} \cdot \text{s/m}} \approx 2 \times 10^{-18} \text{ m}.$$

Strong Nuclear Force

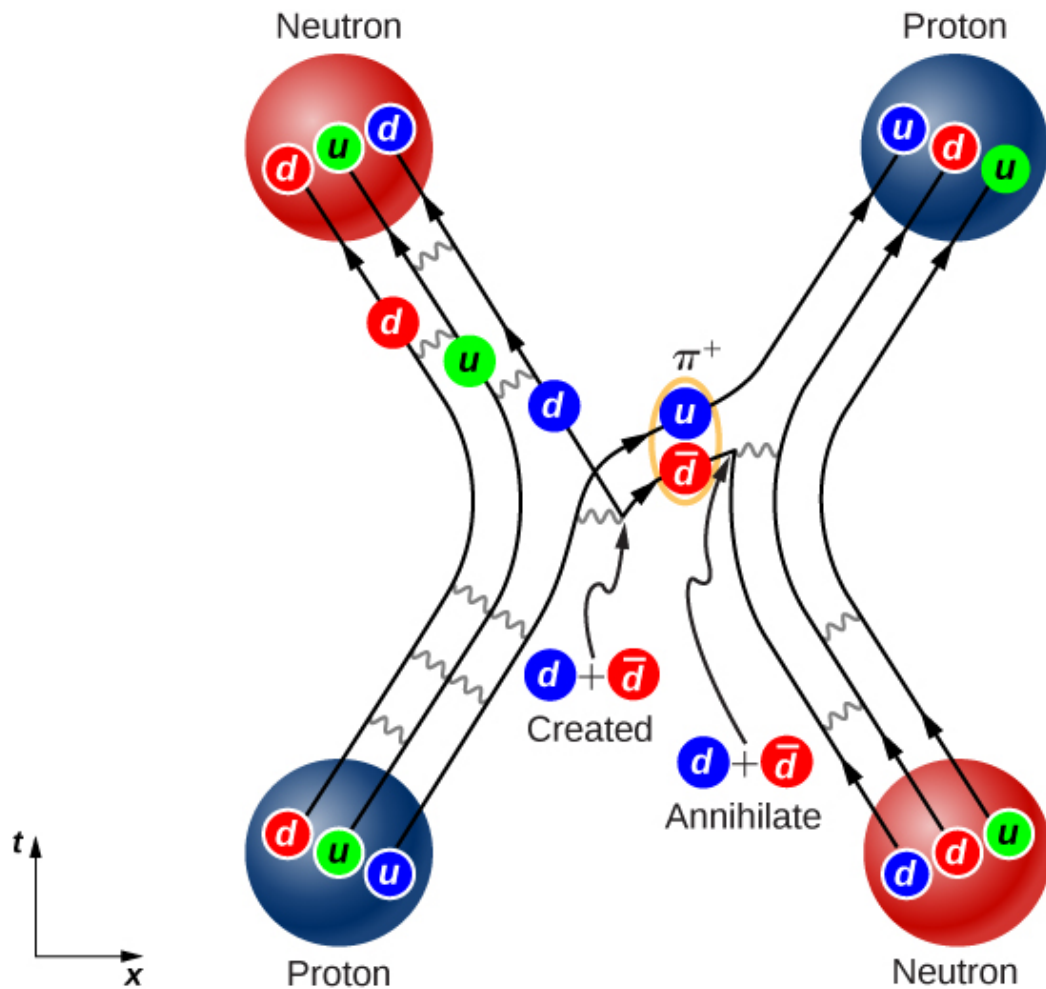
Strong nuclear interactions describe interactions between quarks. Details of these interactions are described by QCD. According to this theory, quarks bind together by sending and receiving gluons. Just as quarks carry electric charge [either $(+2/3)e$ or $(-1/3)e$] that determines the strength of electromagnetic interactions between the quarks, quarks also carry “color charge” (either red, blue, or green) that determines the strength of strong nuclear interactions. As discussed before, quarks bind together in groups in color neutral (or “white”) combinations, such as red-blue-green and red-antired.

Interestingly, the gluons themselves carry color charge. Eight known gluons exist: six that carry a color and anticolor, and two that are color neutral ([\[link\]](#) (a)). To illustrate the interaction between quarks through the exchange of charged gluons, consider the Feynman diagram in part (b). As time increases, a red down quark moves right and a green strange quark moves left. (These appear at the lower edge of the graph.) The up quark exchanges a red-antigreen gluon with the strange quark. (Anticolors are shown as secondary colors. For example, antired is represented by cyan because cyan mixes with red to form white light.) According to QCD, all interactions in this process—identified with the vertices—must be color neutral. Therefore, the down quark transforms from red to green, and the strange quark transforms from green to red.



(a) Eight types of gluons carry the strong nuclear force. The white gluons are mixtures of color-anticolor pairs. (b) An interaction between two quarks through the exchange of a gluon.

As suggested by this example, the interaction between quarks in an atomic nucleus can be very complicated. [\[link\]](#) shows the interaction between a proton and neutron. Notice that the proton converts into a neutron and the neutron converts into a proton during the interaction. The presence of quark-antiquark pairs in the exchange suggest that bonding between nucleons can be modeled as an exchange of pions.



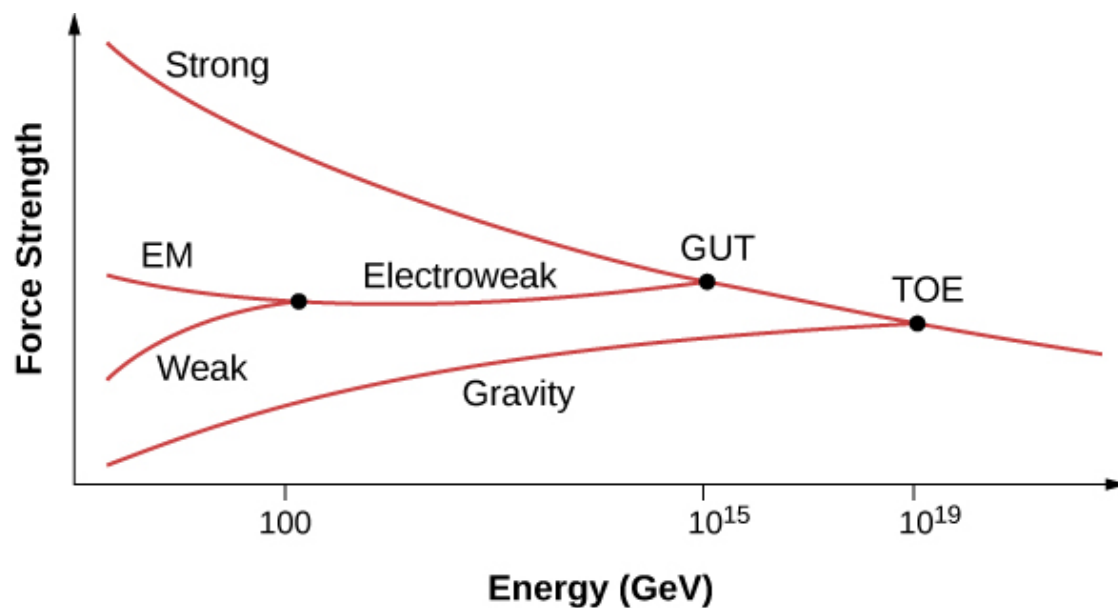
A Feynman diagram that describes a strong nuclear interaction between a proton and a neutron.

In practice, QCD predictions are difficult to produce. This difficulty arises from the inherent strength of the force and the inability to neglect terms in the equations. Thus, QCD calculations are often performed with the aid of supercomputers. The existence of gluons is supported by electron-nucleon scattering experiments. The estimated quark momenta implied by these scattering events are much smaller than we would expect without gluons because the gluons carry away some of the momentum of each collision.

Unification Theories

Physicists have long known that the strength of an interaction between particles depends on the distance of the interaction. For example, two positively charged particles experience a larger repulsive force at a short distance than at a long distance. In scattering experiments, the strength of an interaction depends on the energy of the interacting particle, since larger energy implies both closer and stronger interactions.

Particle physicists now suspect that the strength of all particle interactions (the four forces) merge at high energies, and the details of particle interactions at these energies can be described in terms of a single force ([\[link\]](#)). A unified theory describes what these interactions are like and explains why this description breaks down at low-energy scales. A grand unified theory is a theory that attempts to describe strong and electroweak interaction in terms of just one force. A theory of everything (TOE) takes the unification concept one step further. A TOE combines all four fundamental forces (including gravity) into one theory.



Grand unification of forces at high energies.

Summary

- The Standard Model describes interactions between particles through the strong nuclear, electromagnetic, and weak nuclear forces.
- Particle interactions are represented by Feynman diagrams. A Feynman diagram represents interactions between particles on a space-time graph.
- Electromagnetic forces act over a long range, but strong and weak forces act over a short range. These forces are transmitted between particles by sending and receiving bosons.
- Grand unified theories seek an understanding of the universe in terms of just one force.

Conceptual Questions

Exercise:

Problem:

What is the Standard Model? Express your answer in terms of the four fundamental forces and exchange particles.

Solution:

The Standard Model is a model of elementary particle interactions. This model contains the electroweak theory and quantum chromodynamics (QCD). It describes the interaction of leptons and quarks through the exchange of photons (electromagnetism) and bosons (weak theory), and the interaction of quark through the exchange of gluons (QCD). This model does not describe gravitational interactions.

Exercise:

Problem:

Draw a Feynman diagram to represent annihilation of an electron and positron into a photon.

Exercise:

Problem: What is the motivation behind grand unification theories?

Solution:

To explain particle interactions that involve the strong nuclear, electromagnetic, and weak nuclear forces in a unified way.

Exercise:

Problem:

If a theory is developed that unifies all four forces, will it still be correct to say that the orbit of the Moon is determined by the gravitational force? Explain why.

Exercise:

Problem:

If the Higgs boson is discovered and found to have mass, will it be considered the ultimate carrier of the weak force? Explain your response.

Solution:

No, however it will explain why the W and Z bosons are massive (since the Higgs “imparts” mass to these particles), and therefore why the weak force is short ranged.

Exercise:

Problem:

One of the common decay modes of the Λ^0 is $\Lambda^0 \rightarrow \pi^- + p$. Even though only hadrons are involved in this decay, it occurs through the weak nuclear force. How do we know that this decay does not occur through the strong nuclear force?

Problems

Exercise:

Problem:

Using the Heisenberg uncertainty principle, determine the range of the weak force if this force is produced by the exchange of a Z boson.

Exercise:

Problem:

Use the Heisenberg uncertainty principle to estimate the range of a weak nuclear decay involving a graviton.

Solution:

The graviton is massless, so like the photon is associated with a force of infinite range.

Exercise:

Problem: (a) The following decay is mediated by the electroweak force:

$$p \rightarrow n + e^+ + \nu_e.$$

Draw the Feynman diagram for the decay.

(b) The following scattering is mediated by the electroweak force:

$$\nu_e + e^- \rightarrow \nu_e + e^-.$$

Draw the Feynman diagram for the scattering.

Exercise:**Problem:**

Assuming conservation of momentum, what is the energy of each γ ray produced in the decay of a neutral pion at rest, in the reaction $\pi^0 \rightarrow \gamma + \gamma$?

Solution:

67.5 MeV

Exercise:**Problem:**

What is the wavelength of a 50-GeV electron, which is produced at SLAC? This provides an idea of the limit to the detail it can probe.

Exercise:**Problem:**

The primary decay mode for the negative pion is $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$. (a) What is the energy release in MeV in this decay? (b) Using conservation of momentum, how much energy does each of the decay products receive, given the π^- is at rest when it decays? You may assume the muon antineutrino is massless and has momentum $p = E/c$, just like a photon.

Solution:

a. 33.9 MeV; b. By conservation of momentum, $|p_\mu| = |p_\nu| = p$. By conservation of energy, $E_\nu = 29.8$ MeV, $E_\mu = 4.1$ MeV

Exercise:**Problem:**

Suppose you are designing a proton decay experiment and you can detect 50 percent of the proton decays in a tank of water. (a) How many kilograms of water would you need to see one decay per month, assuming a lifetime of 10^{31} y? (b) How many cubic meters of water is this? (c) If the actual lifetime is 10^{33} y, how long would you have to wait on an average to see a single proton decay?

Glossary

Feynman diagram

space-time diagram that describes how particles move and interact

quantum chromodynamics (QCD)

theory that describes strong interactions between quarks

quantum electrodynamics (QED)

theory that describes the interaction of electrons with photons

Standard Model

model of particle interactions that contains the electroweak theory and quantum chromodynamics (QCD)

virtual particle

particle that exists for too short of time to be observable

The Big Bang

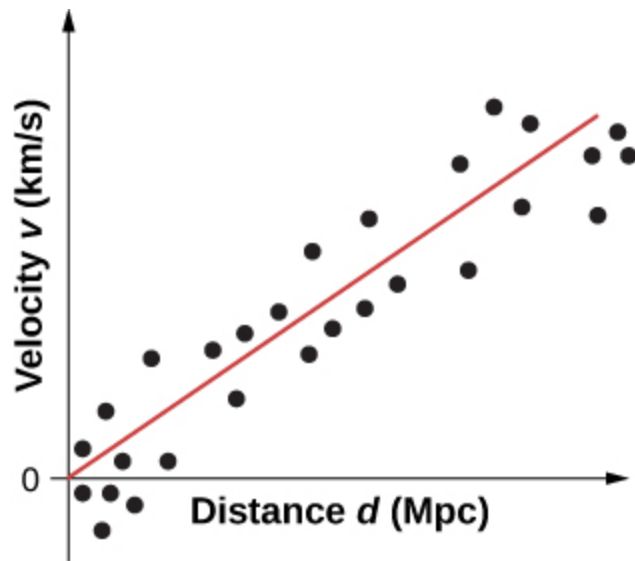
By the end of this section, you will be able to:

- Explain the expansion of the universe in terms of a Hubble graph and cosmological redshift
- Describe the analogy between cosmological expansion and an expanding balloon
- Use Hubble's law to make predictions about the measured speed of distant galaxies

We have been discussing elementary particles, which are some of the smallest things we can study. Now we are going to examine what we know about the universe, which is the biggest thing we can study. The link between these two topics is high energy: The study of particle interactions requires very high energies, and the highest energies we know about existed during the early evolution of the universe. Some physicists think that the unified force theories we described in the preceding section may actually have governed the behavior of the universe in its earliest moments.

Hubble's Law

In 1929, Edwin Hubble published one of the most important discoveries in modern astronomy. Hubble discovered that (1) galaxies appear to move away from Earth and (2) the velocity of recession (v) is proportional to the distance (d) of the galaxy from Earth. Both v and d can be determined using stellar light spectra. A best fit to the sample illustrative data is given in [\[link\]](#). (Hubble's original plot had a considerable scatter but a general trend was still evident.)



This graph of red shift versus distance for galaxies shows a linear relationship, with larger red shifts at greater distances, implying an expanding universe. The slope gives an approximate value for the expansion rate.
(credit: John Cub)

The trend in the data suggests the simple proportional relationship:

Note:

Equation:

$$v = H_0 d,$$

where $H_0 = 70 \text{ km / s / Mpc}$ is known as **Hubble's constant**. (Note: 1 Mpc is one megaparsec or one million parsecs, where one parsec is 3.26

light-years.) This relationship, called **Hubble's law**, states that distant stars and galaxies recede away from us at a speed of 70 km/s for every one megaparsec of distance from us. Hubble's constant corresponds to the slope of the line in [\[link\]](#). Hubble's constant is a bit of a misnomer, because it varies with time. The value given here is only its value *today*.

Note:

Watch this [video](#) to learn more about the history of Hubble's constant.

Hubble's law describes an average behavior of all but the closest galaxies. For example, a galaxy 100 Mpc away (as determined by its size and brightness) typically moves away from us at a speed of

Equation:

$$v = \left(\left(70 \frac{\text{km}}{\text{s}} \right) / \text{Mpc} \right) (100 \text{ Mpc}) = 7000 \text{ km/s}.$$

This speed may vary due to interactions with neighboring galaxies. Conversely, if a galaxy is found to be moving away from us at speed of 100,000 km/s based on its red shift, it is at a distance

Equation:

$$d = v/H_0 = (10,000 \text{ km/s}) / \left(\left(70 \frac{\text{km}}{\text{s}} \right) / \text{Mpc} \right) = 143 \text{ Mpc}.$$

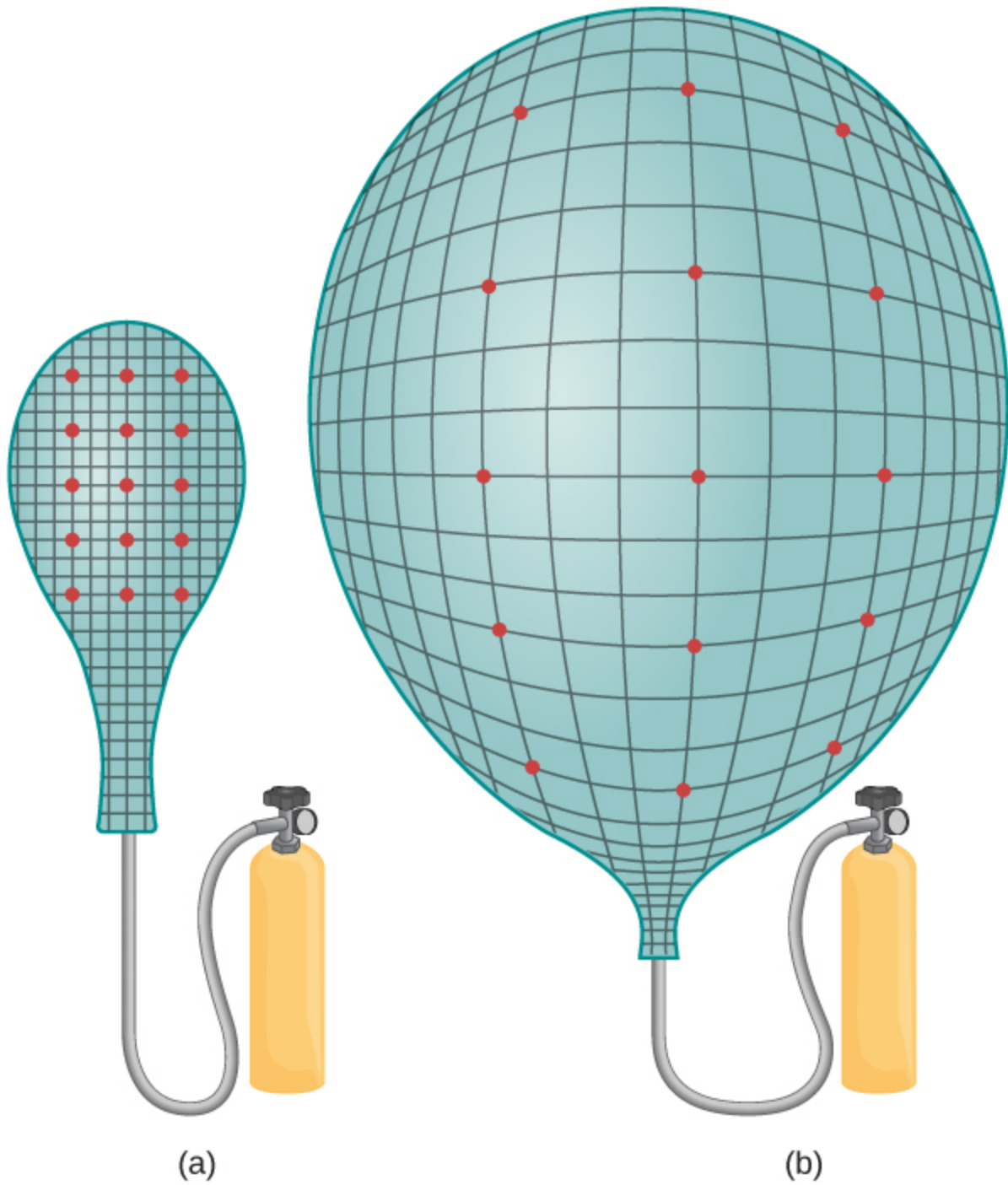
This last calculation is approximate because it assumes the expansion rate was the same 5 billion years ago as it is now.

Big Bang Model

Scientists who study the origin, evolution, and ultimate fate of the universe (**cosmology**) believe that the universe began in an explosion, called the **Big**

Bang, approximately 13.7 billion years ago. This explosion was not an explosion of particles through space, like fireworks, but a rapid expansion of space itself. The distances and velocities of the outward-going stars and galaxies permit us to estimate when all matter in the universe was once together—at the beginning of time.

Scientists often explain the Big Bang expansion using an inflated-balloon model ([\[link\]](#)). Dots marked on the surface of the balloon represent galaxies, and the balloon skin represents four-dimensional space-time ([Relativity](#)). As the balloon is inflated, every dot “sees” the other dots moving away. This model yields two insights. First, the expansion is observed by all observers in the universe, no matter where they are located. The “center of expansion” does not exist, so Earth does not reside at the “privileged” center of the expansion (see [\[link\]](#)).



An analogy to the expanding universe: The dots move away from each other as the balloon expands; compare (a) to (b) after expansion.

Second, as mentioned already, the Big Bang expansion is due to the expansion of space, not the increased separation of galaxies in ordinary (static) three-dimensional space. This cosmological expansion affects all things: dust, stars, planets, and even light. Thus, the wavelength of light (λ) emitted by distant galaxies is “stretched” out. This makes the light appear “redder” (lower energy) to the observer—a phenomenon called cosmological **redshift**. Cosmological redshift is measurable only for galaxies farther away than 50 million light-years.

Example:

Calculating Speeds and Galactic Distances

A galaxy is observed to have a redshift:

Equation:

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{emit}}}{\lambda_{\text{emit}}} = 4.5.$$

This value indicates a galaxy moving close to the speed of light. Using the relativistic redshift formula (given in [Relativity](#)), determine (a) How fast is the galaxy receding with respect to Earth? (b) How far away is the galaxy?

Strategy

We need to use the relativistic Doppler formula to determine speed from redshift and then use Hubble’s law to find the distance from the speed.

Solution

- a. According to the relativistic redshift formula:

Equation:

$$z = \sqrt{\frac{1 + \beta}{1 - \beta}} - 1,$$

where $\beta = v/c$. Substituting the value for z and solving for β , we get $\beta = 0.93$. This value implies that the speed of the galaxy is $2.8 \times 10^8 \text{ m/s}$.

b. Using Hubble's law, we can find the distance to the galaxy if we know its recession velocity:

Equation:

$$d = \frac{v}{H_0} = \frac{2.8 \times 10^8 \text{ m/s}}{73.8 \times 10^3 \text{ m/s per Mpc}} = 3.8 \times 10^3 \text{ Mpc.}$$

Significance

Distant galaxies appear to move very rapidly away from Earth. The redshift of starlight from these galaxies can be used to determine the precise speed of recession, over 90% of the speed of light in this case. This motion is not due to the motion of galaxy through space but by the expansion of space itself.

Note:

Exercise:

Problem:

Check Your Understanding The light of a galaxy that moves away from us is “redshifted.” What occurs to the light of a galaxy that moves toward us?

Solution:

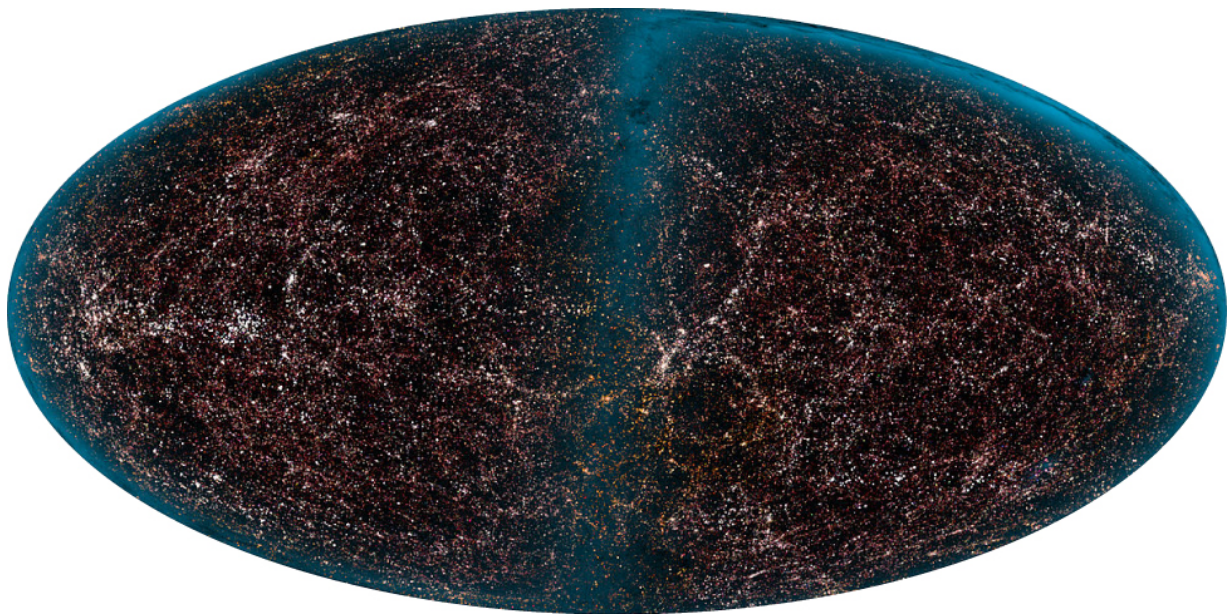
blueshifted

Note:

View this [video](#) to learn more about the cosmological expansion.

Structure and Dynamics of the Universe

At large scales, the universe is believed to be both isotropic and homogeneous. The universe is believed to be isotropic because it appears to be the same in all directions, and homogeneous because it appears to be the same in all places. A universe that is isotropic and homogeneous is said to be smooth. The assumption of a smooth universe is supported by the Automated Plate Measurement Galaxy Survey conducted in the 1980s and 1990s ([link](#)). However, even before these data were collected, the assumption of a smooth universe was used by theorists to simplify models of the expansion of the universe. This assumption of a smooth universe is sometimes called the cosmological principle.



The Automated Plate Measurement (APM) Galaxy Survey. Over 2 million galaxies are depicted in a region 100 degrees across centered toward the Milky Way's south pole. (credit: 2MASS/T. H. Jarrett, J. Carpenter, & R. Hurt)

The fate of this expanding and smooth universe is an open question. According to the general theory of relativity, an important way to characterize the state of the universe is through the space-time metric:

Note:

Equation:

$$ds^2 = c^2 dt^2 - a(t)^2 d\Sigma^2,$$

where c is the speed of light, a is a scale factor (a function of time), and $d\Sigma$ is the length element of the space. In spherical coordinates (r, θ, ϕ) , this length element can be written

Equation:

$$d\Sigma^2 = \frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2),$$

where k is a constant with units of inverse area that describes the curvature of space. This constant distinguishes between open, closed, and flat universes:

- $k = 0$ (flat universe)
- $k > 0$ (closed universe, such as a sphere)
- $k < 0$ (open universe, such as a hyperbola)

In terms of the scale factor a , this metric also distinguishes between static, expanding, and shrinking universes:

- $a = 1$ (static universe)
- $da/dt > 0$ (expanding universe)
- $da/dt < 0$ (shrinking universe)

The scale factor a and the curvature k are determined from Einstein's general theory of relativity. If we treat the universe as a gas of galaxies of density ρ and pressure p , and assume $k = 0$ (a flat universe), then the scale factor a is given by

Equation:

$$\frac{d^2a}{dt^2} = -\frac{4\pi G}{3}(\rho + 3p)a,$$

where G is the universal gravitational constant. (For ordinary matter, we expect the quantity $\rho + 3p$ to be greater than zero.) If the scale factor is positive ($a > 0$), the value of the scale factor “decelerates” ($d^2a/dt^2 < 0$), and the expansion of the universe slows down over time. If the numerator is less than zero (somehow, the pressure of the universe is negative), the value of the scale factor “accelerates,” and the expansion of the universe speeds up over time. According to recent cosmological data, the universe appears to be expanding. Many scientists explain the current state of the universe in terms of a very rapid expansion in the early universe. This expansion is called inflation.

Summary

- The universe is expanding like a balloon—every point is receding from every other point.
- Distant galaxies move away from us at a velocity proportional to its distance. This rate is measured to be approximately 70 km/s/Mpc. Thus, the farther galaxies are from us, the greater their speeds. These “recessional velocities” can be measure using the Doppler shift of light.
- According to current cosmological models, the universe began with the Big Bang approximately 13.7 billion years ago.

Conceptual Questions

Exercise:

Problem:

What is meant by cosmological expansion? Express your answer in terms of a Hubble graph and the red shift of distant starlight.

Solution:

Cosmological expansion is an expansion of space. This expansion is different than the explosion of a bomb where particles pass rapidly *through* space. A plot of the recessional speed of a galaxy is proportional to its distance. This speed is measured using the red shift of distant starlight.

Exercise:**Problem:**

Describe the balloon analogy for cosmological expansion. Explain why it only *appears* that we are at the center of expansion of the universe.

Exercise:**Problem:**

Distances to local galaxies are determined by measuring the brightness of stars, called Cepheid variables, that can be observed individually and that have absolute brightnesses at a standard distance that are well known. Explain how the measured brightness would vary with distance, as compared with the absolute brightness.

Solution:

With distance, the absolute brightness is the same, but the apparent brightness is inversely proportional to the square of its distance (or by Hubble's law recessional velocity).

Problems

Exercise:**Problem:**

If the speed of a distant galaxy is $0.99c$, what is the distance of the galaxy from an Earth-bound observer?

Solution:

$$(0.99)(299792 \text{ km/s}) = \left((70 \frac{\text{km}}{\text{s}}) / \text{Mpc} \right) (d), \quad d = 4240 \text{ Mpc}$$

Exercise:**Problem:**

The distance of a galaxy from our solar system is 10 Mpc. (a) What is the recessional velocity of the galaxy? (b) By what fraction is the starlight from this galaxy redshifted (that is, what is its z value)?

Exercise:**Problem:**

If a galaxy is 153 Mpc away from us, how fast do we expect it to be moving and in what direction?

Solution:

$$1.0 \times 10^4 \text{ km/s away from us.}$$

Exercise:**Problem:**

On average, how far away are galaxies that are moving away from us at 2.0% of the speed of light?

Exercise:

Problem:

Our solar system orbits the center of the Milky Way Galaxy. Assuming a circular orbit 30,000 ly in radius and an orbital speed of 250 km/s, how many years does it take for one revolution? Note that this is approximate, assuming constant speed and circular orbit, but it is representative of the time for our system and local stars to make one revolution around the galaxy.

Solution:

$$2.26 \times 10^8 \text{ y}$$

Exercise:**Problem:**

(a) What is the approximate velocity relative to us of a galaxy near the edge of the known universe, some 10 Gly away? (b) What fraction of the speed of light is this? Note that we have observed galaxies moving away from us at greater than $0.9c$.

Exercise:**Problem:**

(a) Calculate the approximate age of the universe from the average value of the Hubble constant, $H_0 = 20 \text{ km/s} \cdot \text{Mly}$. To do this, calculate the time it would take to travel 0.307 Mpc at a constant expansion rate of 20 km/s. (b) If somehow acceleration occurs, would the actual age of the universe be greater or less than that found here? Explain.

Solution:

a. $1.5 \times 10^{10} \text{ y} = 15 \text{ billion years}$; b. Greater, since if it was moving slower in the past it would take less more to travel the distance.

Exercise:

Problem:

The Andromeda Galaxy is the closest large galaxy and is visible to the naked eye. Estimate its brightness relative to the Sun, assuming it has luminosity 10^{12} times that of the Sun and lies 0.613 Mpc away.

Exercise:**Problem:**

Show that the velocity of a star orbiting its galaxy in a circular orbit is inversely proportional to the square root of its orbital radius, assuming the mass of the stars inside its orbit acts like a single mass at the center of the galaxy. You may use an equation from a previous chapter to support your conclusion, but you must justify its use and define all terms used.

Solution:

$$v = \sqrt{\frac{GM}{r}}$$

Glossary**Big Bang**

rapid expansion of space that marked the beginning of the universe

cosmology

study of the origin, evolution, and ultimate fate of the universe

Hubble's constant

constant that relates speed and distance in Hubble's law

Hubble's law

relationship between the speed and distance of stars and galaxies

redshift

lengthening of the wavelength of light (or reddening) due to cosmological expansion

Evolution of the Early Universe

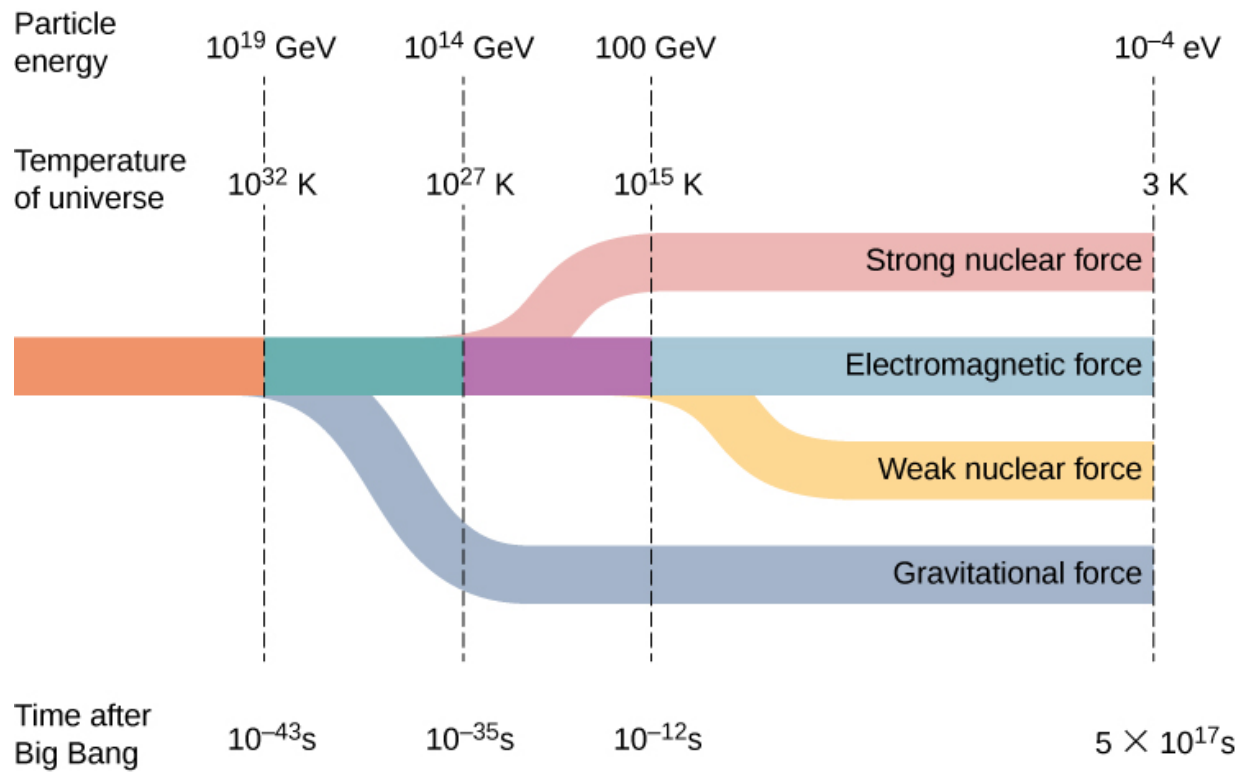
By the end of this section, you will be able to:

- Describe the evolution of the early universe in terms of the four fundamental forces
- Use the concept of gravitational lensing to explain astronomical phenomena
- Provide evidence of the Big Bang in terms of cosmic background radiation
- Distinguish between dark matter and dark energy

In the previous section, we discussed the structure and dynamics of universe. In particular, the universe appears to be expanding and even accelerating. But what was the universe like at the beginning of time? In this section, we discuss what evidence scientists have been able to gather about the early universe and its evolution to present time.

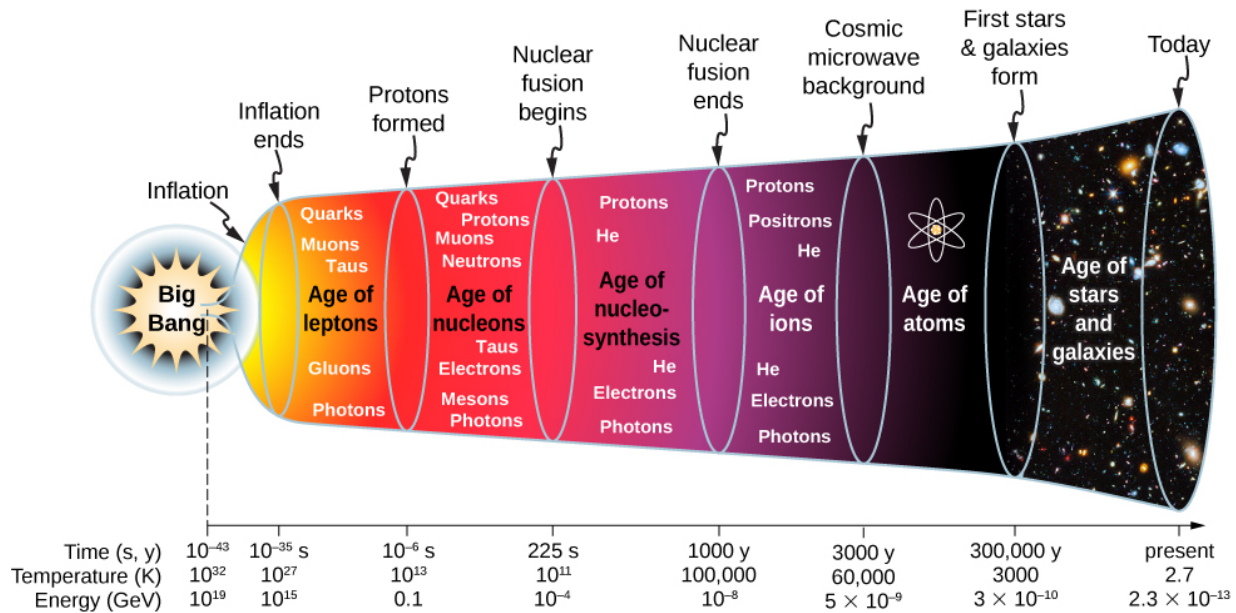
The Early Universe

Before the short period of cosmic inflation, cosmologists believe that all matter in the universe was squeezed into a space much smaller than an atom. Cosmologists further believe that the universe was extremely dense and hot, and interactions between particles were governed by a single force. In other words, the four fundamental forces (strong nuclear, electromagnetic, weak nuclear, and gravitational) merge into one at these energies ([\[link\]](#)). How and why this “unity” breaks down at lower energies is an important unsolved problem in physics.



The separation of the four fundamental forces in the early universe.

Scientific models of the early universe are highly speculative. [\[link\]](#) shows a sketch of one possible timeline of events.



An approximate timeline for the evolution of the universe from the Big Bang to the present.

1. *Big Bang* ($t < 10^{-43}$ s) : The current laws of physics break down. At the end of the initial Big Bang event, the temperature of the universe is approximately $T = 10^{32}$ K.
2. *Inflationary phase* ($t = 10^{-43}$ to 10^{-35} s) : The universe expands exponentially, and gravity separates from the other forces. The universe cools to approximately $T = 10^{27}$ K.
3. *Age of leptons* ($t = 10^{-35}$ to 10^{-6} s) : As the universe continues to expand, the strong nuclear force separates from the electromagnetic and weak nuclear forces (or electroweak force). Soon after, the weak nuclear force separates from the electromagnetic force. The universe is a hot soup of quarks, leptons, photons, and other particles.
4. *Age of nucleons* ($t = 10^{-6}$ to 225 s) : The universe consists of leptons and hadrons (such as protons, neutrons, and mesons) in thermal equilibrium. Pair production and pair annihilation occurs with equal ease, so photons remain in thermal equilibrium:

Equation:

$$\gamma + \gamma \leftrightarrow e^- + e^+$$

$$\gamma + \gamma \leftrightarrow p + \bar{p}$$

$$\gamma + \gamma \leftrightarrow n + \bar{n}.$$

The number of protons is approximately equal to the number of neutrons through interactions with neutrinos:

Equation:

$$\nu_e + n \leftrightarrow e^- + p$$

$$\bar{\nu}_e + p \leftrightarrow e^+ + n.$$

The temperature of the universe settles to approximately 10^{11}K —much too cool for the continued production of nucleon-antinucleon pairs. The numbers of protons and neutrons begin to dominate over their anti-particles, so proton-antiproton ($p\bar{p}$) and neutron-antineutron ($n\bar{n}$)

annihilations decline. Deuterons (proton-neutron pairs) begin to form.

5. *Age of nucleosynthesis* ($t = 225\text{s}$ to 1000 years): As the universe continues to expand, deuterons react with protons and neutrons to form larger nuclei; these larger nuclei react with protons and neutrons to form still larger nuclei. At the end of this period, about $1/4$ of the mass of the universe is helium. (This explains the current amount of helium in the universe.) Photons lack the energy to continue electron-positron production, so electrons and positrons annihilate each other to photons only.
6. *Age of ions* ($t = 1000$ to 3000 years): The universe is hot enough to ionize any atoms formed. The universe consists of electrons, positrons, protons, light nuclei, and photons.
7. *Age of atoms* ($t = 3000$ to $300,000$ years): The universe cools below 10^5K and atoms form. Photons do not interact strongly with neutral atoms, so they “decouple” (separate) from atoms. These photons constitute the **cosmic microwave background radiation** to be discussed later.
8. *Age of stars and galaxies* ($t = 300,000$ years to present): The atoms and particles are pulled together by gravity and form large lumps. The atoms and particles in stars undergo nuclear fusion reaction.

Note:

Watch this [video](#) to learn more about Big Bang cosmology.

To describe the conditions of the early universe quantitatively, recall the relationship between the average thermal energy of particle (E) in a system of interacting particles and equilibrium temperature (T) of that system:

Equation:

$$E = k_B T,$$

where k_B is Boltzmann's constant. In the hot conditions of the early universe, particle energies were unimaginably large.

Example:**What Was the Average Thermal Energy of a Particle just after the Big Bang?****Strategy**

The average thermal energy of a particle in a system of interacting particles depends on the equilibrium temperature of that system [\[link\]](#). We are given this approximate temperature in the above timeline.

Solution

Cosmologists think the temperature of the universe just after the Big Bang was approximately $T = 10^{32}\text{K}$. Therefore, the average thermal energy of a particle would have been

Equation:

$$k_B T \approx (10^{-4} \text{ eV/K})(10^{32}\text{K}) = 10^{28} \text{ eV} = 10^{19} \text{ GeV}.$$

Significance

This energy is many orders of magnitude larger than particle energies produced by human-made particle accelerators. Currently, these accelerators operate at energies less than 10^4 GeV .

Note:

Exercise:

Problem:

Check Your Understanding Compare the abundance of helium by mass 10,000 years after the Big Bang and now.

Solution:

about the same

Nucleons form at energies approximately equal to the rest mass of a proton, or 1000 MeV. The temperature corresponding to this energy is therefore

Equation:

$$T = \frac{1000 \text{ MeV}}{8.62 \times 10^{11} \text{ MeV} \cdot \text{K}^{-1}} = 1.2 \times 10^{13} \text{ K}.$$

Temperatures of this value or higher existed within the first second of the early universe. A similar analysis can be done for atoms. Atoms form at an energy equal to the ionization energy of ground-state hydrogen (13 eV). The effective temperature for atom formation is therefore

Equation:

$$T = \frac{13 \text{ eV}}{8.62 \times 10^5 \text{ eV} \cdot \text{K}^{-1}} = 1.6 \times 10^5 \text{ K}.$$

This occurs well after the four fundamental forces have separated, including forces necessary to bind the protons and neutrons in the nucleus (strong nuclear force), and bind electrons to the nucleus (electromagnetic force).

Nucleosynthesis of Light Elements

The relative abundances of the light elements hydrogen, helium, lithium, and beryllium in the universe provide key evidence for the Big Bang. The data suggest that much of the helium in the universe is primordial. For instance, it turns out that that 25% of the matter in the universe is helium, which is too high an abundance and cannot be explained based on the production of helium in stars.

How much of the elements in the universe were created in the Big Bang? If you run the clock backward, the universe becomes more and more compressed, and hotter and hotter. Eventually, temperatures are reached that permit **nucleosynthesis**, the period of formation of nuclei, similar to what occurs at the core of the Sun. Big Bang nucleosynthesis is believed to have occurred within a few hundred seconds of the Big Bang.

How did Big Bang nucleosynthesis occur? At first, protons and neutrons combined to form deuterons, ${}^2\text{H}$. The deuteron captured a neutron to form triton, ${}^3\text{H}$ —the nucleus of the radioactive hydrogen called tritium. Deuterons also captured protons to make helium ${}^3\text{He}$. When ${}^3\text{H}$ captures a proton or ${}^3\text{He}$ captures a neutron, helium ${}^4\text{He}$ results. At this stage in the Big Bang, the ratio of protons to neutrons was about 7:1. Thus, the process of conversion to ${}^4\text{He}$ used up almost all neutrons. The process lasted about 3 minutes and almost 25% of all the matter turned into ${}^4\text{He}$, along with small percentages of ${}^2\text{H}$, ${}^3\text{H}$, and ${}^3\text{He}$. Tiny amounts of ${}^7\text{Li}$ and ${}^7\text{Be}$ were also formed. The expansion during this time cooled the universe enough that the nuclear reactions stopped. The abundances of the light nuclei ${}^2\text{H}$, ${}^4\text{He}$, and ${}^7\text{Li}$ created after the Big Bang are very dependent on the matter density.

The predicted abundances of the elements in the universe provide a stringent test of the Big Bang and the Big Bang nucleosynthesis. Recent experimental estimates of the matter density from the Wilkinson Microwave Anisotropy Probe (WMAP) agree with model predictions. This agreement provides convincing evidence of the Big Bang model.

Cosmic Microwave Background Radiation

According to cosmological models, the Big Bang event should have left behind thermal radiation called the cosmic microwave background radiation (CMBR). The intensity of this radiation should follow the blackbody radiation

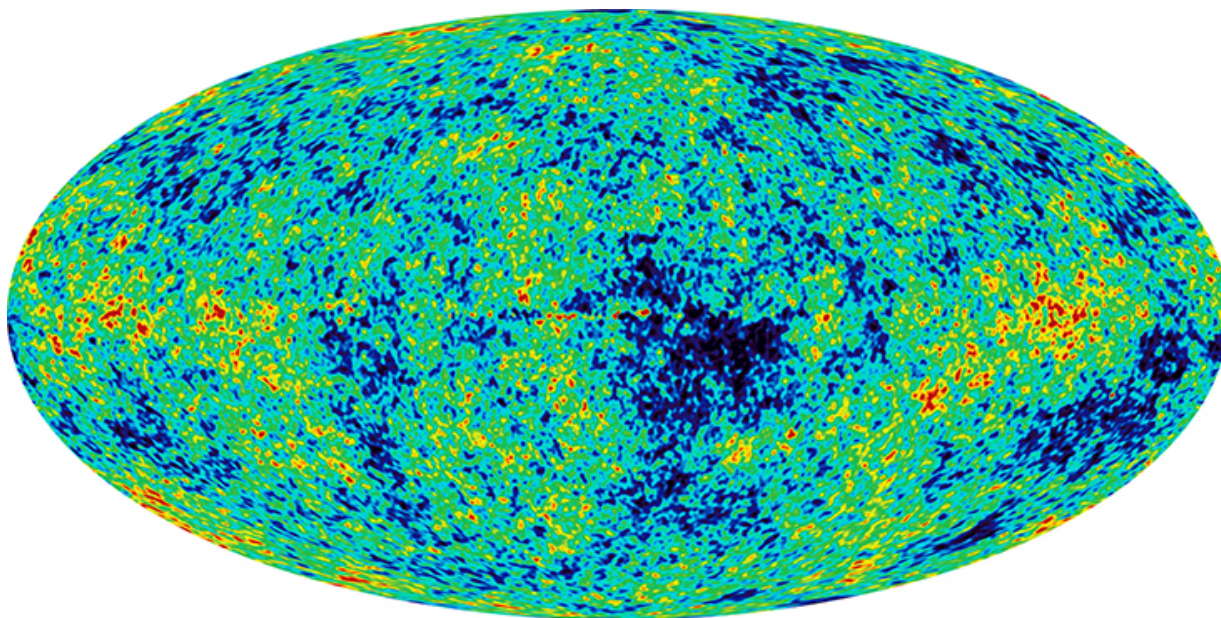
curve ([Photons and Matter Waves](#)). Wien's law states that the wavelength of the radiation at peak intensity is

Equation:

$$\lambda_{\text{max}} = \frac{2.898 \times 10^{-3} \text{ m-K}}{T},$$

where T is temperature in kelvins. Scientists expected the expansion of the universe to “stretch the light,” and the temperature to be very low, so cosmic background radiation should be long-wavelength and low energy.

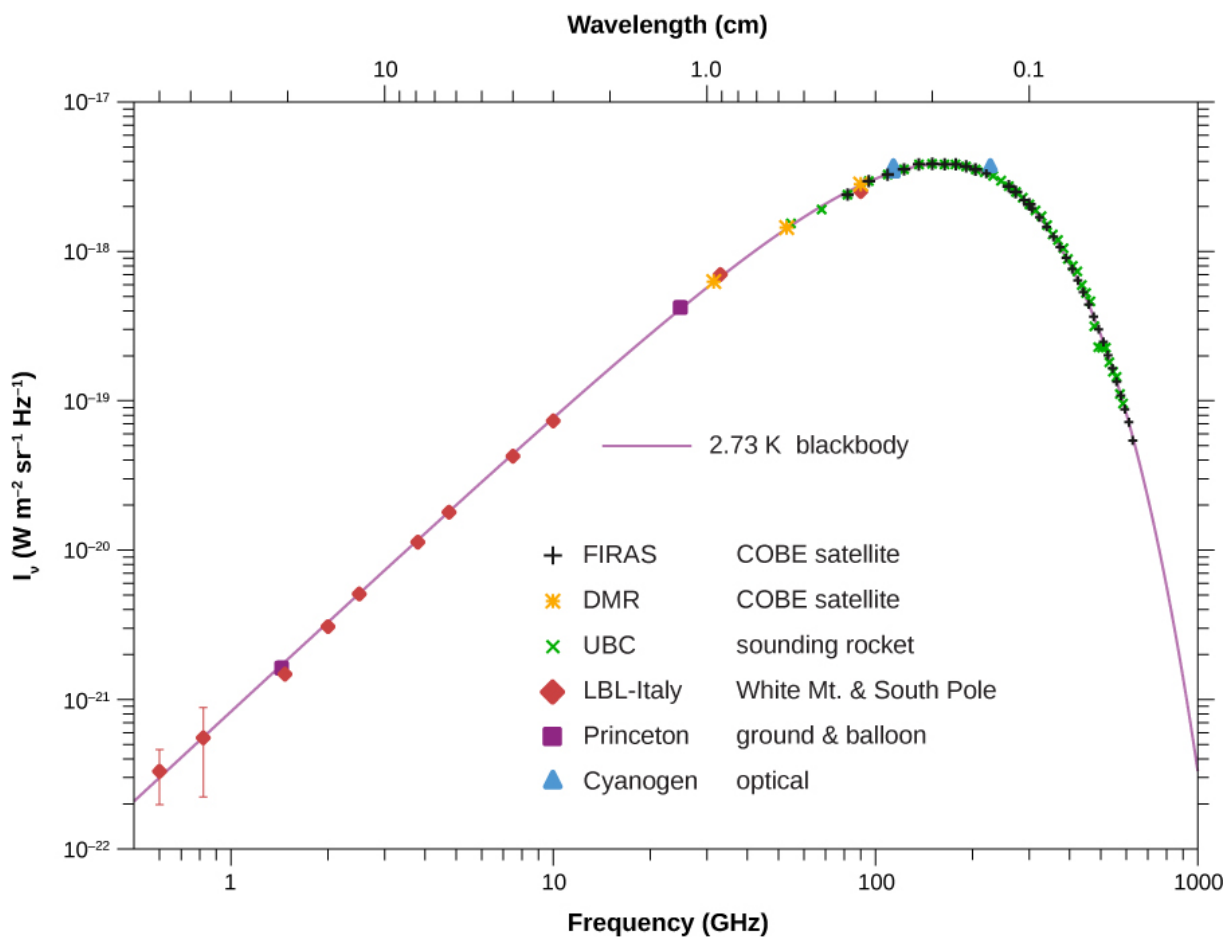
In the 1960s, Arno Penzias and Robert Wilson of Bell Laboratories noticed that no matter what they did, they could not get rid of a faint background noise in their satellite communication system. The noise was due to radiation with wavelengths in the centimeter range (the microwave region). Later, this noise was associated with the cosmic background radiation. An intensity map of the cosmic background radiation appears in [\[link\]](#). The thermal spectrum is modeled well by a blackbody curve that corresponds to a temperature $T = 2.7\text{K}$ ([\[link\]](#)).



This map of the sky uses color to show fluctuations, or wrinkles, in the

cosmic microwave background observed with the WMAP spacecraft.

The Milky Way has been removed for clarity. Red represents higher temperature and higher density, whereas blue indicates lower temperature and density. This map does not contradict the earlier claim of smoothness because the largest fluctuations are only one part in one million. (credit: NASA/WMAP Science Team)



Intensity distribution of cosmic microwave background radiation. The model predictions (the line) agree extremely well with the experimental results (the dots). Frequency and brightness values are shown on a log axis. (credit: modification of work by George Smoot/NASA COBE Project)

The formation of atoms in the early universe makes these atoms less likely to interact with light. Therefore, photons that belong to the cosmic background radiation must have separated from matter at a temperature T associated with 1 eV (the approximate ionization energy of an atom) . The temperature of the universe at this point was

Equation:

$$k_B T \sim 1 \text{ eV} \Rightarrow T = \frac{1 \text{ eV}}{8.617 \times 10^5 \text{ eV/K}} \sim 10^4 \text{ K}.$$

According to cosmological models, the time when photons last scattered off charged particles was approximately 380,000 years after the Big Bang. Before that time, matter in the universe was in the plasma form and the photons were “thermalized.”

Antimatter and Matter

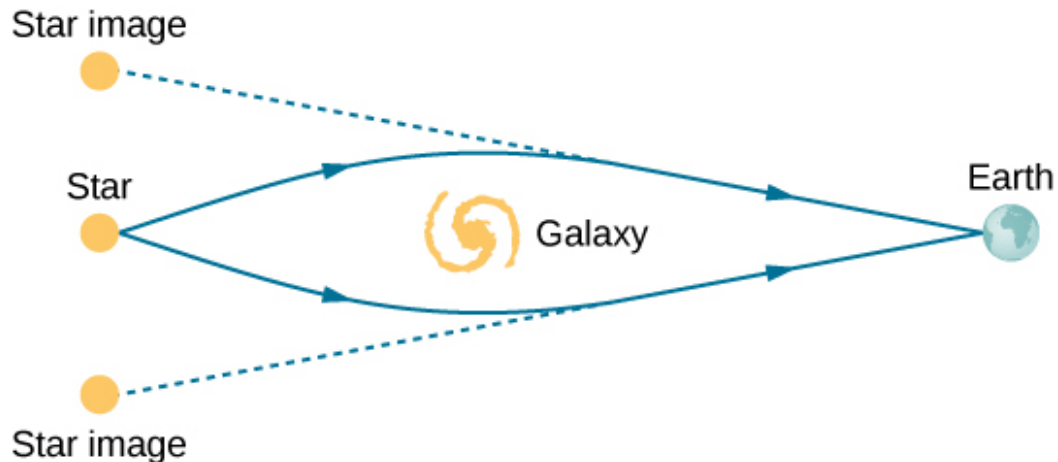
We know from direct observation that antimatter is rare. Earth and the solar system are nearly pure matter, and most of the universe also seems dominated by matter. This is proven by the lack of annihilation radiation coming to us from space, particularly the relative absence of 0.511-MeV γ rays created by the mutual annihilation of electrons and positrons. (Antimatter in nature is created in particle collisions and in β^+ decays, but only in small amounts that quickly annihilate, leaving almost pure matter surviving.)

Despite the observed dominance of matter over antimatter in the universe, the Standard Model of particle interactions and experimental measurement suggests only small differences in the ways that matter and antimatter interact. For example, neutral kaon decays produce only slightly more matter than antimatter. Yet, if through such decay, slightly more matter than antimatter was produced in the early universe, the rest could annihilate pair by pair, leaving mostly ordinary matter to form the stars and galaxies. In this way, the vast number of stars we observe may be only a tiny remnant of the original matter created in the Big Bang.

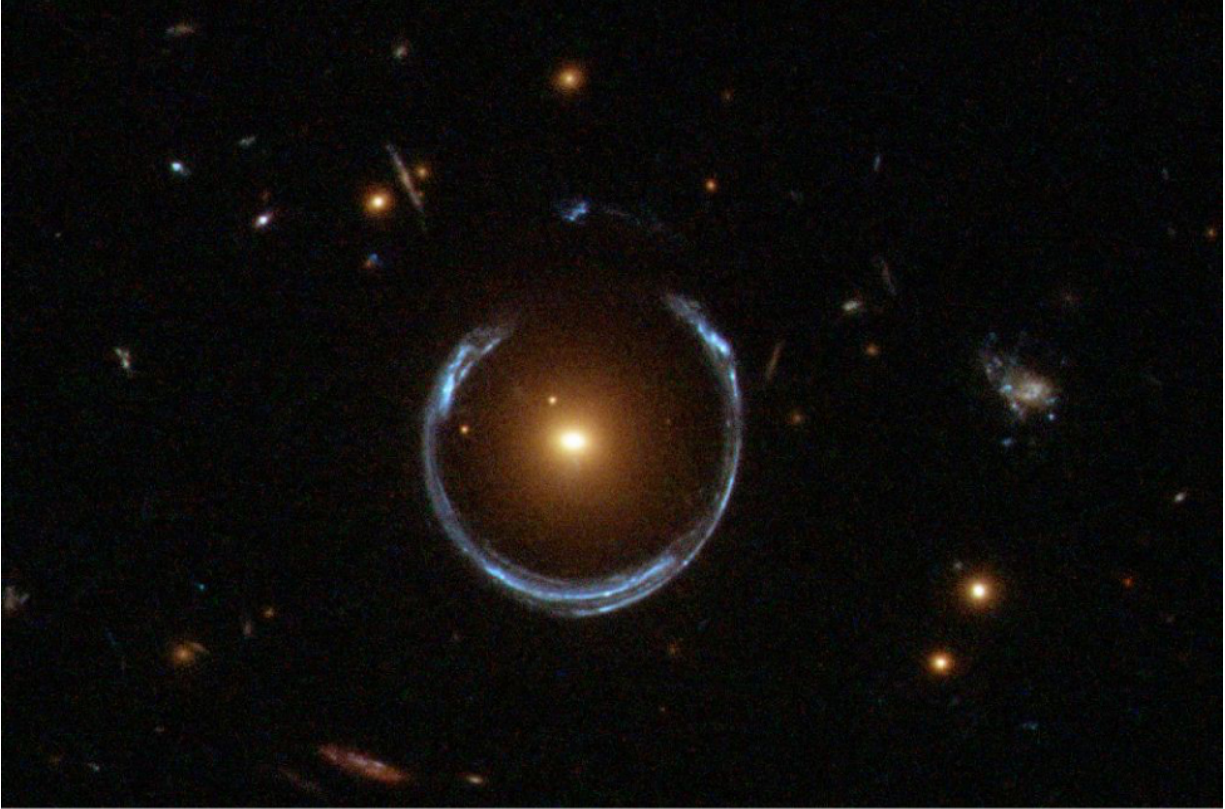
Dark Matter and Dark Energy

In the last two decades, new and more powerful techniques have revealed that the universe is filled with **dark matter**. This type of matter is interesting and important because, currently, scientists do not know what it is! However, we can infer its existence by the deflection of distant starlight. For example, if light from a distant galaxy is bent by the gravitational field of a clump of dark matter between us and the galaxy, it is possible that two images of the same galaxy can be produced ([\[link\]](#)). The bending of light by the gravitational field of matter is called gravitational lensing. In some cases, the starlight travels to an observer by multiple paths around the galaxy, producing a ring ([\[link\]](#)).

Based on current research, scientists know only that dark matter is cold, slow moving, and interacts weakly with ordinary matter. Dark matter candidates include neutralinos (partners of Z bosons, photons, and Higgs bosons in “supersymmetry theory”) and particles that circulate in tiny rings set up by extra spatial dimensions.



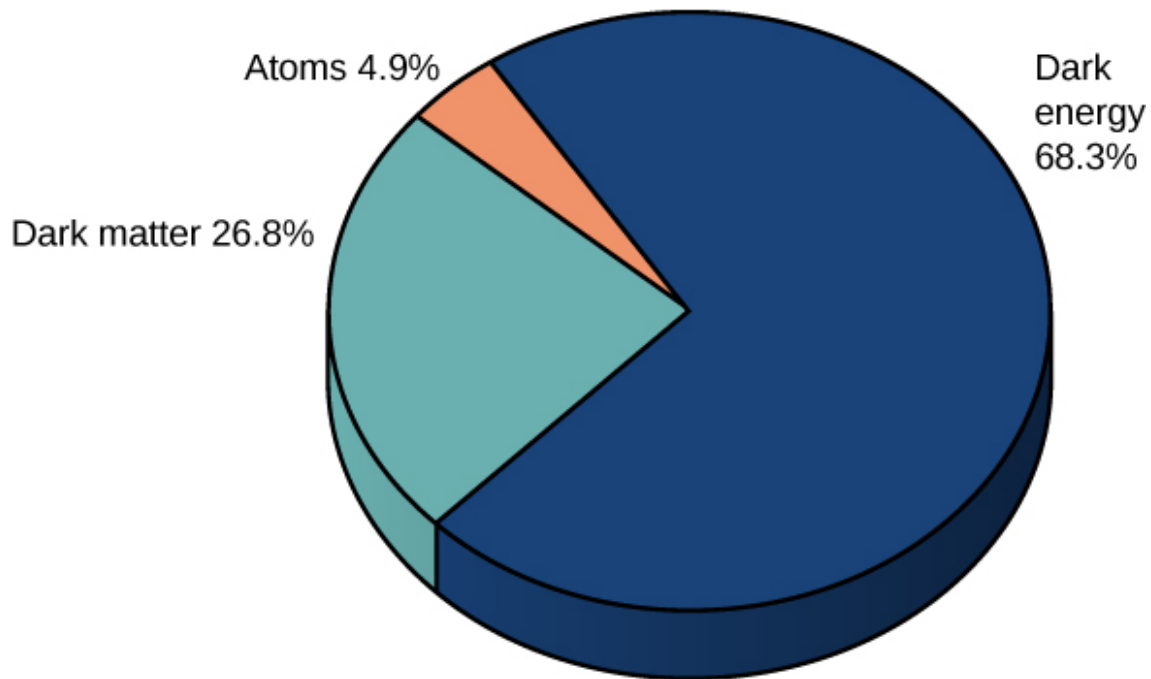
Light from a distant star is bent around a galaxy. Under the right conditions, two duplicate images of the same star can be seen.



Light from a distant star is bent around a galaxy. Under the right conditions, we can see a ring of light instead of a single star. (credit: modification of work by ESA/Hubble & NASA)

Increasingly precise astronomical measurements of the expanding universe also reveal the presence of a new form of energy called **dark energy**. This energy is thought to explain larger-than-expected values for the observed galactic redshifts for distant galaxies. These redshifts suggest that the universe is not only expanding, but expanding at an increasing rate. Virtually nothing is known about the nature and properties of dark energy. Together, dark energy and dark matter represent two of the most interesting and unsolved puzzles of modern physics. Scientists attribute 68.3% of the energy of the universe to dark energy, 26.8% to dark matter, and just 4.9% to the mass-energy of ordinary particles ([\[link\]](#)). Given the current great mystery over the nature of dark matter and dark energy, Isaac Newton's humble words are as true now as they were centuries ago:

“I do not know what I may appear to the world, but to myself I seem to have been only like a boy playing on the sea-shore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.”



Estimated distribution of matter and energy in the universe. (credit: NASA/WMAP Science Team)

Summary

- The early universe was hot and dense.
- The universe is isotropic and expanding.
- Cosmic background radiation is evidence for the Big Bang.
- The vast portion of the mass and energy of the universe is not well understood.

Key Equations

Momentum of a charged particle in a cyclotron	$p = 0.3Br$
Center-of-mass energy of a colliding beam machine	$W^2 = 2[E_1E_2 + (p_1c)(p_2c)] + (m_1c^2)^2 + (m_2c^2)^2$
Approximate time for exchange of a virtual particle between two other particles	$\Delta t = \frac{h}{E}$
Hubble's law	$v = H_0d$
Cosmological space-time metric	$ds^2 = c^2dt^2 - a(t)^2d\Sigma^2$

Conceptual Questions

Exercise:

Problem:

What is meant by a “cosmological model of the early universe?” Briefly describe this model in terms of the four fundamental forces.

Exercise:

Problem:

Describe two pieces of evidence that support the Big Bang model.

Solution:

The observed expansion of the universe and the cosmic background radiation spectrum.

Exercise:

Problem:

In what sense are we, as Newton once said, “a boy playing on the sea-shore”? Express your answer in terms of the concepts of dark matter and dark energy.

Exercise:

Problem:

If some unknown cause of redshift—such as light becoming “tired” from traveling long distances through empty space—is discovered, what effect would that have on cosmology?

Solution:

If light slow down, it takes long to reach Earth than expected. We conclude that the object is much closer than it really is. Thus, for every recessional velocity (based on the frequency of light, which we assume is not disturbed by the slowing), the distance is smaller than the “true” value, Hubble’s constant is larger than the “true” value, and the age of the universe is smaller than the “true” value.

Exercise:

Problem:

In the past, many scientists believed the universe to be infinite. However, if the universe is infinite, then any line of sight should eventually fall on a star’s surface and the night sky should be very bright. How is this paradox resolved in modern cosmology?

Additional Problems

Exercise:

Problem:

Experimental results suggest that a muon decays to an electron and photon. How is this possible?

Exercise:**Problem:**

Each of the following reactions is missing a single particle. Identify the missing particle for each reaction.

- (a) $p + \bar{p} \rightarrow n + ?$
- (b) $p + p \rightarrow p + \Lambda^0 + ?$
- (c) $\pi^- + p \rightarrow \Sigma^- + ?$
- (d) $K^- + n \rightarrow \Lambda^0 + ?$
- (e) $\tau^+ \rightarrow e^+ + \nu_e + ?$
- (f) $\bar{\nu}_e + p \rightarrow n + ?$

Solution:

a. \bar{n} ; b. K^+ ; c. K^+ ; d. π^- ; e. $\bar{\nu}_\tau$; f. e^+

Exercise:**Problem:**

Because of energy loss due to synchrotron radiation in the LHC at CERN, only 5.00 MeV is added to the energy of each proton during each revolution around the main ring. How many revolutions are needed to produce 7.00-TeV (7000 GeV) protons, if they are injected with an initial energy of 8.00 GeV?

Exercise:

Problem:

A proton and an antiproton collide head-on, with each having a kinetic energy of 7.00 TeV (such as in the LHC at CERN). How much collision energy is available, taking into account the annihilation of the two masses? (Note that this is not significantly greater than the extremely relativistic kinetic energy.)

Solution:

$$14.002 \text{ TeV} \approx 14.0 \text{ TeV}$$

Exercise:**Problem:**

When an electron and positron collide at the SLAC facility, they each have 50.0-GeV kinetic energies. What is the total collision energy available, taking into account the annihilation energy? Note that the annihilation energy is insignificant, because the electrons are highly relativistic.

Exercise:**Problem:**

The core of a star collapses during a supernova, forming a neutron star. Angular momentum of the core is conserved, so the neutron star spins rapidly. If the initial core radius is $5.0 \times 10^5 \text{ km}$ and it collapses to 10.0 km, find the neutron star's angular velocity in revolutions per second, given the core's angular velocity was originally 1 revolution per 30.0 days.

Solution:

$$964 \text{ rev/s}$$

Exercise:

Problem:

Using the solution from the previous problem, find the increase in rotational kinetic energy, given the core's mass is 1.3 times that of our Sun. Where does this increase in kinetic energy come from?

Exercise:**Problem:**

(a) What Hubble constant corresponds to an approximate age of the universe of 10^{10} y? To get an approximate value, assume the expansion rate is constant and calculate the speed at which two galaxies must move apart to be separated by 1 Mly (present average galactic separation) in a time of 10^{10} y. (b) Similarly, what Hubble constant corresponds to a universe approximately 2×10^{10} years old?

Solution:

$$\text{a. } H_0 = \frac{30 \text{ km/s}}{1 \text{ Mly}} = 30 \text{ km/s} \cdot \text{Mly}; \text{ b. } H_0 = \frac{15 \text{ km/s}}{1 \text{ Mly}} = 15 \text{ km/s} \cdot \text{Mly}$$

Challenge Problems**Exercise:****Problem:**

Electrons and positrons are collided in a circular accelerator. Derive the expression for the center-of-mass energy of the particle.

Exercise:**Problem:**

The intensity of cosmic ray radiation decreases rapidly with increasing energy, but there are occasionally extremely energetic cosmic rays that create a shower of radiation from all the particles they create by striking a nucleus in the atmosphere. Suppose a cosmic ray particle having an energy of 10^{10} GeV converts its energy into particles with masses averaging $200 \text{ MeV}/c^2$.

(a) How many particles are created? (b) If the particles rain down on a 1.00-km^2 area, how many particles are there per square meter?

Solution:

a. 5×10^{10} ; b. divide the number of particles by the area they hit:
 $5 \times 10^4 \text{ particles/m}^2$

Exercise:

Problem:

(a) Calculate the relativistic quantity $\gamma = \frac{1}{\sqrt{1-v^2/c^2}}$ for 1.00-TeV protons produced at Fermilab. (b) If such a proton created a π^+ having the same speed, how long would its life be in the laboratory? (c) How far could it travel in this time?

Exercise:

Problem:

Plans for an accelerator that produces a secondary beam of K mesons to scatter from nuclei, for the purpose of studying the strong force, call for them to have a kinetic energy of 500 MeV. (a) What would the relativistic quantity $\gamma = \frac{1}{\sqrt{1-v^2/c^2}}$ be for these particles? (b) How long would their average lifetime be in the laboratory? (c) How far could they travel in this time?

Solution:

a. 2.01; b. $2.50 \times 10^{-8} \text{ s}$; c. 6.50 m

Exercise:

Problem:

In supernovae, neutrinos are produced in huge amounts. They were detected from the 1987A supernova in the Magellanic Cloud, which is about 120,000 light-years away from Earth (relatively close to our Milky Way Galaxy). If neutrinos have a mass, they cannot travel at the speed of light, but if their mass is small, their velocity would be almost that of light. (a) Suppose a neutrino with a $7\text{-eV}/c^2$ mass has a kinetic energy of 700 keV. Find the relativistic quantity $\gamma = \frac{1}{\sqrt{1-v^2/c^2}}$ for it. (b) If the neutrino leaves the 1987A supernova at the same time as a photon and both travel to Earth, how much sooner does the photon arrive? This is not a large time difference, given that it is impossible to know which neutrino left with which photon and the poor efficiency of the neutrino detectors. Thus, the fact that neutrinos were observed within hours of the brightening of the supernova only places an upper limit on the neutrino's mass. (*Hint:* You may need to use a series expansion to find v for the neutrino, since its γ is so large.)

Exercise:**Problem:**

Assuming a circular orbit for the Sun about the center of the Milky Way Galaxy, calculate its orbital speed using the following information: The mass of the galaxy is equivalent to a single mass 1.5×10^{11} times that of the Sun (or 3×10^{41} kg), located 30,000 ly away.

Solution:

$$\frac{mv^2}{r} = \frac{GMm}{r^2} \Rightarrow$$

$$v = \left(\frac{GM}{r} \right)^{1/2} = \left[\frac{(6.67 \times 10^{-11} \text{ N}\cdot\text{m}^2/\text{kg}^2)(3 \times 10^{41} \text{ kg})}{(30,000 \text{ ly})(9.46 \times 10^{15} \text{ m/ly})} \right] = 2.7 \times 10^5 \text{ m/s}$$

Exercise:

Problem:

(a) What is the approximate force of gravity on a 70-kg person due to the Andromeda Galaxy, assuming its total mass is 10^{13} that of our Sun and acts like a single mass 0.613 Mpc away? (b) What is the ratio of this force to the person's weight? Note that Andromeda is the closest large galaxy.

Exercise:**Problem:**

(a) A particle and its antiparticle are at rest relative to an observer and annihilate (completely destroying both masses), creating two γ rays of equal energy. What is the characteristic γ -ray energy you would look for if searching for evidence of proton-antiproton annihilation? (The fact that such radiation is rarely observed is evidence that there is very little antimatter in the universe.) (b) How does this compare with the 0.511-MeV energy associated with electron-positron annihilation?

Solution:

a. 938.27 MeV; b. 1.84×10^3

Exercise:**Problem:**

The peak intensity of the CMBR occurs at a wavelength of 1.1 mm. (a) What is the energy in eV of a 1.1-mm photon? (b) There are approximately 10^9 photons for each massive particle in deep space. Calculate the energy of 10^9 such photons. (c) If the average massive particle in space has a mass half that of a proton, what energy would be created by converting its mass to energy? (d) Does this imply that space is "matter dominated"? Explain briefly.

Exercise:

Problem:

(a) Use the Heisenberg uncertainty principle to calculate the uncertainty in energy for a corresponding time interval of 10^{-43} s. (b) Compare this energy with the 10^{19} GeV unification-of-forces energy and discuss why they are similar.

Solution:

a. $3.29 \times 10^{18} \text{ GeV} \approx 3 \times 10^{18} \text{ GeV}$; b. 0.3; Unification of the three forces breaks down shortly after the separation of gravity from the unification force (near the Planck time interval). The uncertainty in time then becomes greater. Hence the energy available becomes less than the needed unification energy.

Glossary

cosmic microwave background radiation (CMBR)

thermal radiation produced by the Big Bang event

dark energy

form of energy believed to be responsible for the observed acceleration of the universe

dark matter

matter in the universe that does not interact with other particles but that can be inferred by deflection of distance star light

nucleosynthesis

creation of heavy elements, occurring during the Big Bang

Units

Quantity	Common Symbol	Unit	Unit in Terms of Base SI Units
Acceleration	\vec{a}	m/s ²	m/s ²
Amount of substance	n	mole	mol
Angle	θ, ϕ	radian (rad)	
Angular acceleration	$\vec{\alpha}$	rad/s ²	s ⁻²
Angular frequency	ω	rad/s	s ⁻¹
Angular momentum	\vec{L}	kg · m ² /s	kg · m ² /s
Angular velocity	$\vec{\omega}$	rad/s	s ⁻¹
Area	A	m ²	m ²
Atomic number	Z		
Capacitance	C	farad (F)	A ² · s ⁴ /kg · m ²
Charge	q, Q, e	coulomb (C)	A · s
Charge density:			
Line	λ	C/m	A · s/m
Surface	σ	C/m ²	A · s/m ²

Quantity	Common Symbol	Unit	Unit in Terms of Base SI Units
Volume	ρ	C/m ³	A · s/m ³
Conductivity	σ	1/Ω · m	A ² · s ³ /kg · m ³
Current	I	ampere	A
Current density	\vec{J}	A/m ²	A/m ²
Density	ρ	kg/m ³	kg/m ³
Dielectric constant	κ		
Electric dipole moment	\vec{p}	C · m	A · s · m
Electric field	\vec{E}	N/C	kg · m/A · s ³
Electric flux	Φ	N · m ² /C	kg · m ³ /A · s ³
Electromotive force	ε	volt (V)	kg · m ² /A · s ³
Energy	E, U, K	joule (J)	kg · m ² /s ²
Entropy	S	J/K	kg · m ² /s ² · K
Force	\vec{F}	newton (N)	kg · m/s ²
Frequency	f	hertz (Hz)	s ⁻¹
Heat	Q	joule (J)	kg · m ² /s ²
Inductance	L	henry (H)	kg · m ² /A ² · s ²
Length:	ℓ, L	meter	m
Displacement	$\Delta x, \Delta \vec{r}$		

Quantity	Common Symbol	Unit	Unit in Terms of Base SI Units
Distance	d, h		
Position	x, y, z, \vec{r}		
Magnetic dipole moment	$\vec{\mu}$	N · J/T	A · m ²
Magnetic field	\vec{B}	tesla (T) = Wb/m ²	kg/A · s ²
Magnetic flux	Φ_m	weber (Wb)	kg · m ² /A · s ²
Mass	m, M	kilogram	kg
Molar specific heat	C	J/mol · K	kg · m ² /s ² · mol · K
Moment of inertia	I	kg · m ²	kg · m ²
Momentum	\vec{p}	kg · m/s	kg · m/s
Period	T	s	s
Permeability of free space	μ_0	N/A ² = (H/m)	kg · m/A ² · s ²
Permittivity of free space	ϵ_0	C ² /N · m ² = (F/m)	A ² · s ⁴ /kg · m ³
Potential	V	volt (V) = (J/C)	kg · m ² /A · s ³
Power	P	watt (W) = (J/s)	kg · m ² /s ³
Pressure	p	pastcal (P) = N/m ²	kg/m · s ²
Resistance	R	ohm (Ω) = (V/A)	kg · m ² /A ² · s ³

Quantity	Common Symbol	Unit	Unit in Terms of Base SI Units
Specific heat	c	J/kg · K	$\text{m}^2/\text{s}^2 \cdot \text{K}$
Speed	v	m/s	m/s
Temperature	T	kelvin	K
Time	t	second	s
Torque	$\vec{\tau}$	N · m	$\text{kg} \cdot \text{m}^2/\text{s}^2$
Velocity	\vec{v}	m/s	m/s
Volume	V	m^3	m^3
Wavelength	λ	m	m
Work	W	joule (J) = (N · m)	$\text{kg} \cdot \text{m}^2/\text{s}^2$

Units Used in Physics (Fundamental units in bold)

Conversion Factors

	m	cm	km
1 meter	1	10^2	10^{-3}
1 centimeter	10^{-2}	1	10^{-5}
1 kilometer	10^3	10^5	1
1 inch	2.540×10^{-2}	2.540	2.540×10^{-5}
1 foot	0.3048	30.48	3.048×10^{-4}
1 mile	1609	1.609×10^4	1.609
1 angstrom	10^{-10}		
1 fermi	10^{-15}		
1 light-year			9.460×10^{12}
	in.	ft	mi
1 meter	39.37	3.281	6.214×10^{-4}
1 centimeter	0.3937	3.281×10^{-2}	6.214×10^{-6}
1 kilometer	3.937×10^4	3.281×10^3	0.6214
1 inch	1	8.333×10^{-2}	1.578×10^{-5}
1 foot	12	1	1.894×10^{-4}
1 mile	6.336×10^4	5280	1

Length

Area

$$1 \text{ cm}^2 = 0.155 \text{ in.}^2$$

$$1 \text{ m}^2 = 10^4 \text{ cm}^2 = 10.76 \text{ ft}^2$$

$$1 \text{ in.}^2 = 6.452 \text{ cm}^2$$

$$1 \text{ ft}^2 = 144 \text{ in.}^2 = 0.0929 \text{ m}^2$$

Volume

$$1 \text{ liter} = 1000 \text{ cm}^3 = 10^{-3} \text{ m}^3 = 0.03531 \text{ ft}^3 = 61.02 \text{ in.}^3$$

$$1 \text{ ft}^3 = 0.02832 \text{ m}^3 = 28.32 \text{ liters} = 7.477 \text{ gallons}$$

$$1 \text{ gallon} = 3.788 \text{ liters}$$

	s	min	h	day	yr
1 second	1	1.667×10^{-2}	2.778×10^{-4}	1.157×10^{-5}	3.169×10^{-8}
1 minute	60	1	1.667×10^{-2}	6.944×10^{-4}	1.901×10^{-6}
1 hour	3600	60	1	4.167×10^{-2}	1.141×10^{-4}
1 day	8.640×10^4	1440	24	1	2.738×10^{-3}
1 year	3.156×10^7	5.259×10^5	8.766×10^3	365.25	1

Time

	m/s	cm/s	ft/s	mi/h
1 meter/second	1	10^2	3.281	2.237
1 centimeter/second	10^{-2}	1	3.281×10^{-2}	2.237×10^{-2}
1 foot/second	0.3048	30.48	1	0.6818
1 mile/hour	0.4470	44.70	1.467	1

Speed

Acceleration

$$1 \text{ m/s}^2 = 100 \text{ cm/s}^2 = 3.281 \text{ ft/s}^2$$

$$1 \text{ cm/s}^2 = 0.01 \text{ m/s}^2 = 0.03281 \text{ ft/s}^2$$

$$1 \text{ ft/s}^2 = 0.3048 \text{ m/s}^2 = 30.48 \text{ cm/s}^2$$

$$1 \text{ mi/h} \cdot \text{s} = 1.467 \text{ ft/s}^2$$

	kg	g	slug	u
1 kilogram	1	10^3	6.852×10^{-2}	6.024×10^{26}
1 gram	10^{-3}	1	6.852×10^{-5}	6.024×10^{23}
1 slug	14.59	1.459×10^4	1	8.789×10^{27}
1 atomic mass unit	1.661×10^{-27}	1.661×10^{-24}	1.138×10^{-28}	1
1 metric ton	1000			

Mass

	N	dyne	lb
1 newton	1	10^5	0.2248
1 dyne	10^{-5}	1	2.248×10^{-6}
1 pound	4.448	4.448×10^5	1

Force

	Pa	dyne/cm²	atm	cmHg	lb/in.²
1 pascal	1	10	9.869×10^{-6}	7.501×10^{-4}	1.450×10^{-4}
1 dyne/centimeter ²	10^{-1}	1	9.869×10^{-7}	7.501×10^{-5}	1.450×10^{-5}
1 atmosphere	1.013×10^5	1.013×10^6	1	76	14.70
1 centimeter mercury*	1.333×10^3	1.333×10^4	1.316×10^{-2}	1	0.1934
1 pound/inch ²	6.895×10^3	6.895×10^4	6.805×10^{-2}	5.171	1
1 bar	10^5				
1 torr				1 (mmHg)	
*Where the acceleration due to gravity is 9.80665 m/s^2 and the temperature is 0°C					

Pressure

	J	erg	ft.lb
1 joule	1	10^7	0.7376
1 erg	10^{-7}	1	7.376×10^{-8}
1 foot-pound	1.356	1.356×10^7	1
1 electron-volt	1.602×10^{-19}	1.602×10^{-12}	1.182×10^{-19}
1 calorie	4.186	4.186×10^7	3.088
1 British thermal unit	1.055×10^3	1.055×10^{10}	7.779×10^2
1 kilowatt-hour	3.600×10^6		
	eV	cal	Btu
1 joule	6.242×10^{18}	0.2389	9.481×10^{-4}
1 erg	6.242×10^{11}	2.389×10^{-8}	9.481×10^{-11}
1 foot-pound	8.464×10^{18}	0.3239	1.285×10^{-3}
1 electron-volt	1	3.827×10^{-20}	1.519×10^{-22}
1 calorie	2.613×10^{19}	1	3.968×10^{-3}
1 British thermal unit	6.585×10^{21}	2.520×10^2	1

Work, Energy, Heat

Power

$$1 \text{ W} = 1 \text{ J/s}$$

$$1 \text{ hp} = 746 \text{ W} = 550 \text{ ft} \cdot \text{lb/s}$$

$$1 \text{ Btu/h} = 0.293 \text{ W}$$

Angle

$$1 \text{ rad} = 57.30^\circ = 180^\circ/\pi$$

$$1^\circ = 0.01745 \text{ rad} = \pi/180 \text{ rad}$$

$$1 \text{ revolution} = 360^\circ = 2\pi \text{ rad}$$

$$1 \text{ rev/min (rpm)} = 0.1047 \text{ rad/s}$$

Fundamental Constants

Quantity	Symbol	Value
Atomic mass unit	u	$1.660\,538\,782\,(83) \times 10^{-27}\,\text{kg}$ $931.494\,028\,(23)\,\text{MeV}/c^2$
Avogadro's number	N_A	$6.02214076 \times 10^{23}$ reciprocal mole(mol^{-1})
Bohr magneton	$\mu_B = \frac{e\hbar}{2m_e}$	$9.274\,009\,15\,(23) \times 10^{-24}\,\text{J/T}$
Bohr radius	$a_0 = \frac{\hbar^2}{m_e e^2 k_e}$	$5.291\,772\,085\,9\,(36) \times 10^{-11}\,\text{m}$
Boltzmann's constant	$k_B = \frac{R}{N_A}$	1.380649×10^{-23} joule per kelvin($\text{J} \cdot \text{K}^{-1}$)
Compton wavelength	$\lambda_C = \frac{h}{m_e c}$	$2.426\,310\,217\,5\,(33) \times 10^{-12}\,\text{m}$
Coulomb constant	$k_e = \frac{1}{4\pi\epsilon_0}$	$8.987\,551\,788\dots \times 10^9\,\text{N} \cdot \text{m}^2/\text{C}^2$ (exact)
Deuteron mass	m_d	$3.343\,583\,20\,(17) \times 10^{-27}\,\text{kg}$ $2.013\,553\,212\,724\,(78)\,\text{u}$ $1875.612\,859\,\text{MeV}/c^2$
Electron mass	m_e	$9.109\,382\,15\,(45) \times 10^{-31}\,\text{kg}$ $5.485\,799\,094\,3\,(23) \times 10^{-4}\,\text{u}$ $0.510\,998\,910\,(13)\,\text{MeV}/c^2$
Electron volt	eV	$1.602\,176\,487\,(40) \times 10^{-19}\,\text{J}$

Quantity	Symbol	Value
Elementary charge	e	$1.602176634 \times 10^{-19} \text{ C}$
Gas constant	R	$8.314\,472\,(15) \text{ J/mol} \cdot \text{K}$
Gravitational constant	G	$6.674\,28\,(67) \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$
Neutron mass	m_n	$1.674\,927\,211\,(84) \times 10^{-27} \text{ kg}$ $1.008\,664\,915\,97\,(43) \text{ u}$ $939.565\,346\,(23) \text{ MeV}/c^2$
Nuclear magneton	$\mu_n = \frac{e\hbar}{2m_p}$	$5.050\,783\,24\,(13) \times 10^{-27} \text{ J/T}$
Permeability of free space	μ_0	$4\pi \times 10^{-7} \text{ T} \cdot \text{m/A (exact)}$
Permittivity of free space	$\epsilon_0 = \frac{1}{\mu_0 c^2}$	$8.854\,187\,817... \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2 \text{ (exact)}$
Planck's constant	h $\hbar = \frac{h}{2\pi}$	$6.62607015 \times 10^{-34} \text{ kg} \cdot \text{m}^2 \cdot \text{s}^{-1}$ $1.05457182 \times 10^{-34} \text{ kg} \cdot \text{m}^2 \cdot \text{s}^{-1}$
Proton mass	m_p	$1.672\,621\,637\,(83) \times 10^{-27} \text{ kg}$ $1.007\,276\,466\,77\,(10) \text{ u}$ $938.272\,013\,(23) \text{ MeV}/c^2$
Rydberg constant	R_H	$1.097\,373\,156\,852\,7\,(73) \times 10^7 \text{ m}^{-1}$
Speed of light in vacuum	c	$2.997\,924\,58 \times 10^8 \text{ m/s (exact)}$

Fundamental Constants*Note:* These constants are the values recommended in 2006 by CODATA, based on a least-squares adjustment of data from different measurements. The numbers in parentheses for the values represent the uncertainties of the last two digits.

Useful combinations of constants for calculations:

$$hc = 12,400 \text{ eV} \cdot \text{\AA} = 1240 \text{ eV} \cdot \text{nm} = 1240 \text{ MeV} \cdot \text{fm}$$

$$\hbar c = 1973 \text{ eV} \cdot \text{\AA} = 197.3 \text{ eV} \cdot \text{nm} = 197.3 \text{ MeV} \cdot \text{fm}$$

$$k_e e^2 = 14.40 \text{ eV} \cdot \text{\AA} = 1.440 \text{ eV} \cdot \text{nm} = 1.440 \text{ MeV} \cdot \text{fm}$$

$$k_B T = 0.02585 \text{ eV at } T = 300 \text{ K}$$

Astronomical Data

Celestial Object	Mean Distance from Sun (million km)	Period of Revolution (d = days) (y = years)	Period of Rotation at Equator	Eccentricity of Orbit
Sun	–	–	27 d	–
Mercury	57.9	88 d	59 d	0.206
Venus	108.2	224.7 d	243 d	0.007
Earth	149.6	365.25 d	23 h 56 min 4 s	0.017
Mars	227.9	687 d	24 h 37 min 23 s	0.093
Jupiter	778.4	11.9 y	9 h 50 min 30 s	0.048
Saturn	1426.7	29.5 y	10 h 14 min	0.054
Uranus	2871.0	84.0 y	17 h 14 min	0.047
Neptune	4498.3	164.8 y	16 h	0.009

Celestial Object	Mean Distance from Sun (million km)	Period of Revolution (d = days) (y = years)	Period of Rotation at Equator	Eccentricity of Orbit
Earth's Moon	149.6 (0.386 from Earth)	27.3 d	27.3 d	0.055
Celestial Object	Equatorial Diameter (km)	Mass (Earth = 1)	Density (g/cm³)	
Sun	1,392,000	333,000.00	1.4	
Mercury	4879	0.06	5.4	
Venus	12,104	0.82	5.2	
Earth	12,756	1.00	5.5	
Mars	6794	0.11	3.9	
Jupiter	142,984	317.83	1.3	
Saturn	120,536	95.16	0.7	
Uranus	51,118	14.54	1.3	
Neptune	49,528	17.15	1.6	
Earth's Moon	3476	0.01	3.3	

Astronomical Data

Other Data:

Mass of Earth: 5.97×10^{24} kg

Mass of the Moon: 7.36×10^{22} kg

Mass of the Sun: 1.99×10^{30} kg

Chemistry

Periodic Table of the Elements

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H 1.008 hydrogen																	2 He 4.003 helium
2	3 Li 6.94 lithium	4 Be 9.012 beryllium											5 B 10.81 boron	6 C 12.01 carbon	7 N 14.01 nitrogen	8 O 16.00 oxygen	9 F 19.00 fluorine	10 Ne 20.18 neon
3	11 Na 22.99 sodium	12 Mg 24.31 magnesium											13 Al 26.98 aluminum	14 Si 28.09 silicon	15 P 30.97 phosphorus	16 S 32.06 sulfur	17 Cl 35.45 chlorine	18 Ar 39.95 argon
4	19 K 39.10 potassium	20 Ca 40.08 calcium	21 Sc 44.96 scandium	22 Ti 47.87 titanium	23 V 50.94 vanadium	24 Cr 52.00 chromium	25 Mn 54.94 manganese	26 Fe 55.85 iron	27 Co 58.93 cobalt	28 Ni 58.69 nickel	29 Cu 63.55 copper	30 Zn 65.38 zinc	31 Ga 69.72 gallium	32 Ge 72.63 germanium	33 As 74.92 arsenic	34 Se 78.97 selenium	35 Br 79.90 bromine	36 Kr 83.80 krypton
5	37 Rb 85.47 rubidium	38 Sr 87.62 strontium	39 Y 88.91 yttrium	40 Zr 91.22 zirconium	41 Nb 92.91 niobium	42 Mo 95.95 molybdenum	43 Tc [97] technetium	44 Ru 101.1 ruthenium	45 Rh 102.9 rhodium	46 Pd 106.4 palladium	47 Ag 107.9 silver	48 Cd 112.4 cadmium	49 In 114.8 indium	50 Sn 118.7 tin	51 Sb 121.8 antimony	52 Te 127.6 tellurium	53 I 126.9 iodine	54 Xe 131.3 xenon
6	55 Cs 132.9 cesium	56 Ba 137.3 barium	57-71 La-Lu *	72 Hf 178.5 hafnium	73 Ta 180.9 tantalum	74 W 183.8 tungsten	75 Re 186.2 rhenium	76 Os 190.2 osmium	77 Ir 192.2 iridium	78 Pt 195.1 platinum	79 Au 197.0 gold	80 Hg 200.6 mercury	81 Tl 204.4 thallium	82 Pb 207.2 lead	83 Bi 209.0 bismuth	84 Po [209] polonium	85 At [210] astatine	86 Rn [222] radon
7	87 Fr [223] francium	88 Ra [226] radium	89-103 Ac-Lr **	104 Rf [267] rutherfordium	105 Db [270] dubnium	106 Sg [271] seaborgium	107 Bh [270] bohrium	108 Hs [277] hassium	109 Mt [276] meitnerium	110 Ds [281] darmstadtium	111 Rg [282] roentgenium	112 Cn [285] copernicium	113 Uut [285] ununtrium	114 Fl [289] flerovium	115 Uup [288] ununpentium	116 Lv [293] livermorium	117 Uus [294] ununseptium	118 Uuo [294] ununoctium

57 La 138.9 lanthanum	58 Ce 140.1 cerium	59 Pr 140.9 praseodymium	60 Nd 144.2 neodymium	61 Pm [145] promethium	62 Sm 150.4 samarium	63 Eu 152.0 europium	64 Gd 157.3 gadolinium	65 Tb 158.9 terbium	66 Dy 162.5 dysprosium	67 Ho 164.9 holmium	68 Er 167.3 erbium	69 Tm 168.9 thulium	70 Yb 173.1 ytterbium	71 Lu 175.0 lutetium
89 Ac [227] actinium	90 Th 232.0 thorium	91 Pa 231.0 protactinium	92 U 238.0 uranium	93 Np [237] neptunium	94 Pu [244] plutonium	95 Am [243] americium	96 Cm [247] curium	97 Bk [247] berkelium	98 Cf [251] californium	99 Es [252] einsteinium	100 Fm [257] fermium	101 Md [258] mendelevium	102 No [259] nobelium	103 Lr [262] lawrencium

Atomic number → 1

Symbol → **H**

Atomic mass → 1.008

Name → hydrogen

Color Code	
<div></div>	Metal
<div></div>	Metalloid
<div></div>	Nonmetal
<div></div>	Solid
<div></div>	Liquid
<div></div>	Gas

The Greek Alphabet

Name	Capital	Lowercase	Name	Capital	Lowercase
Alpha	A	α	Nu	N	ν
Beta	B	β	Xi	Ξ	ξ
Gamma	Γ	γ	Omicron	O	o
Delta	Δ	δ	Pi	Π	π
Epsilon	E	ϵ	Rho	P	ρ
Zeta	Z	ζ	Sigma	Σ	σ
Eta	H	η	Tau	T	τ
Theta	Θ	θ	Upsilon	Υ	υ
Iota	I	ι	Phi	Φ	ϕ
Kappa	K	κ	Chi	X	χ
Lambda	Λ	λ	Psi	ψ	ψ
Mu	M	μ	Omega	Ω	ω

The Greek Alphabet